Quantum Statistics

Ole E. Barndorff-Nielsen
Aarhus University

Richard D. Gill
Leiden University

Mădălin I. Guţă
Nottingham University

Peter E. Jupp
University St. Andrews

September 4, 2008

Preface

The short title of this book, *Quantum Statistics*, could be amplified into: "statistical inference for quantum statistical models", the latter being *stochastic models for data obtained from observation or measurement of quantum systems*.

This explanation is necessary, since the title alone may well be confusing for some readers. In physics, the notion of "quantum statistics" is well established, and means "probability distributions derived from quantum mechanics which differ from their classical counterparts". For instance, the physicist will think of certain non-Poissonian models for counts. However, we are going to pay attention both to statistical modelling and to statistical inference, so the double meaning of the title is not that misleading.

Niels Bohr said that anyone who claims to understand quantum theory, clearly doesn't. And according to John von Neumann, you don't understand a new mathematical theory, you just get accustomed to it. So the best way to get into our subject is to get into a concrete problem as soon as possible. We need to get accustomed to strange notations, possibly unfamiliar mathematics, and remarkable phenomena. But before this, here are some clues to the overall scheme of things.

We are going to study statistical problems, for instance parameter estimation, for statistical models coming from quantum physics. This is meaningful since, according to our point of view, quantum physics is a *stochastic theory*: it tells us the *probability distribution* of data coming from an experiment involving a quantum system. It never predicts the actual outcome (except in some special situations where probabilities are zero or one). Moreover, the probability model for a given quantum experiment will typically depend on unknown parameters, in particular describing some aspects of the state of the quantum system being measured, or of the measurement apparatus.

Here the word "experiment" may be taken in a very broad sense, just as we are used to when we say that probability theory provides a mathematical model for "chance experiments". We will also talk, equally loosely for the time being, about a *measurement* on the quantum system.

So, what is a quantum system? The following definition might be unsatisfactory, but it is the best we can do: it is a physical system which behaves according to the laws of quantum physics. And according to Bohr, von Neumann, and many others, one cannot understand those laws, one can only hope to get used to them.

That quantum system might be an atom, an electron, a photon. It might be an assemblage of such objects. (In cosmology, it is the whole universe, though much of what we are about to say will then become problematic). For the time being, let's not worry about *what* we are talking about, but describe the *language* we will be using. In this language, there will be reference to a quantum system "inside" and a classical real world "outside" (we are not going to discuss cosmology). On its own, the quantum system will evolve, in time, according to certain laws. Without any interaction between the system and the outside world there is nothing interesting to talk about. We only have any experience of the quantum system through its effects on the outside world. At the same time, the interaction between quantum system and outside world will change the quantum system in some way.

Measuring a quantum system can therefore generate, according to some probability distribution, real data in the outside world. The distribution of the data depends on two aspects of the situation: the *state* of the quantum system, and *how* it is measured. In principle, there could be unknown parameters in both of these components. But from this point on, everything is classical, even though some of the most natural and beautiful statistical models are not familiar to statisticians, and often not well understood by physicists either.

So if we are just going to look at classical statistical models, why make something special out of quantum statistics?

Three things need to be said in answer. Firstly, the laws of quantum physics delineate rather sharply all possible experiments which can be done. Problems in the *design of experiments* can therefore be formulated very precisely, and often (because of the nice geometry of the class of all possible experiments) they have rather beautiful, surprising, and useful solutions.

Secondly, this same fact, that the collection of possible experiments can be described so sharply, means that the different probability models for all potential experiments are interrelated in a very tight and beautiful way. They are constrained by a rich mathematical structure which in a pure mathematical sense involves a *generalization* of familiar structures from classical probability theory. It pays to recognise this structure and to exploit it. Going further, from probability to statistics, the structure implies the existence of special interrelations between statistical concepts connected to the various possible experiments, and this interrelation can be formulated in terms of mathematical generalizations of various fundamental notions from statistics: for instance, score function, Fisher information, and so on.

Thirdly, quantum physics exhibits many strange features (superposition, complementarity, entanglement) and these features turn out also to have statistical repercussions. The real motivation for this book is to explore these wonderful phenomena.

Acknowledgements

Here we need to acknowledge a lot of people's help and support! Also some organisations, e.g., various EU programmes.

Contents

1

Introduction

Quantum mechanics is the foundation of much of modern physics. The theory is intrinsically stochastic: it does not predict what will happen in any given physical situation, but in principle only allows one to compute a probability distribution over a number of possibilities.

However, the interpretation of this distribution truly as a *probability* distribution only becomes inescapable when the theory is applied to a single quantum system, or to a small number of identical copies, and not to a *huge* number of identical copies (an ensemble, as physicists say). It is only very recently that this situation has become achievable in the laboratory, and also of burning interest, both to theoreticians and to experimentalists.

The reason for this new interest is because of the recent birth of a new field in physics called *quantum information*. The field is developing at a breathtaking pace, and this development is at least partly motivated by a glimpse of extraordinary potential technology.

Section 1.1 of this chapter provides a snapshot of "what is going on" right now, though we are fully aware that this picture will already be quite out of date in a year from now. We use the snapshot to describe in a nontechnical way the kind of problems we are going to study. As we will see, within the general field of quantum information one can in particular identify statistical problems, problems concerning quantum statistical information. Section 1.2 contains a specific example of such a problem, which we will study in depth, later (to be precise, in Section 5.2) . In Section 1.3 we will give preliminary mathematical definitions of the key notions of state, measurement, operation, and instrument (to be studied in depth in Chapter 2). Finally, Section 1.4 of this chapter provides an outline of the rest of the book.

1.1 The ultimate laptop

Gordon Moore, co-founder of Intel, noted in 1965 that the density of transistors on integrated circuits had doubled every year since the integrated circuit was invented. He saw no reason for this trend to slow down. Nowadays one is a little more conservative: Moore's law states that data density doubles approximately every 18 months. It has been doing this for 50 years and the trend is expected to continue

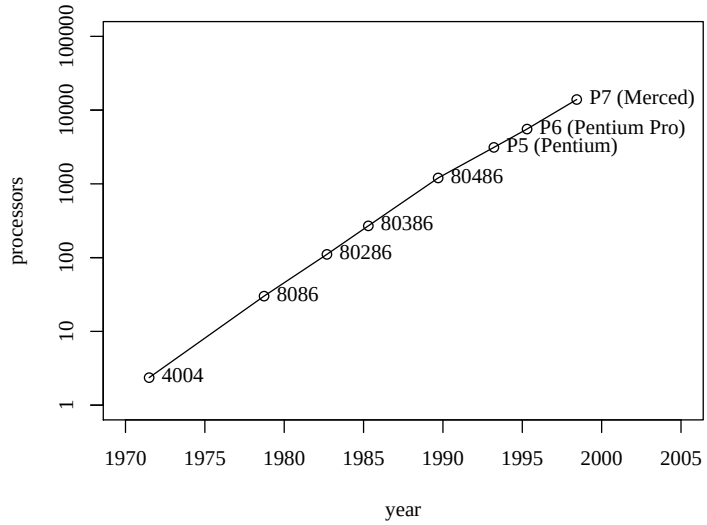

Fig. 1.1. Moore's Law—starting to slow down? Number of transistors on an Intel CPU (in thousands), versus time.

for another two decades. According to Figure 1.1 the doubling time is closer to two years, and the rate is beginning to markedly slow down.

Within twenty years however, computers operating according to Moore's law will have atomic level components. At this level the laws of physics are the laws of quantum physics and these involve uncertainty in an intrinsic way. There are no longer definite particles always at definite places as time evolves. Quantum physics only allows one to write down the probability to find a particle at some place at some time. A computer in which $1+1 \mod 2$ is sometimes 0, sometimes 1, does not seem like a good idea.

However, there are also features of quantum physics which might be useful for computation. According to quantum physics one has to describe a particle with a wave function. The squared amplitude of the wave is the probability density to find the particle at that position. The wave function evolves in time in a deterministic way. The evolution is linear so that a superposition of waves evolves as the superposition of its components. This promises a kind of parallelism: simultaneously doing the same computation on a whole collection of different inputs.

There is another striking feature called entanglement, according to which several particles together have different behaviour from anything which is attainable by simply adding together behaviours of separate particles. The total is more than the sum of its parts. Entanglement is connected to non-locality. It is as though one particle feels what is simultaneously happening to another, distant particle, with which it has interacted in the past. Unfortunately (or fortunately, since quantum mechanics should not contradict relativity theory), the particle finds it hard to express what it feels, and it is only later, when one compares notes, that one sees that

Fig. 1.2. Erwin Schrödinger (from MacTutor Archive, St. Andrews.)

Fig. 1.3. Richard Feynmann (from MacTutor Archive, St. Andrews).

it was talking sense. The situation is reminiscent of the oracles of antiquity who always correctly predicted the future, but did it in such a way that only in retrospect could their predictions be understood.

Entanglement remains a tantalizing and a controversial subject. It was discovered and named by Erwin Schrödinger (see Figure 1.2), one of the founding fathers of quantum physics. Schrödinger thought it was "schrecklich" (terrible). About quantum physics as a whole he said "I don't like it, and I'm sorry I ever had anything to do with it". Richard Feynmann (Figure 1.3, who already in the fifties started thinking about quantum computers, said "Don't try to understand quantum mechanics or you will fall into a black hole and never be heard from again".

Fortunately you don't have to understand quantum physics in order to understand quantum statistics (more precisely, quantum statistical inference). There is an elegant and simple mathematical model, which one can accept just as one might accept the rules of some new game of chess: just start to play the game, follow one's instinct, and see where it brings one. In view of Feynmann's advice it is probably best to postpone consideration of "what it all means" till after one is familiar with what comes out of it.

In the nineties theoretical physicists started thinking seriously about quantum

Fig. 1.4. Peter Shor (from his webpage).

computers. A few problems were invented (for instance, by David Deutsch) in which quantum computation offered a mild advantage over classical computation, by exploiting the parallelism of quantum physics (163). Moreover, computer scientists and mathematicians got involved. And things started moving really fast when the computer scientist Peter Shor (Figure 1.4), in 1994, showed how a computer architecture in which the basic elements of the memory were not "bits" but "qubits", could be used to factor large integers in polynomial time (344). Here, entanglement is used in a big way.

As everyone realises, it is easy, in principle, to multiply two large numbers together. One follows an algorithm due to Euclid, which used to be taught at primary school, and it just takes some time and a large piece of paper. If one thinks about it for a moment one will realise that the time it takes to do the calculation is not more than quadratic in the total length (number of decimals) of the numbers involved.

On the other hand, and also well known, the inverse problem is rather more difficult. Given a thousand digit number which is the product of two five hundred digit primes, it takes a much, much longer time to discover what those two factors are, than it takes to multiply them and get their product. Though it is not known for sure, mathematicians and computer scientists believe that this fact is intrinsic to the problem of factoring; it is not that we just weren't smart enough to come up with a good algorithm yet. The time it takes the best algorithm increases exponentially in the length of the number to be factored. And on this fact depends the security of the internet and of internet banking, of modern telecommunication systems, and of a huge amount more of modern technology.

These facts are intrinsic to the problem of factoring *together with* the basic limitations of classical computation. Analogue computation with classical physics does not help either. Computation is a physical process and according to classical physics, factoring is hard. Shor's algorithm uses clever tricks from algebra and group theory to convert the problem of factoring integers into a problem of computing a discrete Fourier transform. This problem turns out to be something which an

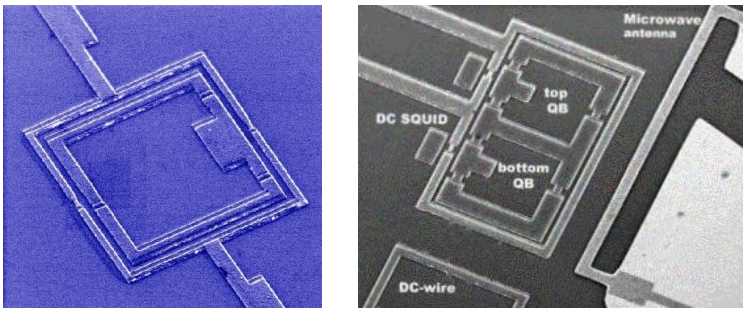

Fig. 1.5. Delft Qubits (from the Delft group's webpages)

array of quantum bits, on which a small number of basic quantum operations can be applied, can solve quite easily. The mathematics is beautiful and it's a brilliant discovery.

Shor's discovery led to an explosion of activity and of hype. Take a look for instance at this quotation from the respectable British popular science journal *New Scientist*, June 2002:

*The prize is a machine powerful enough to take on life, the Universe and everything. Justin Mullins comment*ates *on the race to build a quantum computer.*

Is it just hype? No, and there really is a race on. Albert Mooij's group in Delft is one of the world leaders. Figure 1.5 shows pictures of one, and of a pair, of Delft qubits. Each is about 5 μm (micrometers) across. A micrometer is one thousandth of a millimeter. The atomic scale, nanometers, is one thousand times smaller still. Such an apparatus is called a SQUID: a Semi-conducting Quantum Interference Device. It is actually an aluminium circuit, brought to very low temperature. It is broken in three places by Josephson junctions—these are a kind of dirty connections, across which electric current only flows with great difficulty. About a billion electrons in the circuit behave together as a single fundamental particle and can either be in a state of clockwise or counter-clockwise motion around the ring. Moreover, since they behave according to quantum mechanics, it is possible to get them into a superposition of those two states:

$$|\psi\rangle \;=\; \big(|\circlearrowleft\rangle + |\circlearrowright\rangle\big)/\sqrt{2}$$

The next aim is to create entanglement, such as the following state of two qubits: both loops clockwise, superimposed with both loops counterclockwise:

$$|\psi\rangle \;=\; \big(|\circlearrowleft\rangle \otimes |\circlearrowleft\rangle + |\circlearrowright\rangle \otimes |\circlearrowright\rangle\big)/\sqrt{2}$$

There truly is a race on: note the dates in the following two citations: Chiorescu, Nakamura, Harmans, and Mooij, Coherent Quantum Dynamics of *a* Superconducting Flux Qubit, *Science* **299** 1869-1871, March 21 (2003); and the American competition, Berkley et al., Entangled Macroscopic Quantum States in *Two* Superconducting Qubits, *Science Online* 10845281-0, May 15 (2003).

We emphasize that although a billion electrons are involved here, the quantum system being built is a two-level system, the quantum analogue of a bit (in the

second picture, two two-level systems; two bits). For a probabilist, a good analogy is a coin-toss, for which an appropriate mathematical model is the Bernoulli trial, depending on one parameter p.

From the wave-like nature of quantum mechanics it turns out that for a quantum coin toss (a measurement on a two level system), not one but three real numbers are needed to fully describe the state of the system. Moreover there are a continuum of different ways to toss the quantum coin, not just one, leading to a continuum of Bernoulli models, each one with a p depending on the state of the system, and the measurement which was done on it.

Despite this continuum of possible states and possible measurements, one can only extract one bit of classical information from such a system. The future quantum computer starts with an array of N qubits, altogether in one of 2^N basic states (each qubit ↄ for 0, or ↄ for 1). Their joint state evolves deterministically and after some fixed time one looks at each qubit and reads off a 0 or a 1. Just like a classical computer, in fact.

Now a two-bit computer cannot do very much. It is going to take quite a long time before we are able to build even 1000 bit quantum calculating machines, let alone the million bit memory of a modern PC. However a beautiful thing about quantum physics is that the basic mathematical rules are the same, whatever the physical system involved. Quantum computation, communication, and cryptographic protocols have already been implemented on ions in ion traps, photons from a laser travelling 15 Km through glass fibre cables, and in nuclear spins in molecules in solution in water. For instance, the number 15 has been factored using an enormously noisy quantum computation being done simultaneously on about 10^{22} 7 qubit quantum computers (7 nuclear spins in a chloroform molecule; the answer was ...). In the meantime there are a great deal of practical and theoretical problems to be solved, the main one being decoherence: interaction of the qubits of the quantum computer with its environment causes the fragile entangled state to decohere into boring, separate, classical states of the separate bits. On the other hand, reading data in and out of the computer, without which it is no use whatsover, requires ... interaction with the environment. Some physicists—among them, the recent Nobel-prizewinner Gerard 't Hooft—think the problems here may be intrinsic. Resolving this issue, which means finding out whether quantum mechanics really does apply to larger and larger systems, is the real scientific drive behind the efforts of many of the experimentalists and theorists in this field. The dream of a quantum computer has been a useful hype to secure publicity and research funding, but the emperor's new clothes are wearing thin these days.

If we ever do get a quantum computer, what will it actually be able to do? Most people felt ten years ago that Shor's breakthrough had simply exposed the tip of an iceberg. Today *Factoring*, tomorrow *The Travelling Salesman* ...! Despite huge efforts however, only a handful of new problems have been found to be amenable to "exponential speed up" on a quantum computer. All of those problems are very close in mathematical nature, despite appearances, to factoring, and the new algorithms are close relatives of Shor's. Quantum mechanics is good at doing Fourier transforms. A realistic outlook at the moment, is that within 10 years we will

have 30 qubit quantum computers, which are able to do ... theoretical quantum mechanical calculations exponentially faster than any classical computer can. The hardware will actually be a lattice of ions riding in the standing waves of interfering laser beams. The theory is being pioneered by Ignatio Cirac of the Max Planck Institute for Quantum Optics at Garching, near Munich. The scope of application might seem limited, but it would have enormous practical impact in molecular biology and nanotechnology in general.

The title of this section comes from some papers of the last couple of years by Seth Lloyd (MIT) which the reader might like to find by searching for the author at `http://arxiv.org/find/quant-ph`. He asks the question what is the maximum computing capacity allowed by the laws of physics, which can be done on one liter of matter, of mass one kilo? The ultimate laptop ... . The answer is ... that under Moore's law, it will take another 200 years to hit the ultimate boundary. By that time the CPU will be a little black hole and the user interface is going to be a bit tricky ... .

1.2 An example: entanglement assisted estimation of a quantum transformation

Here is a brief preview of the kind of problem we are going to study—problems where statistics and probability are deeply involved. We consider here a concrete problem studied by **(author?)** (16). The problem comes from experimental quantum optics and the theoretical results described here are having impact on planned future experiments.

Suppose you have a "quantum blackbox" which does something to the state of polarization of a single photon. You could feed in various different input states, do various different kinds of measurements on the output, and so reconstruct what the blackbox is doing. Quantum mechanics in fact puts rather precise limitations to the "what can be done to one qubit". It also limits rather precisely how much information can be extracted about its state, a statistical version of the famous uncertainty relations.

Now it turns out that if one feeds into the blackbox one of a pair of maximally entangled photons, then the joint state of the two output photons (one of which has been through the blackbox, the other which may have been nowhere near) contains within it the complete specification of what the blackbox does. So one does not need to feed in all different states, one only needs to feed in photons in a single state. One can see this as an instance of quantum parallelism—because of the entanglement one is actually feeding in all different input states simultaneously.

But now one has to measure the two output photons, not just one, and the question is now, what kind of measurement is best. In particular, can one gain by bringing the output photon and its partner into quantum interaction *before* measurement, or is it enough to measure each one separately in sufficiently many different ways (obviously, looking at the correlations between the results)?

(author?) (16) has discovered, for the important subclass of unitary blackboxes, that the best "joint" measurement of the two outputs is exactly three times as effi-

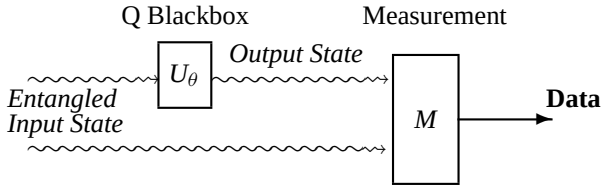

Fig. 1.6. Using an entangled probe state on an unknown quantum operation

cient as the best one can do, when one only measures the outputs separately. We know precisely what is the best measurement to do. It turns up in a number of extremely important quantum algorithms, including teleportation. Unfortunately, the best the quantum opticians can do at the moment is a kind of surrogate measurement which only succeeds once in four times! But they are working on it ...

1.3 States, Measurements, Operations, Instruments

In this section we get down to business by describing the elements of the mathematical model of quantum information. We will give working mathematical definitions of states, measurements, operations, and instruments. In the next chapter we will go into further depth and give alternative, more fundamental definitions. The provisional definitions here, which are concrete but unintuitive, allow a kind of quick-start, and we will already be able to study some simple examples in this section.

The model allows us to convert the boxes and arrows of Figure 1.6 into a formal mathematical model. On the one hand one has quantum systems, which are described or represented by their *states*. On the other hand, quantum systems can be operated on in various ways, transforming them or getting information out of them. The technical term for the most general kind of transformation which is allowed by quantum physics is an *instrument*. An instrument will in general both extract information from the state, and transform the state to a new one. By "information" we mean here: data, which could be observed by a physicist; it will be random, and the mathematical model just tells us what its probability distribution is. One often speaks of *classical* information, to further underline the character of this part of the output of the instrument.

As special cases we have instruments which deliver only one of the two kinds of outputs: quantum and classical. At the one extreme we have an instrument which does not yield any classical information but only transforms the quantum state; this kind of instrument is called an *operation*. At the other end of the spectrum, an instrument which produces data but no output state is called a *measurement*.

We think of operations, measurements and more generally instruments, as boxes which can be plugged into one another, with one-way connecting cables of the two

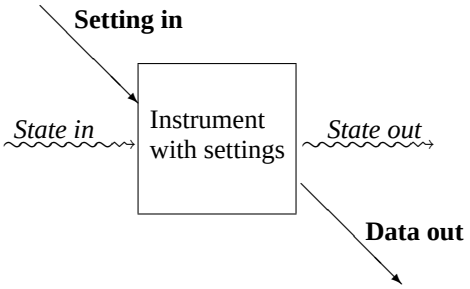

Fig. 1.7. An instrument with settings

different kinds: quantum, and classical. A simple example is provided by Figure 1.6. Note the wavy lines for quantum states and straight lines for classical data. The output quantum systems can in principle be fed into new instruments. Not only this; also, the output data could in principle be manipulated and used to fix a *setting* on another instrument. When we say that an instrument has a setting, we mean that the internal configuration of the instrument can be altered by pushing some buttons or turning knobs on the outside. Strictly speaking, an instrument with settings is a family of instruments, each one corresponding to a different value of the settings. In diagrams, an "instrument with settings" will be a box with both classical and quantum, input and output, channels, see Figure 1.7.

In this way we can build up complex arrangements with many boxes and connections, forming a directed, acyclic graph. "Applying the same operation several times" is represented by repeating the same box several times in the figure. Examples are given in Section 2.7 of Chapter 2. The end result can be extremely complex but it remains some kind of a caricature of what quantum physics is all about. Continuous time has been replaced by the discrete steps through our figure. However, the "true" continuous time picture which we might eventually want to study, can always be thought of as the limiting result of such discrete time diagrams, with more and more instruments each one having a smaller and smaller effect.

First we mathematically define the objects on which measurements, operations and instruments all act: states.

1.3.1 States

The state of a quantum system is fixed by specifying its dimension d together with a $d \times d$ matrix ρ of complex numbers, which must be nonnegative (this notion to be explained shortly) and have trace (sum of the diagonal elements) equal to 1. The matrix ρ is often simply called *the state*. A matrix with these properties is called a *density matrix*. We restrict attention here to the finite dimensional case, where already the most simple nontrivial case, $d = 2$, leads to interesting statistical models. That two-dimensional case, $d = 2$, goes under a variety of names, such as: the qubit; the two-level system; the spin-half system. This model applies to the polarization of a single photon, the spin of a single electron, and to an atom which

can either be in its ground state or its first excited state. It can also be used to model the presence or absence of a particle. Our purpose in subsequent subsections is to explain how the probabilities of measurement results (for instance, a measurement to determine whether the atom is in its ground state or its excited state) can be read off the state.

At the same time as explaining what it means for a matrix to be nonnegative, we introduce some further notation: Dirac's bra and ket notation for vectors, much used by physicists.

Let $|\psi\rangle$ stand for a column vector in $\mathbb{C}^d$. It is called a *ket*. We write $\langle\psi|$ for its Hermitian conjugate, in other words the row vector of complex conjugates of the same numbers; it is called a *bra*. The expression $\langle\psi|\rho|\psi\rangle$ therefore stands for a complex number. When we say that ρ is nonnegative, we mean that for all $|\psi\rangle \in \mathbb{C}^d$, the number $\langle\psi|\rho|\psi\rangle$ is real and nonnegative.

It is a simple exercise to prove that ρ nonnegative implies that ρ is self-adjoint, i.e., $\rho = \rho^*$, where ρ^* is the Hermitian conjugate of ρ (transpose and elementwise complex conjugate). For future reference, we note the following fact about self-adjoint matrices: a d-dimensional self-adjoint matrix has d real eigenvalues, and one can find d corresponding eigenvectors forming an orthornormal basis of $\mathbb{C}^d$. We shall refer to $\mathbb{C}^d$ as the *state-space* of the quantum system, often denoting it by $\mathscr{H}$ (for Hilbert-space).

Returning to the bra-ket notation: at this stage, one might prefer to just write ψ and ψ^* instead of $|\psi\rangle$ and $\langle\psi|$; the number $\langle\psi|\rho|\psi\rangle$ being just $\psi^*\rho\psi$. The brackets can be completely removed; a smattering of stars inserted judiciously instead. However as we will see, the bra-ket notation enables a powerful short-hand, which can be very useful in more complex situations.

Letting $\mathbf{0}$ denote the $d \times d$ zero matrix, we can summarize the defining properties of density matrix as $\rho \geq \mathbf{0}$, $\mathrm{trace}(\rho) = 1$. For free we also get $\rho = \rho^*$. It is not difficult to check the following fact: *ρ is a $d \times d$ density matrix if and only if the d diagonal elements of ρ, expressed with respect to an arbitrary orthonormal basis of $\mathbb{C}^d$, form a probability distribution over* $\{1, \ldots, d\}$. This fact will be connected in Section 1.3.2 to measurements on the quantum system. In fact: choosing an orthornormal basis corresponds to choosing a measurement within a certain basic class of measurements.

Exploiting the bra-ket notation, let us denote by $|1\rangle$, $|2\rangle$, $\ldots$, $|d\rangle$ some orthonormal basis of $\mathbb{C}^d$. Thus, these are d column vectors satisfying the property $\langle i|j\rangle = \delta_{ij}$ (Kronecker's delta) for all i and j from 1 to d. Arrange the row vectors $\langle i|$ as rows of a matrix U. The orthornomality of the basis is expressed by saying $UU^* = \mathbf{1}$, the identity matrix. It follows (when d is finite) that we also have $U^*U = \mathbf{1}$; the matrix U is called *unitary*. Note that $U|j\rangle$ is the column vector containing a 1 at the jth position and 0's elsewhere. So U acting (on U's right) on an arbitrary ket, expresses that vector in the new basis. As we will see later, the matrix ρ may also be thought of as an operator, also acting on a ket on ρ's right. Since $U(\rho|\psi\rangle) = (U\rho U^*)(U|\psi\rangle)$ we see that $U\rho U^*$ is the matrix representation of the operator ρ with respect to the basis defined by the $|i\rangle$'s.

To summarize: ρ is a density matrix if and only if the diagonal elements of

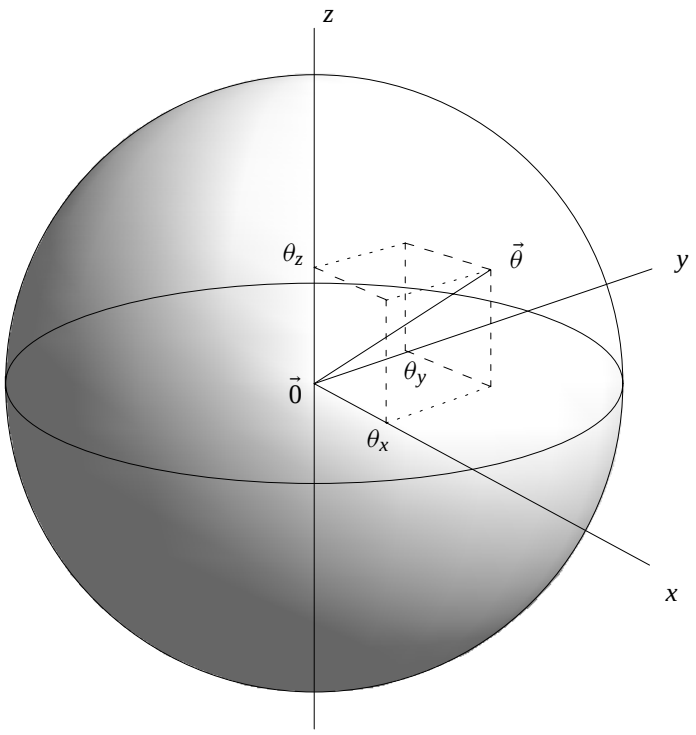

Fig. 1.8. The Bloch sphere: an unknown mixed qubit $\rho(\vec{\theta})$.

$U\rho U^*$ form a probability distribution *for any unitary matrix* U, i.e., with respect to any orthonormal basis of $\mathbb{C}^d$. As we will see in the next subsection (on measurements), each unitary U also corresponds to a certain measurement with possible outcomes $\{1, \ldots, d\}$. The diagonal of $U\rho U^*$ is the probability distribution of the outcome of that measurement.

First we conclude this subsection on states by presenting our most basic example: the case $d = 2$. We will return to this example in each of the subsequent subsections, as well as providing further details in the next chapter.

Example 1.1 (The Qubit: states)**.** Suppose ρ is a 2×2 density matrix. The properties $\rho \geq \mathbf{0}$ and $\text{trace}(\rho) = 1$ imply that ρ_{11} and ρ_{22} are probabilities adding to one, while $\rho_{12} = \overline{\rho}_{21}$ (overline denoting complex conjugate) is a complex number. The matrix being nonnegative also implies that its determinant is nonnegative, thus $\rho_{11}\rho_{22} - |\rho_{12}|^2 \geq 0$ where the absolute value of a complex number c is defined by $|c| = \sqrt{(\Re c)^2 + (\Im c)^2} = \sqrt{c\overline{c}}$. We may write $\rho_{21} = \frac{1}{2}(\theta_1 + i\theta_2)$, $\rho_{12} = \frac{1}{2}(\theta_1 - i\theta_2)$, $\rho_{11} = \frac{1}{2}(1 + \theta_3)$, $\rho_{22} = \frac{1}{2}(1 - \theta_3)$, where θ_1, θ_2, θ_3 are real numbers. The nonnegativity of the determinant turns out to be equivalent to $\theta_1^2 + \theta_2^2 + \theta_3^2 \leq 1$. It is easy to check that these conditions are not only necessary but also sufficient for ρ to be a 2×2 density matrix.

Let $\vec{\theta} \in \mathbb{R}^3$ be a real vector with components θ_i, and such that $|\vec{\theta}| \leq 1$; thus $\vec{\theta}$ is a point in the unit ball in real three-dimensional space. Let $\rho(\vec{\theta})$ be the corre-

Fig. 1.9. A measurement

sponding density matrix. We have hereby defined the quantum statistical model for "a completely unknown state of one qubit" and shown how it can be parametrised by a point $\vec{\theta}$ in the unit ball, which in this context is called *the Bloch sphere*, see Figure 1.8. The three axes are conventionally labelled the x-, y-, and z-axes, and the three components of $\vec{\theta}$ often relabelled as θ_x, θ_y, and θ_z.

It turns out that the effect of a change of basis in $\mathbb{C}^2$ is equivalent to a rotation of the unit ball in $\mathbb{R}^3$. This fact allows us to read off the probability distribution of outcomes of a basic class of measurements from this picture.

Note that the projection of the point $\vec{\theta}$ onto the vertical axis splits the vertical diameter of the ball (of total length 2), into two parts, of lengths $1+\theta_z$ and $1-\theta_z$ respectively. We already noted that the diagonal elements of ρ are exactly $\frac{1}{2}$ times these quantities. In general therefore, a basic measurement on the qubit corresponds to the choice of a direction in $\mathbb{R}^3$, or if you prefer, to the choice of a diameter of the unit ball. This measurement has a binary outcome. *The probabilities of the two outcomes are given by the ratio in which the projection of the state (seen as a point in the ball) onto the measurement (seen as an axis of the ball, or a direction through it) divides the corresponding diameter in two parts.* □

1.3.2 Measurements

A measurement M on a quantum system, see Figure 1.9, is defined by specifying a collection of nonnegative matrices $m(x)$, indexed by potential outcomes x in some sample space $\mathscr{X}$, such that $\sum_x m(x) = \mathbf{1}$. The rule for obtaining the probabilities of the different outcomes is known as the trace rule:

$$p(x|\rho, M) \;=\; \mathrm{trace}(\rho m(x)). \tag{1.1}$$

It is an exercise to the reader, to check that this prescription indeed defines a probability distribution: nonnegative real numbers adding to 1.

If the state is parametrized by $\theta \in \Theta$ then, given the measurement, we obtain a parametric statistical model, $p(x|\theta, M) = \mathrm{trace}(\rho(\theta)m(x))$. A central aim of quantum statistics is to design the measurement M, taking account of constraints on experimental resources, to optimize the amount of statistical information which the experiment will provide about θ.

We describe a special kind of measurements, called *simple measurements*, in detail. Let $|1\rangle$, $|2\rangle$, ..., $|d\rangle$ denote as before some orthonormal basis of $\mathbb{C}^d$. The matrices $|i\rangle\langle i|$, $i = 1, \ldots, d$, are the projection operators onto the d orthogonal one-dimensional subspaces generated by each of the d basis vectors. They add to the identity: $\sum_i |i\rangle\langle i| = \mathbf{1}$. In fact this formula is simply a rewriting of the equation $U^*U = \mathbf{1}$, where U is the matrix with $\langle i|$ as its ith row.

We have hereby defined a measurement M with sample space $\mathscr{X} = \{1, \ldots, d\}$, and with $m(x) = |x\rangle\langle x|$ for each $x \in \mathscr{X}$. Applying the trace rule to compute the probabilities of the measurement outcomes, we find $p(x|\rho, M) = \mathrm{trace}(\rho m(x)) = \langle x|\rho|x\rangle$. This is nothing else than the (x, x) element of the matrix $U\rho U^*$. As we announced before, the diagonal of ρ expressed with respect to any particular orthonormal basis contains the probability distribution of the outcomes of a particular measurement on that state.

Example 1.2 (The Qubit: measurements)**.** In Example 1.1 we analysed the density matrix of a 2-dimensional quantum system. Defining

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.2}$$

the result there could be expressed $\rho(\vec{\theta}) = \frac{1}{2}(\mathbf{1} + \vec{\theta} \cdot \vec{\sigma})$ with $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, $\vec{\theta} = (\theta_x, \theta_y, \theta_z)$ and '$\cdot$' denoting the inner product of two 3-vectors. The same analysis shows that an arbitrary 2×2 self-adjoint matrix must be of the form $a\mathbf{1} + \vec{b} \cdot \vec{\sigma}$ with a a real number and $\vec{b}$ a real 3-vector.

The three self-adjoint matrices σ_x, σ_y and σ_z are called the Pauli matrices. They satisfy the famous commutation relations

$$\begin{aligned} \sigma_x\sigma_y &= i\sigma_z = -\sigma_y\sigma_x, \\ \sigma_y\sigma_z &= i\sigma_x = -\sigma_z\sigma_y, \\ \sigma_z\sigma_x &= i\sigma_y = -\sigma_x\sigma_z, \end{aligned} \tag{1.3}$$

and moreover

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \mathbf{1}. \tag{1.4}$$

They each have trace zero. Since their square is the identity, they must have eigenvalues ± 1. It follows from the properties of the Pauli matrices, that for any 3-vector of unit length $\vec{u}$, $\vec{u} \cdot \vec{\sigma}$ has trace zero and its square is the identity. It therefore also also has eigenvalues ± 1. The eigenvalues of a generic 2-dimensional self-adjoint matrix $a\mathbf{1} + \vec{b} \cdot \vec{\sigma}$ are therefore $a \pm |\vec{b}|$ and the matrix is nonnegative if and only if $a \geq |\vec{b}|$.

These preliminaries enable us to add simple measurements to the Bloch sphere picture of the qubit, Figure 1.8. For $\vartheta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$ let

$$\vec{u} = \vec{u}(\vartheta, \varphi) = (\sin\vartheta\cos\varphi, \sin\vartheta\sin\varphi, \cos\vartheta) \tag{1.5}$$

denote the unit 3-vector whose polar coordinates are (ϑ, φ), see Figure 1.10. Define

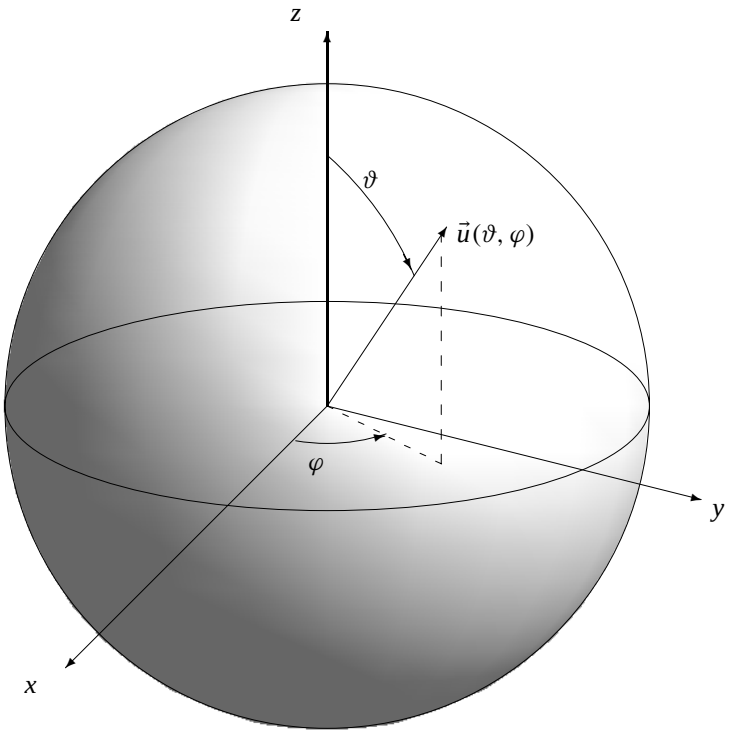

Fig. 1.10. The Bloch Sphere: polar coordinates of pure states.

the ket

$$|\vec{u}\rangle \;=\; |\vartheta, \varphi\rangle \;=\; \begin{pmatrix} e^{-\varphi/2}\cos(\vartheta/2) \\ e^{+\varphi/2}\sin(\vartheta/2) \end{pmatrix}. \tag{1.6}$$

By simple trigonometry one finds the important relation

$$|\vartheta, \varphi\rangle\langle\vartheta, \varphi| \;=\; \frac{1}{2}(\mathbf{1} + \vec{u}(\vartheta, \varphi) \cdot \vec{\sigma}). \tag{1.7}$$

We recognise here a special case of the formula for a 2-dimensional density matrix. For a given unit 3-vector $\vec{u}$, the two states $\rho(\pm\vec{u})$ are located at opposite ends of a diameter of the Bloch ball in the direction $\vec{u}$. Such states are called *pure* states.The eigenvalues of both of these density matrices are 1 and 0, the eigenvectors of both are $|\vec{u}\rangle$ and $|-\vec{u}\rangle$. The density matrices are simultaneously the projector operators onto the *orthogonal* one-dimensional subspaces spanned by the kets $|\vec{u}\rangle = |\vartheta, \varphi\rangle$ and $|-\vec{u}\rangle = |\pi - \vartheta, \pi + \varphi\rangle$.

It follows that a simple measurement M with outcomes, say, $+1$ and -1, has two measurement components or elements of the form $m(\pm 1) = \frac{1}{2}(\mathbf{1} \pm \vec{u} \cdot \vec{\sigma})$. These matrices are of course nonnegative and add to the identity; moreover, being projection matrices, they are idempotent (equal to their squares). The measurement

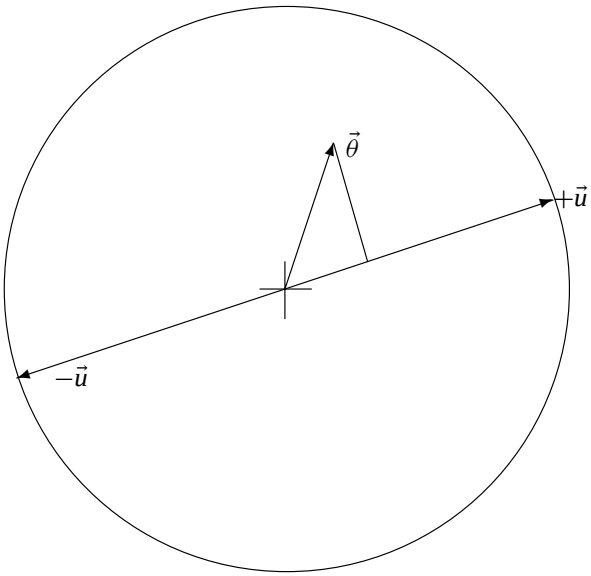

Fig. 1.11. Geometric picture of measurement probabilities. Simple measurement in the direction $\vec{u}$ of a qubit in state $\vec{\theta}$. Probabilities of outcomes ± 1 are half the lengths of the two segments of the diameter.

M is completely determined by choosing a direction $\vec{u}$ in the Bloch sphere. By a simple calculation using (1.3) and (1.4) together with the fact that a Pauli matrix is traceless, we find from the trace rule, for the state $\rho(\vec{\theta}) = \frac{1}{2}(\mathbf{1}+\vec{\theta}\cdot\vec{\sigma})$, the probabilities $p(\pm 1|\vec{\theta}, M) = \frac{1}{2}(1 \pm \vec{u}\cdot\vec{\theta})$. These probabilities can be found geometrically as follows: in the Bloch sphere, project the state $\vec{\theta}$ onto the diameter in the direction $\vec{u}$. This splits the diameter into two parts. The odds on the outcomes ± 1 stand in the ratio of the lengths $(1 \pm \vec{u}\cdot\vec{\theta})$ of the two parts; see Figure 1.11. □

1.3.3 Operations

An operation R on a quantum system, see Figure 1.12, is defined by specifying matrices r_i, satisfying $\sum_i r_i^* r_i = \mathbf{1}$. The result of applying R to the input state ρ is the output state $R(\rho) = \sum_i r_i \rho r_i^*$. The reader should verify that this does define a quantum state: nonnegative, trace 1.

A very special case results when the index i takes on a single value only. Discarding the now superfluous index, and renaming r to U, we have that the input state ρ is transformed into the output state $U\rho U^*$ where $U^*U = \mathbf{1}$. In the finite dimensional case it follows that we also have $UU^* = \mathbf{1}$; the matrix U is unitary, and the transformation R is called unitary too.

Example 1.3 (The Qubit: operations)**.** It can be shown that the generic quantum operation on a single qubit can be represented in the Bloch sphere picture as an affine map of the unit ball, into the unit ball; but with reflection excluded. It can

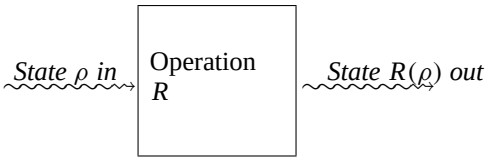

Fig. 1.12. An operation

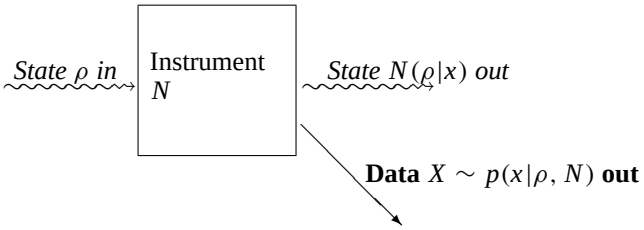

Fig. 1.13. An instrument

therefore be composed of a rotation, shrinking in three orthogonal directions, and finally a shift, in such a way that the resulting ellipsoid remains inside the unit ball.

Of special interest are the rotations (and reflections). These turn out to correspond to the unitary transformations. We postpone further discussion to the next chapter, in which we will introduce the functional calculus of operators, which will enable us to investigate the relations between self-adjoint operators and unitary operators in more depth. □

1.3.4 Instruments

An instrument N on a quantum system, see Figure 1.13, is defined by specifying a collection of matrices $n_i(x)$ where $x \in \mathscr{X}$ is a possible outcome (data) of applying the instrument to the quantum system. The matrix elements of the instrument have to satisfy $\sum_i \sum_x n_i^*(x) n_i(x) = \mathbf{1}$.

For a given instrument N define a measurement M with the same outcome space by $m(x) = \sum_i n_i^*(x) n_i(x)$. The instrument acts on a quantum system in state ρ as follows: with probability $p(x|\rho, N) = \mathrm{trace}(\rho m(x))$ the instrument yields data x; in other words, the instrument applies the measurement M to the quantum system. However, that is not all: the instrument also transforms the input state into an output state. Given the output data x, the output state is $N(\rho|x) = \sum_i n_i(x) \rho n_i^*(x) / p(x|\rho, N)$.

An important special case of an instrument is obtained when the index i only

takes on a single value and when the matrices $n_i(x)$ are the projectors $|x\rangle\langle x|$ onto the one-dimensional spaces spanned by elements of an orthonormal basis $|x\rangle$, $x \in \mathscr{X} = \{1, \ldots, d\}$. Applying the rules, we discover that the instrument yields the outcome x with probability $\langle x|\rho|x\rangle$, while given this outcome, the output state is $|x\rangle\langle x|$. We call such an instrument a *simple instrument*.

Example 1.4 (The Qubit: instruments)**.** By the discussion above, a simple instrument on a qubit is defined by choosing a direction $\vec{u}$ in the unit sphere. We may take the outcome space to be $\mathscr{X} = \{+1, -1\}$. When the instrument is applied to a system in state $\rho(\vec{\theta})$, the outcome state is $\rho(\pm\vec{u})$, depending on whether the outcome ± 1 is observed. The probabilities of these two outcomes are $\frac{1}{2}(1 \pm \vec{\theta} \cdot \vec{u})$. They are proportional to the lengths of the two line segments, formed by projecting $\vec{\theta}$ onto the diameter of the sphere in the direction $\vec{u}$. □

1.4 Outline of this book

In the previous section we have defined a quantum state and various operations which can be applied to a quantum state, yielding sometimes data, sometimes an output state, and sometimes both. We saw that the smallest non-trivial case, a 2-dimensional quantum system, admits of a parametrization using the unit ball in real, three-dimensional space. We see that if we are given many identical copies of a qubit in the same, unknown, state, then one way in which the state can be reconstructed, is by doing simple measurements on one third each of the copies, in the x-, y- and z-directions respectively. Each subset of the measurements gives us statistical information about θ_x, θ_y and θ_z respectively; the three components of the parameter $\vec{\theta}$ of the density matrix, its so-called Bloch vector.

Are there other, better, measurement schemes? And what if we have prior knowledge about the state; for instance, what if we know that its Bloch vector lies on the surface of the Bloch sphere?

The purpose of the book is to answer these and similar questions. To begin with, we must describe important classes of measurement schemes. Instead of measuring each of N copies of a quantum state in the same way, we can consider more complex schemes whereby the outcome of measuring one state is used to control which measurement to apply to the next. So far we only considered one quantum system at a time. According to quantum physics, quantum systems can interact together according to an extension of the rules we have discussed so far. We need to introduce notions of *product system* and of *entanglement* between states. This leads to a notion of joint or collective measurements, going beyond the possibilities which are covered so far. We are going to need an operator calculus for dealing more efficiently with the various different kinds of matrices (operators) which we have met so far.

The definitions of measurement, operation and instrument which we gave above, should be considered as provisional only. A main aim of the next chapter, Chapter 2, is to present alternative and more fundamental definitions. This also requires consideration of product systems and entanglement.

Chapter 3 will extend the modelling from finite dimensional quantum systems, and discrete outcome spaces, to infinite dimensional spaces and to arbitrary outcome spaces. Regarding outcomes of measurements, this corresponds to moving from discrete probability theory to general (measure-theoretic based) probability. The extension is largely a question of notation. On the quantum side, going to infinite dimensional spaces involves many technical complications and subtelties. At first reading, this chapter could be skipped, especially the material on infinite-dimensional states.

In Chapter 4 we will study quantum statistical models in depth, mainly for finite-dimensional states, but allowing arbitrary measurements. We will introduce quantum exponential families and quantum transformation models. A central tool will be the quantum Cramér-Rao inequality together with notions of quantum score and quantum information. We will obtain "large N" answers to some of our main questions: what is the best class of measurements to use? What difference does it make if we have prior knowledge about the state of the quantum system?

In Chapter 6 we will study some quantum estimation problems with an infinite dimensional state space. A basic example here is called "quantum tomography". It results in inverse statistical problems related to classical tomography, where the aim is to reconstruct an arbitrary function of two variables given information about all one-dimensional projections of the function. Methodology is needed from curve- and density-estimation in classical statistics. A main question concerns what rate of convergence is possible, and whether prior knowledge of the parameter (smoothness) allows better convergence rates.

In Chapter 7 we will study continuous time observation of a quantum system, leading to stochastic process models both of diffusion type and of counting type, for the observed outcome process.

In Chapter 8 we study various paradoxical and sometimes problematic features of quantum physics. Could there be a classical physical explanation "behind the scenes" which explains the randomness of quantum measurement outcomes, simply through their dependence on "hidden variables" of a classical nature? If "measurement" is a physical process, why do we have to make some kind of divide between a classical and a quantum level of description? Surely a measurement apparatus is also just a quantum system, and measurement is a quantum process. But then, what is real? We will study experiments which have been made to probe the most thought-provoking consequences of quantum physics, and consider questions of optimal statistical design, and how to take account of various imperfections ("loopholes") in the present-day implementation of these experiments.

The final Chapter 9 will collect together some material on various advanced, or less statistical topics, from quantum physics and quantum information theory.

1.5 Problems, extensions, and bibliographic notes

```
PROBLEMS:
show that the trace rule does yield probabilities
show that nonnegative implies self-adjoint
```

do the spin-half computations
every measurement is part of an instrument
coarsening of instruments
composition of instruments

REFERENCES to mathematical basics (appendix?):
eigenvalues and eigenvectors of self-adjoint matrices
circularity of trace

REFERENCES to literature for basic definitions.
History?
Connections to usual quantum physics:
von Neumann measurement and projection postulate
Schrödinger evolution
Born’s law
Stern-Gerlach

2

Discrete Quantum Probability

In this chapter we study our modelling framework of quantum states, and various kinds of operations which can be applied to them, in depth. To begin with we refine our notion of quantum state by distinguishing between pure and mixed states. The fact that a mixed state truly may be thought of as the physical result of a classical, probabilistic mixing, has far-reaching consequences for the properties of measurements, operations and instruments on quantum states. A second and equally far-reaching contribution of this chapter, is the notion of joint or product or composite quantum systems, leading to the notion of product states and to entanglement. Again, this physical notion has consequences for the properties of measurements, operations and instruments. We show how our preliminary definitions of these notions, which were concrete but unintuitive, can now be replaced by intuitive definitions based on the physical implications of probabilistic mixing and of the possible formation (and dissolution) of product systems. The first definitions we gave, are transformed into the conclusions of representation theorems, giving an explicit characterization of everything that is possible under quantum physics, as characterized by the behaviour of physical systems under mixing and entanglement.

We are now also able to describe a rich hierarchy of classes of measurements on N copies of a quantum system. A main aim of quantum statistics, and in particular of asymptotic theory, is to study this hierarchy. How much do we lose when we restrict ourselves to more basic but more easy to implement measurements?

The reason for the title of this chaper, discrete quantum probability, is linked to the final topic of the chapter. We will study pure states, simple measurements, and unitary operations in more depth, connecting to the quantum physical notions of wave function, observable, von Neumann collapse of the wave function, Born's law, and Schrödinger evolution. We will show how measurements, operations and instruments in general can be thought of as being built of these basic ingredients, acting on an enlarged state space, formed by taking the product of the system of interest with an auxilliary or ancillary system, which might be thought of as representing the measurement apparatus itself, or the macroscopic environment of the quantum system of interest. We show how the trace rule and a functional calculus of observables (self-adjoint operators) leads to a probability-like calculus of

observables and states, generalizing the classical probability calculus of random variables and probabity measures.

Finally, we will give further examples concerning the qubit, as we go along.

2.1 Pure states and mixed states

Recall that the state of a quantum system is represented by a *density matrix*: a $d \times d$ nonnegative matrix of trace 1. It follows that a convex combination of density matrices is again a density matrix. This corresponds physically to classical probabilistic mixing. Suppose with probability p_1 we prepare a quantum system in state ρ_1, and with probability p_2 in state ρ_2, then the resulting system is in state $\rho = p_1\rho_1 + p_2\rho_2$.

Of special interest are the states which are extreme with respect to mixing, i.e., which cannot themselves be represented (in a nontrivial way) as mixtures of other states. Such states are called *pure states*.

A density matrix ρ has eigenvalues, which form a probability distribution over $\{1, \ldots, d\}$, and eigenvectors which form an orthornormal basis of $\mathbb{C}^d$. Denote the eigenvalues by $p_1, \ldots, p_d$ and the eigenvectors by $|1\rangle, \ldots, |d\rangle$. It follows that we can write $\rho = \sum_i p_i |i\rangle\langle i|$. Each of the matrices $|i\rangle\langle i|$ is itself a density matrix (nonnegative, trace 1). It is also equal to the projection operator, which projects onto the one-dimensional subspace spanned by $|i\rangle$.

We see from these considerations that a density matrix is the density matrix of a pure state if and only if its spectrum (its eigenvalues) is equal to $(1, 0, \ldots, 0)$, and if and only if it is idempotent, $\rho^2 = \rho$. All other density matrices are proper mixtures, and the eigenvalue-eigenvector decomposition represents just one way in which they can be written as a mixture of different states.

A pure state has density matrix of the form $|\psi\rangle\langle\psi|$ for some unit vector $|\psi\rangle$, which is then called the *state vector*. Actually, one can multiply the vector by a complex number $e^{i\phi}$ of modulus 1, without changing the density matrix, so the state vector of a pure state is only unique up to an arbitrary *phase factor*. Still, we shall often name a pure state by calling it "the state $|\psi\rangle$". Pure states are often also called vector states.

Example 2.1 (The Qubit: pure states, mixed states)**.** Figures 1.8 and 1.10 gave two pictures of the states of a two-dimensional quantum system. Recall that any 2×2 density matrix can be written as $\rho = \frac{1}{2}(\mathbf{1} + \vec{\theta} \cdot \vec{\sigma})$ where $\vec{\theta}$ is a real 3-vector of length less than or equal to 1, and $\vec{\sigma}$ is the vector of the three Pauli matrices defined in (1.2). It follows that the pure states are those states whose Bloch vector $\vec{\theta}$ has length 1. Probabilistic mixing of states corresponds in this picture with forming the centre of gravity of some mass distribution over the ball. Any mixed state can be represented as a mixture of other states, and in particular as a mixture of pure states, in a multitude of different ways. In particular, the state $\rho = \frac{1}{2}(\mathbf{1} + \vec{\theta} \cdot \vec{\sigma})$ is a mixture of the two pure states $\frac{1}{2}(\mathbf{1} \pm \vec{\theta} \cdot \vec{\sigma}/\|\vec{\theta}\|)$ according to the probabilities $\frac{1}{2}(1 \pm \|\vec{\theta}\|)$.

The pure states can be conveniently parametrized by their polar coordinates, see Figure 1.10. □

2.2 Measurements as positive, normalized linear maps

If a state $\rho = p_1\rho_1 + p_2\rho_2$ can be thought of a probabilistic mixture of states ρ_1 and ρ_2, then it must be the case that when we measure the state with some measurement device M, the results will also be a mixture of the results of measuring ρ_1 with M, and of measuring ρ_2 with M, according to the same probabilities p_1 and p_2. Fortunately, this does follow from our provisional definition of measurements in the Section 1.3.2: the trace rule $p(x|\rho, M) = \text{trace}(\rho m(x))$ *does* guarantee that the result of measuring a mixture of states is the same mixture of probability distributions of measurement outcomes.

In fact this property can be used as a alternative definition of *measurement*. In other words, the trace rule is the inevitable consequence of the physical (probabiliistic) interpretation of mixing. A measurement M is a mapping from states ρ to probability distributions of outcomes $(p(x|\rho, M) : x \in \mathscr{X})$ in some outcome space $\mathscr{X}$. The mapping is linear with respect to convex combinations. We can extend it, in a unique way, to a mapping from all self-adjoint matrices to signed measures on the outcome space $\mathscr{X}$, by insisting on linearity. As such, it is *positive*, in the sense that it maps a nonnegative self-adjoint matrix to a nonnegative measure, and *normalized*, in the sense that it maps a matrix of trace 1 to a measure assigning mass 1 to the whole outcome space $\mathscr{X}$. It is not difficult to prove the converse (and this is left as an exercise to the reader): every positive, normalized linear map from states to discrete measures on $\mathscr{X}$ can be represented through a collection of non-negative matrices $m(x)$ adding to the identity; the mapping becomes $\rho \mapsto (\text{trace}(\rho m(x)) : x \in \mathscr{X})$.

2.3 Operations as positive, normalized linear maps?

Again, by the very interpretation of mixing, the result of applying an operation to a mixture of states should be the same mixture of outcomes of applying the operation to each input state separately. It is clear that our provisional definition of operations from Section 1.3.3 does respect linearity, just as was the case for measurements. A natural conjecture would be that we could define quantum operations abstractly, as being those operations which (when extended, by linearity, to arbitrary self-adjoint matrices) are linear, positive (map nonnegative matrices to nonnegative matrices) and *trace preserving* (or normalized): map matrices of trace 1 to matrices of trace 1. However, this conjecture is *false*. A simple counterexample (see exercises) is the "operation" of taking the *transpose* of a density matrix—which does obviously possess all the properties we just listed.

In fact, "transpose" turns out to be an operation which is disallowed by the laws of quantum physics. In order to explain why, we need to introduce product systems and the notion of entanglement.

2.4 Product systems and entanglement

Often we need to consider a quantum system which is composed of a number of subsystems. This could correspond to different particles, different locations, or different properties of the same particle. The mathematical model for the composition of a joint system from a number of subsystems is through the *tensor product*.

Let $\mathscr{H} = \mathbb{C}^d$ and $\mathscr{K} = \mathbb{C}^{d'}$ be the state spaces of the two composing parts of one joint quantum system. The state space for the joint system is the tensor product space $\mathscr{H} \otimes \mathscr{K}$. In particular, if we bring together two quantum systems in states $\rho_{\mathscr{H}}$ and $\rho_{\mathscr{K}}$ together and form one combined system from them, then the joint state of the two components is $\rho_{\mathscr{H}} \otimes \rho_{\mathscr{K}}$: this is a $(d.d') \times (d.d')$ matrix, whose rows and columns are each indexed by a pair of row indices or a pair of column indices from the two components.

A measurement, operation, or instrument defined on just one component of this composite system can be extended in the natural way to the joint system. For instance, a measurement M with matrix elements $m(x)$ on the system $\mathscr{H}$ can be extended to $\mathscr{H} \otimes \mathscr{K}$ by defining matrix elements $m(x) \otimes \mathbf{1}$ on the product space (nonnegative, add to the identity $\mathbf{1} \otimes \mathbf{1}$). An extended measurement, operation of instrument acts on a system in a *product state*, i.e., of the form $\rho_{\mathscr{H}} \otimes \rho_{\mathscr{K}}$, just as it should, namely, by ignoring the second component completely.

However the whole point of considering joint or composite systems is that many more states are possible than just product states, and many more measurements, operations and instruments are possible, than separate operations on the separate components.

A joint state ρ in $\mathscr{H} \otimes \mathscr{K}$ is called *entangled* if it cannot be written as a mixture of product states. In particular, an entangled pure state in $\mathscr{H} \otimes \mathscr{K}$ has state vector which is *not* the tensor product of separate state vectors in $\mathscr{H}$ and $\mathscr{K}$.

Example 2.2 (Entangled states of two qubits)**.** Let us denote by $|0\rangle$ and $|1\rangle$ the standard basis of $\mathbb{C}^2$: thus

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{2.1}$$

They are located at North and South pole respectively of the Bloch sphere. The notation comes from quantum computing; we think of the two states as representing the two possible binary states, 0 and 1, of a bit in an internal memory register of a computer. To be more precise: we can input such states, by preparation of a quantum system before computation starts, and we will read out such states, by measurement, at the end of the computation, but between reading in the data, and reading out the results, several such systems will have been brought, by joint quantum operations, into states which are *not* products of the "classical" states $|0\rangle$ and $|1\rangle$.

In particular we can consider joint states of two qubits, having state space $\mathbb{C}^2 \otimes \mathbb{C}^2 = \mathbb{C}^4$. A natural orthonormal basis of the joint space is found by taking all tensor products of elements of orthonormal bases of the two components. In this case, it consists of the four vectors $|0\rangle \otimes |0\rangle$, $|0\rangle \otimes |1\rangle$, $|1\rangle \otimes |0\rangle$, and $|1\rangle \otimes |1\rangle$.

We denote these basis vectors by $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. This basis is called the *computational basis* in quantum computing.

The vector $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle$ has length 1, and cannot be written as a tensor product of kets belonging to the two separate subsystems. The corresponding state is called the *singlet state* and is probably the most famous entangled state in quantum physics. In fact the four states $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle$, $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle$, $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle$ and $\frac{1}{\sqrt{2}}(|00\rangle - |11\rangle$ form an orthonormal basis of *maximally entangled states* of $\mathbb{C}^2 \otimes \mathbb{C}^2$ (this terminology will be explained later), called the *Bell basis*.

An example is provided in Section 1.2: **(author?)** (16) has shown that the optimal way to measure an unknown unitary operation on one qubit in the scheme of Figure 1.6 is by using a maximally entangled pair of probe qubits, in the singlet state, and by measuring the two output qubits in the Bell basis. □

Suppose a composite quantum system is in a joint state ρ, but we are only interested in measurements, operations, instruments on one of the two components. It turns out that the parts of the joint system behave on their own, as quantum systems in their own right. In other words we can compute for each subsystem a "marginal state", and applying the measurement $M \otimes \mathbf{1}$ on the joint system in a given joint state, produces the same results as applying M to the first subsystem in its own marginal state. The mathematical representation of "discarding a subsystem" is through the operation of *partial trace*. If ρ is a density matrix of a composite system, then we define the marginal state of the first component, $\rho_{\mathscr{H}}$, to be the partial trace, over the second component, of the joint state:

$$\rho_{\mathscr{H}} \;=\; \mathrm{trace}_{\mathscr{K}}(\rho) \tag{2.2}$$

where

$$(\rho_{\mathscr{H}})_{ij} \;=\; (\mathrm{trace}_{\mathscr{K}}(\rho))_{ij} \;=\; \sum_k \rho_{ik,jk}. \tag{2.3}$$

Consider a composite system in a pure state $|\psi\rangle \in \mathscr{H} \otimes \mathscr{K}$. It is a theorem that one can choose orthonormal bases of the two subsystems, let us denote them by $|i\rangle_{\mathscr{H}}$ and $|j\rangle_{\mathscr{K}}$ respectively, such that

$$|\psi\rangle \;=\; \sum_i a_i |ii\rangle \;=\; \sum_i a_i |i\rangle_{\mathscr{H}} \otimes |j\rangle_{\mathscr{K}} \tag{2.4}$$

where the numbers a_i are real and nonnegative and ordered from large to small, and $\sum_i a_i^2 = 1$. This representation is called the *Schmidt decomposition* of a pure state in the composite system. Note that the two bases depend on the given state $|\psi\rangle$. The reduced state of the first subsystem then turns out to be the mixed state $\sum_i a_i^2 |i\rangle_{\mathscr{H}}\langle i|_{\mathscr{H}}$. As an example, the "Schmidt coefficients" of the singlet state $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle$ are $(1/\sqrt{2}, 1/\sqrt{2})$, as are also those of any element of the Bell basis, defined in Example 2.2. It follows that the reduced state, of either component, of any of these four Bell states, is the "completely mixed state" $\frac{1}{2}\mathbb{1}$ on $\mathbb{C}^2$, i.e., the state with Bloch vector $\vec{0}$, at the centre of the Bloch sphere.

One says that a pure, entangled state, in a product space where each component has the same dimension d, is maximally entangled, if its Bloch coefficients are

all equal to $1/\sqrt{d}$. Equivalently, a vector state in a product space is maximally entangled if and only if its reduced state in each component is the completely mixed state $\frac{1}{d}\mathbf{1}$.

In fact, any mixed state can always be thought of as the reduced state of a composite system in a pure state. If $\rho_{\mathscr{H}} = \sum_i p_i |i\rangle_{\mathscr{H}}\langle i|_{\mathscr{H}}$ is a mixed state on $\mathscr{H}$, expressed with respect to an orthonormal basis $|i\rangle_{\mathscr{H}}$, then this state is the partial trace of a pure state $\psi = \sum_i \sqrt{p_i}|i\rangle_{\mathscr{H}} \otimes |i\rangle_{\mathscr{K}}$ on some product system. This fact is the starting point of a collection of results, whereby mixtures, measurements, operations and instruments can all be thought of as being merely a reflection of something much more simple going on at the level of a much larger state space. It even leads to a philosophical point of view concerning quantum physics, which is sometimes called the dogma of *The Church of the Larger Hilbert Space*!

2.5 Operations as completely positive, normalized linear maps

Though we apply a quantum operation to one quantum system, that quantum system might be part of a larger system. Several particles may have interacted with one another in the past, yet we only consider one of them.

Consider a quantum operation R as a mapping from quantum states to quantum states. As we argued before, it has to be positive, normalized (trace-preserving) and linear. Given a quantum operation R defined on quantum systems with state-space $\mathscr{H} = \mathbb{C}^d$, we can naturally extend it to joint systems of which $\mathscr{H}$ is a component, by defining R on a product system, to transform the first component as is given, while leaving the second component unchanged, and in particular, leaving the joint system in a product state. Let us denote the extended operation by $R \otimes \mathbf{1}$.

Whatever the dimension of the auxiliary system, the extended operation $R \otimes \mathbf{1}$ must be positive, trace-preserving and linear. Clearly these properties are satisfied by operations of the form

$$R(\rho) = \sum_i r_i \rho r_i^*, \qquad \text{with} \quad \sum_i r_i^* r_i = \mathbf{1}. \tag{2.5}$$

According the *Kraus representation theorem*, the converse is also true.

We say that a mapping R from self-adjoint operators to self-adjoint operators is *completely positive* if $R \otimes \mathbf{1}$ maps positive operators to positive operators, whatever the dimension of the auxiliary system. The point here is that there are many more positive operators, than operators of the form $X \otimes Y$, where X and Y are both positive. An example is given by an entangled pure state density matrix $|\psi\rangle\langle\psi|$, which is a positive operator on the product system, but not a positive combination of products of positive operators. Thus being completely positive is a much stronger property than being positive. Thanks to the Kraus theorem, we could *define* a quantum operation to be a completely positive, trace preserving, linear map. The *Kraus representation* (2.5) is then a convenient corollary.

2.6 Instruments as completely positive, normalized linear maps

Finally, we explain how also quantum instruments can be alternatively defined as completely positive, normalized, linear maps. Consider an instrument N with outcome space $\mathscr{X}$. Given input state ρ it generates data x with probability $p(x|\rho, N)$ and the state is then transformed into $N(\rho|x)$. Now from the collection of *non-normalized states* $\sigma(x|\rho, N) = p(x|\rho, N)N(\rho|x)$ we can recover both the probabilities and the output states by taking the trace, and renormalizing: $p(x|\rho, N) = \text{trace}(\sigma(x|\rho, N))$ and $N(\rho|x) = \sigma(x|\rho, N)/p(x|\rho, N)$.

We can therefore consider the instrument as a mapping from input states ρ to *nonnormalized output states indexed by output data*: the state ρ is mapped to the collection $(\sigma(x|\rho, N) : x \in \mathscr{X})$. By the interpretation of mixed states as probabilistic mixtures, it is not difficult to see, as an application of Bayes' formula, that the mapping from input state ρ to data-indexed nonnormalized output states has to be *linear*. It can therefore be extended to a mapping from self-adjoint operators to data-indexed vectors of self-adoint operators, and as such it has to be *positive*, mapping positive operators to vectors of positive operators, and *normalized* or *trace preserving*: the sum of the traces of the outputs is equal to the trace of the input.

From consideration of the fact that the instrument might operate on one component only of a product system, leaving the other component untouched, we see that an instrument has to be *completely positive*.

Again, the *Kraus representation theorem* states that these properties exactly characterise instruments of the form

$$\sigma(x|\rho, N) = \sum_i n_i(x)\rho n_i^*(x), \qquad \text{with} \quad \sum_{i,x} n_i^*(x)n_i(x) = \mathbf{1}. \tag{2.6}$$

In our definition of instrument and of operation, we silently assumed that the output quantum state has the same dimension as the input state. However, everything we have said about operations and instruments remains true, also when the two dimensions are taken to be arbitrary. This allows us to consider a measurement as a special case of an instrument: take the output quantum state to have dimension $d = 1$. Conversely, a little trick allows us to consider instruments as a special case of operations. We store the outcome x of a measurement in an auxiliary quantum system, having $|x\rangle : x \in \mathscr{X}$ as an orthonormal basis. The output of the instrument is taken to be a product system in the state $\sum_x \sigma(x|\rho, N) \otimes |x\rangle\langle x| = \sum_x p(x|\rho, N)N(\rho|x) \otimes |x\rangle\langle x|$. The second way of writing the state expresses it as a mixture over the outcomes x, of each separate output state together with "the measurement device" being in the pure state $|x\rangle$. Because these state vectors are orthogonal they are perfectly distinguishable and able to represent "classical" data.

The distinction between measurements, operations and instruments is just a matter of convenience. All we have are completely positive, trace-preserving, linear maps between different quantum systems.

2.7 The hierarchy of Joint Measurements

Suppose we are given N identical copies of a quantum system in state $\rho(\theta)$. The word "copy" should be thought of in the following way: some apparatus is used to create a quantum system; and the preparation is repeated N times. How should we measure the N states so as to gain the maximum amount of information about θ?

Clearly there may be experimental limitations on the resources which we can use for this task. As we restrict ourselves to smaller and less complex classes of measurement designs, we can extract less and less information about θ. On the other hand, the experiment presumably becomes easier to carry out in the laboratory.

In this section we describe a hierarchy of measurements on N quantum systems. A major task of quantum statistics is to determine how much gain there is, if any, as we move up the hierarchy to more complex experiments. Is it sufficient to restrict ourselves to some very simple kinds of measurements?

For given, finite N the answer to this question will be difficult. There always will be some gain in going to a larger class of experiments; how much the gain can be, will depend in general on many details of the quantum statistical model, the measurement-class, on which parameters are interest parameters, which are nuisance, or more precisely, if it is available, what is the loss function; the answer will also depend typically on what actually is the true value of the parameter θ. However, just as in classical statistics, one may expect that for large N the picture simplifies, at least, as long as we are only interested in approximate (or asymptotic) answers. This indeed turns out to be the case. In Section 5.2 of Chapter 4 we will see that asymptotically there is a dramatic collapse of the hierarchy.

Here we just introduce the hierarchy of measurements, starting with the largest class possible, a completely general *joint* or *collective* measurement on N copies of a quantum state, Figure 2.1.

The N quantum systems, each in state $\rho(\theta)$ with state space $\mathscr{H} = \mathbb{C}^d$, are thought of as components of one large system in state $\rho(\theta)^N$ with state space $\mathscr{H}^N = \mathbb{C}^{d^N}$. An arbitrary measurement M on this system is described by a collection of $d^N \times d^N$ nonnegative matrices $m(x) : x \in \mathscr{X}$, summing to the identity. In general such measurements can be called *entangled* since they require non-trivial quantum interaction between the subsystems.

Note in Figure 2.1 that we have considered the data-processing as a separate phase in the measurement process. From the mathematical point of view we could just as well absorb the data-processing into the measurement procedure, and consider here only those joint measurements, whose outcomes are guesses of the parameter θ of the quantum statistical model. However, we will see that it often pays to keep the roles of the physicist (measurement) and the statistician (data-processing) separate. In the experimental design phase, we (the statisticians) will advise the physicist on measurement arrangements which will guarantee as much statistical information as possible, typically (for large N) measured in terms of the expected Fisher information in the measurement outcomes. After the experiment is done, the processing of the data can be done according to any reasonable statistical

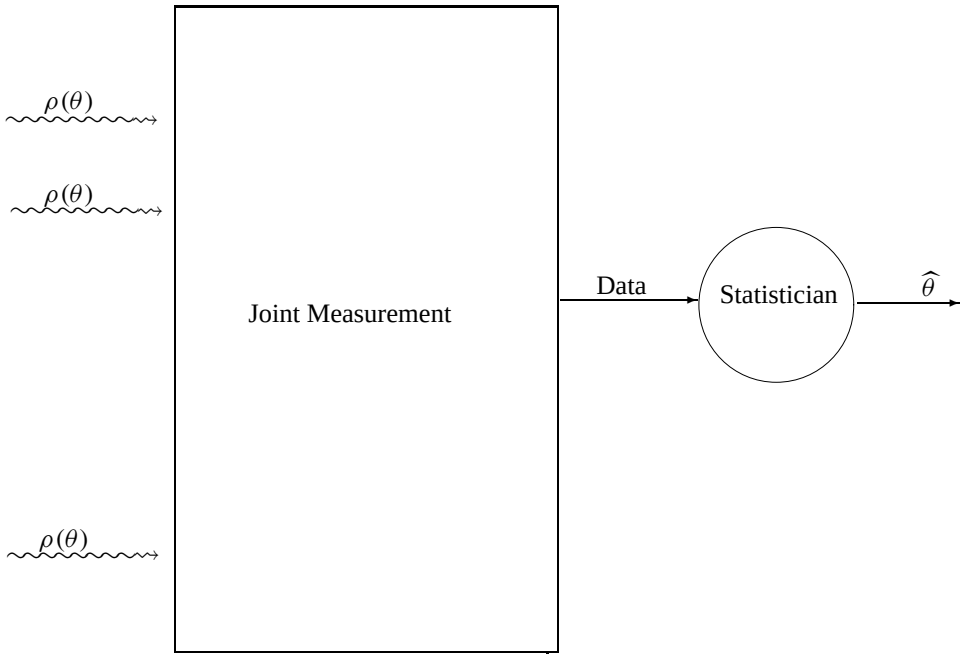

Fig. 2.1. A joint measurement on N copies of a quantum system.

approach (e.g.: maximum likelihood), which will automatically produce estimates with the maximal possible amount of accuracy, given the information in the data.

Now we move to smaller classes of measurements. The first subclass we mention, is a mathematically defined class of measurements called *separable*. A measurement M on the joint system $\mathscr{H}^N$ is separable, if each component $m(x)$ is a sum of products of *positive* matrices, $m(x) = \sum_i \bigotimes_{j=1}^N m_{ij}(x)$, $m_{ij}(x) \geq \mathbb{0}$ for all i, j, and x. So far, no one has found an operational definition of this class of measurements—it is not possible to give a figure, illustrating the meaning of this class! However, as we will see, this class is extremely convenient when carrying out a mathematical analysis of optimality.

A somewhat larger class than separable is the class of PPT, *positive partial transpose*, measurements. This means that each $m(x)$ remains positive, after transposing the elements corresponding to indices belonging to any given subset of subsystems. So far this class has not played a useful role in quantum statistics, though it has turned up in quantum communication. `[Reference: Werner?]`

Strictly smaller than separable, is the class of LOCC measurements, illustrated by Figure 2.2. "LOCC" stands for *local (quantum) operations with classical communication.* The idea is that each subsystem is measured separately, but that the results of the measurements can be used to fix settings on subsequent measurements of the same or other subsystems. In fact we are not talking about separate

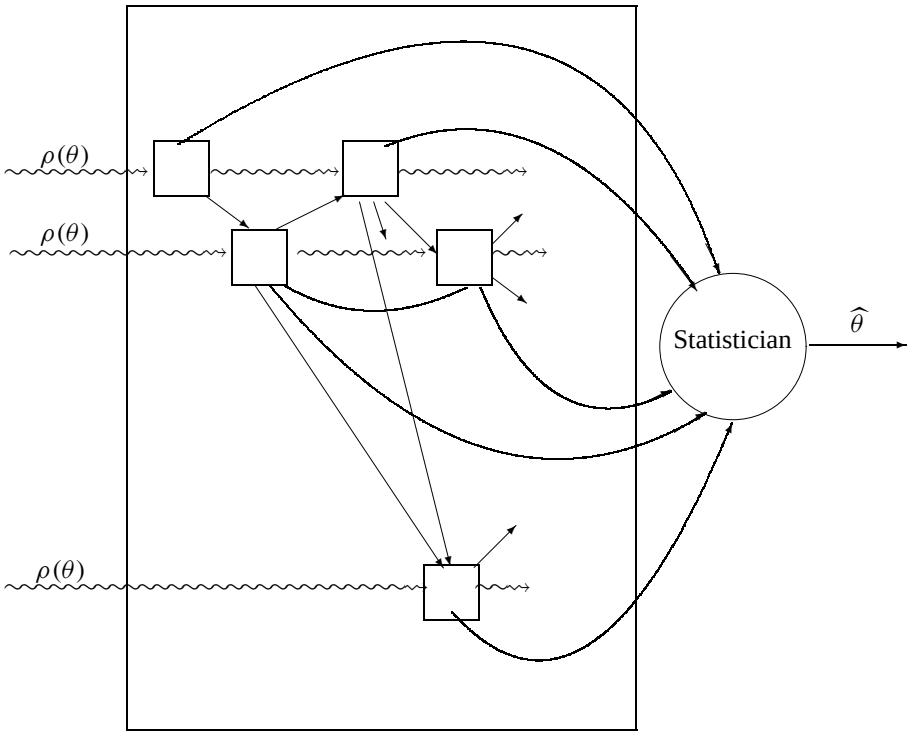

Fig. 2.2. An LOCC, or "multipass adaptive" measurement on N copies of a quantum system.

measurements here, but separate *instruments*, since the individual quantum systems are not destroyed but transformed, and can be measured again. Each subsystem can be measured many times. Without loss of generality, one could agree to measure the subsystems cyclically in the order $1, 2, \ldots, N$, and then back to the beginning. The words "local" and "communication" come from quantum computation theory, where the separate quantum systems might be in physically distinct locations; the only interaction between them comes from communication of classical information, in particular, measurement results, from one location to another. One could also call these measurements *adaptive, multipass*.

It is not difficult to see that all LOCC measurements are separable. The fact that there exist separable measurements which are not LOCC was established by **(author?)** (83), who found the first example of such a measurement. Previously, the two classes were conjectured to be identical. It is still an important open question, to give a concise mathematical characterization of LOCC measurements, and a physical characterization of separable measurements. The strict inclusion remains true, if we would extend the definition of LOCC by taking the closure, in an appropriate sense, of the measurements described so far (thus allowing in some sense also an infinite number of repeat measurements).

We now move to a much smaller class of measurements. Each subsystem is measured just once, the results being used to fix settings on the measurements of

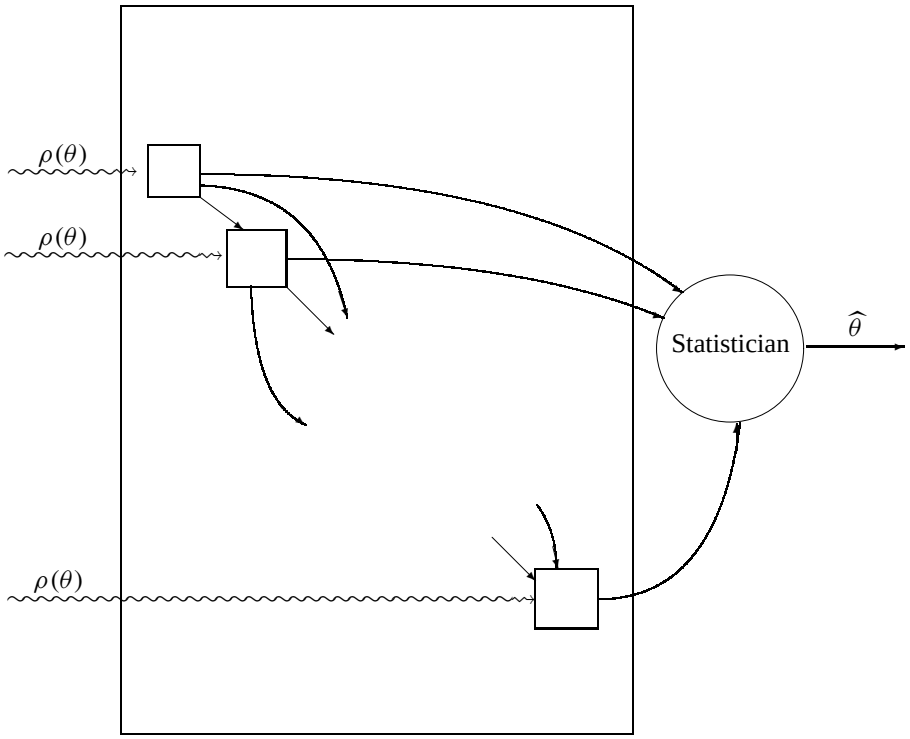

Fig. 2.3. An adaptive (single-pass) measurement on N copies of a quantum system.

the remaining systems, see Figure 2.3. These *adaptive* measurements have been applied in the laboratory by **(author?)** (226), following theory developed by **(author?)** (175). Starting with a uniform prior distribution on an unknown pure state of a qubit (the excitation of an ion in an ion trap), each successive qubit was measured "greedily" using the Bayes optimal measurement, as if the next measurement would be the last one; optimality is defined relative to a certain figure of merit, or loss function, called *fidelity*. For pure states, this is just the absolute value of the inner-product between the true and the guessed state vectors.

Finally, we can consider measuring each quantum system separately (and non-adaptively), but using possibly different measurements on each system, see Figure 2.4. More restrictively, we would use the *same* measurement on each subsystem, Figure 2.5.

2.8 Pure states, Observables

Consider a self-adjoint operator X on a finite-dimensional Hilbert space $\mathscr{H} = \mathbb{C}^d$. It has real eigenvalues, and the whole space can be decomposed into orthogonal eigenspaces (an eigenspace of dimension larger than 1 corresponding to an eigenvalue of the same multiplicity). Let us denote the eigenvalues of the operator by $x \in \mathbb{R}$, and write $[X = x]$ for the eigenspace corresponding to eigenvalue x. For $x \in \mathbb{R}$ which are not eigenvalues, we let $[X = x]$ denote the trivial subspace, of

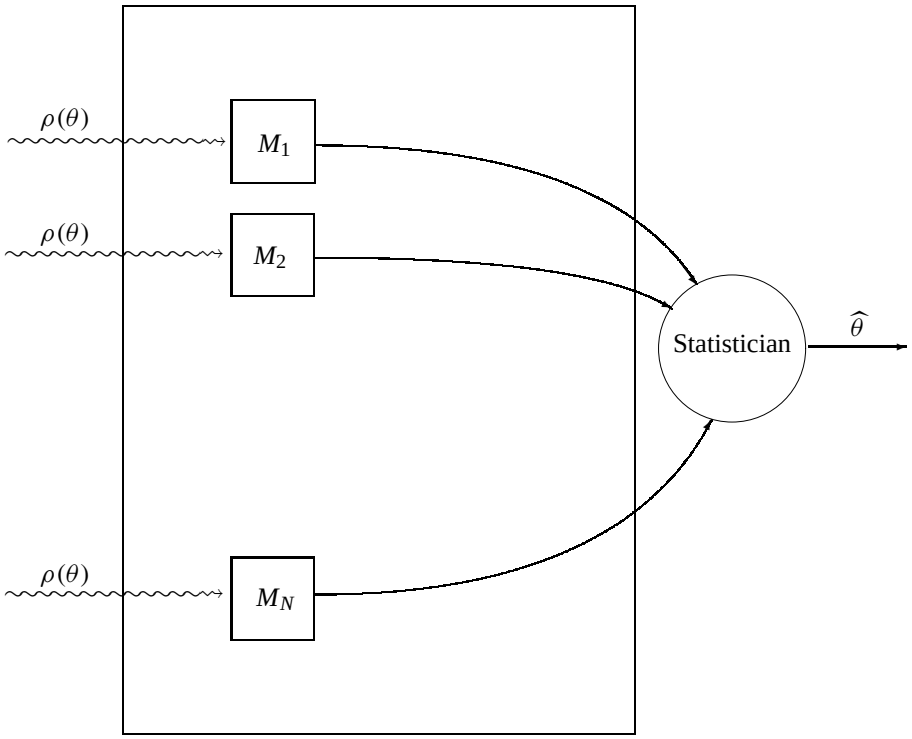

Fig. 2.4. N separate measurements on N copies of a quantum system.

dimension 0, consisting just of the zero vector $\{0\}$. Denote by $\Pi_{[X=x]}$ the operator which projects onto the eigenspace $[X = x]$. We can write $X = \sum_x x\Pi_{[X=x]}$, and we also have $\sum_x \Pi_{[X=x]} = \mathbf{1}$.

For Borel subsets B of the real line, define the subspace $[X \in B] = \bigoplus_{x\in B}[X = x]$. In words, $[X \in B]$ is the subspace of $\mathscr{H}$ generated by the eigenvectors of X with eigenvalue in B. The projection operator onto this subspace is given by $\Pi_{[X\in B]} = \sum_{x\in B} \Pi_{[X=x]}$.

The operator X corresponds to an instrument N_X, by taking the outcome space of the instrument to be the spectrum of X, and by taking the Kraus representation of the instrument to have components $n_i(x) = \Pi_{[X=x]}$, where the index i takes on a single value only. Since a projection operator is self-adjoint and idempotent, the instrument N_X, acting on the state ρ, yields the value x with probability $\mathrm{trace}(\rho\Pi_{[X=x]})$, and the state is then converted into $\Pi_{[X=x]}\rho\Pi_{[X=x]}/\mathrm{trace}(\rho\Pi_{[X=x]})$. In particular, if the state ρ is the pure state $|\psi\rangle\langle\psi|$, then the probability of the outcome x is $\langle\psi|\Pi_{[X=x]}|\psi\rangle = \|\Pi_{[X=x]}|\psi\rangle\|^2$, and the output (unnormalized) state is then $\Pi_{[X=x]}|\psi\rangle\langle\psi|\Pi_{[X=x]}$, which is the pure state with (unnormalized) state vector $\Pi_{[X=x]}|\psi\rangle$.

We call such an instrument a *simple instrument*, slightly enlarging our earlier use of this terminology, where each projection operator projected to a one-dimensional subspace. In this context, X is called an *observable*. "Measuring an observable on a pure state" has a very simple description: the state gets projected into one of

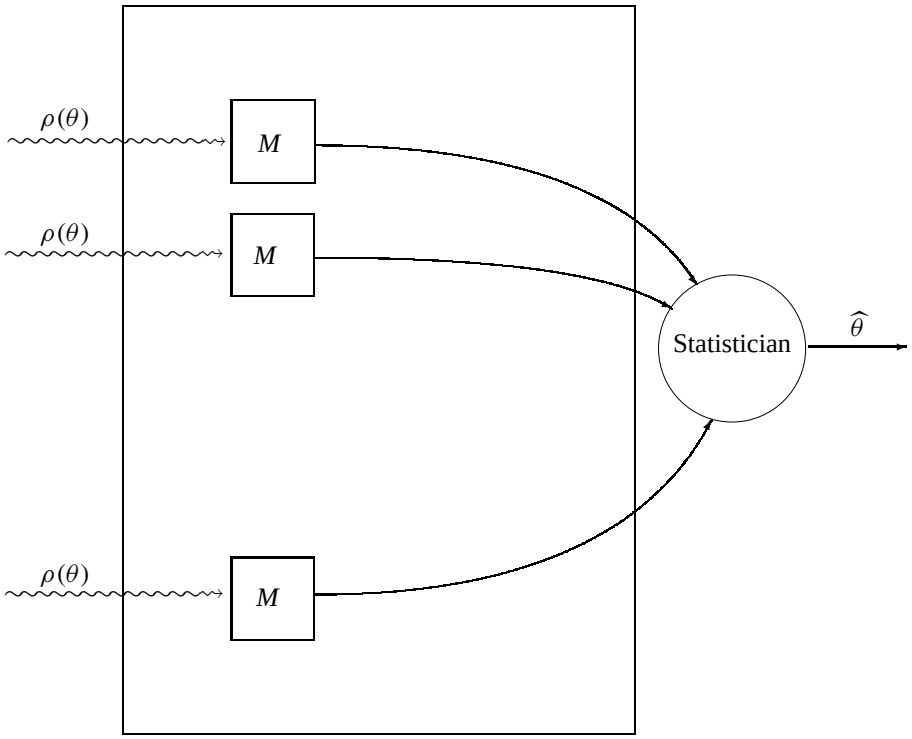

Fig. 2.5. N separate, identical, measurements on N copies of a quantum system.

the eigenspaces of the observable, and the probability of each event is equal to the squared length of the projection. The probabilities are nonnegative and add to 1, by Pythagoras.

Suppose now f is any real function defined on the spectrum of X. We define the operator $f(X) = \sum_x f(x)\Pi_{[X=x]}$. In words: the eigenvalues of X are transformed by the function f, the eigenvectors left unchanged. If the function f is many-to-one, the corresponding eigenspaces are merged. For functions such as "inverse" and "square" the new definitions of X^{-1} and X^2 coincide with the usual definitions. Observe that for the indicator function of a Borel subset of the line, denoted by $\mathbb{1}_B$, we have $\mathbb{1}_B(X) = \Pi_{[X\in B]}$.

The probability of the outcome x was $\mathrm{trace}(\rho\Pi_{[X=x]})$. Let us denote by $\mathrm{meas}(X)$, the random variable with this probability distribution. Multiplying each probability by $f(x)$ and summing over x, we obtain the following version of the trace rule for the expected value of a function of the outcome of measuring X on the state ρ:

$$\mathrm{E}_\rho(f(\mathrm{meas}(X))) \;=\; \mathrm{trace}(\rho f(X)). \tag{2.7}$$

In particular, notice that the mean value of a simple measurement of $f(X)$ is the same as the mean value of f of the outcome of measuring X, even though the posterior state is not necessarily the same, if the function f is many-to-one.

An illustration of this result is the following expression for the variance of the

Fig. 2.6. John von Neumann (from MacTutor Archive, St. Andrews).

outcome of measuring an observable X:

$$\mathrm{var}_\rho(\mathrm{meas}(X)) = \mathrm{trace}(\rho X^2) - (\mathrm{trace}(\rho X))^2. \tag{2.8}$$

We can extend these results to several commuting operators. By a celebrated result of von Neumann (see Figure 2.6), a number of self-adjoint operators all commute with one another if and only if they are all functions of a single self-adjoint operator. In words, commuting operators can be simultaneously diagonalized.

Suppose X and Y commute. We can define a third operator Z such that X and Y are both functions of Z, say $X = f(Z)$, $Y = g(Z)$, and we can choose Z with the smallest possible number of eigenstates for this purpose. The eigenspaces of X and of Y are either eigenspaces of Z, or sumspaces of several eigenspaces of Z.

Now we can consider three different sequential measurement scenarios: first measure X, then measure Y; first measure Y, then measure X; measure Z obtaining an outcome z and report $f(z)$, $g(z)$ as values for X and Y respectively. It is not difficult to check that the joint probability distribution of the pair of outcomes (x, y) is the same under each of these three scenarios, and the posterior state given the outcome (x, y) also coincides in every case. We may talk about a "simultaneous" measurement of X and Y, and we can justly claim that measuring X does not disturb the measurement of Y, and vice versa.

A generalization of the trace rule for expectation values is

$$\mathrm{E}_\rho(\, f(\mathrm{meas}(X, Y))) \;=\; \mathrm{trace}(\rho f(X, Y)). \tag{2.9}$$

where X and Y are commuting observable and f is an arbitrary function of two variables, the operator $f(X, Y)$ defined as $\sum_{x,y} f(x, y)\Pi_{[X=x,Y=y]}$ where $[X = x, Y = y] = [X = x] \cap [Y = y]$. An important corollary is that the probability distribution of a measurement of one observable, or function thereof, is not altered by measuring it jointly together with any number of compatible observables.

Commuting observables are called, in physics, *compatible* observables. All functions $f(X)$ of a given observable are compatible with X, and as we said before, compatible observables can be thought of as both functions of a third observable. An important example of compatible observables is supplied by observables on different subsystems of a composite system. Consider observables on a product space $\mathscr{H} \otimes \mathscr{K}$. If X is an observable for $\mathscr{H}$ and Y an observable for $\mathscr{K}$, then we can define the observables in a natural way on the product space, replacing X by

$X \otimes \mathbf{1}$, and Y by $\mathbf{1} \otimes Y$. The eigenvalues remain unchanged, while the eigenspaces are replaced by the products of the original eigenspaces with the second space in its entireity.

We see that $X \otimes \mathbf{1}$ and $\mathbf{1} \otimes Y$ commute with one another, and therefore we can measure them jointly or in either order, getting the same (distributional) results in each case.

As an example, consider the entangled pure state $(|00\rangle + |11\rangle)/\sqrt{2}$ of two qubits. Suppose we measure the observable σ_z on the first qubit. With probability $1/2$ we observe the value ± 1 and the joint state of the two qubits collapses to the state $|00\rangle$ or to the state $|11\rangle$, according to the outcome. Now measure the second qubit in the same way and with certainty we will observe the same value, ± 1, without any further change to the state. Doing the measurements in the reverse order would result in the same final results. In particular, we notice that the marginal probability distribution of any measurement on the second component, is not altered by doing any measurement whatsoever (or none at all) on the first component. Another way to express this is through the notion of reduced state: the reduced state of the second component of a composite system is not altered by previously subjecting the first component to any quantum instrument. Without this property, the formalism we have developed would be useless as a physical theory. The different components of a composite quantum system are often taken to correspond to physically spatially separated physical systems. The formalism does not allow *action at a distance*: we cannot tell at one location, which operation is being done to another part of the system at another, distant, location.

The projection of the state to an eigenspace of the observable is called the von Neumann-Lüders projection postulate. Though measurement results are not sensitive to operations done at a distance, it does appear that the state of the system as a whole does react globally to operations on subsystems. This raises the philosophical question, whether the state of a quantum system, and in particular the state vector of a system in a pure state, has physical reality. We further discuss these questions in Chapter 8.

A pure state vector is often called in physics a *wave function*; when the Hilbert space becomes infinite dimensional it can indeed be identified with a function. In the finite dimensional space we can consider the state vector as a function from the index set $\{1, \ldots, d\}$ to the complex numbers. When we measure the observable whose eigenvectors are the natural basis vectors of $\mathbb{C}^d$, we obtain the outcomes $i = 1, \ldots, d$ with probabilities $|\psi_i|^2$, the squares of the absolute values of the components of (in other words, the values of) the wave function. This version of the trace rule goes back to the beginning of quantum mechanics, where it is known as *Born's law* (see Figure 2.7).

2.9 Unitary evolution

We briefly mentioned unitary operations as being a special kind of quantum operations in Section 1.3.3. Here we pay them the attention they deserve, connecting

Fig. 2.7. Max Born (from Max Born Institute, Berlin).

them (just as we did for simple measurements) to the most fundamental (and oldest) part of quantum mechanics.

It is a theorem that any unitary matrix U can be written in the form $U = \exp(iA)$ for some self-adjoint matrix A; conversely, if A is self-adjoint, then $\exp(iA)$ is unitary. We can embed a unitary matrix U into a time continuous family $U(t)$ as follows. Consider the differential equation

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle \;=\; H\,|\psi(t)\rangle, \qquad |\psi(0)\rangle \text{ given.} \tag{2.10}$$

where H is a self-adjoint operator, called the *Hamiltonian*, whose elements have the units of energy, and $\hbar$, having units time time energy is *Planck's constant*. The ket $|\psi(0)\rangle$ is an initial, pure state, of a quantum system. The solution of this equation is

$$|\psi(t)\rangle \;=\; U(t)\,|\psi(0)\rangle, \qquad U(t) \;=\; \exp\Big(\frac{1}{i\hbar}Ht\Big). \tag{2.11}$$

Equation (2.10) is called *Schrödinger's equation* and it describes the way in which a quantum system is supposed to evolve, in continuous time, when isolated from any environment. The Hamiltonian H describes the energy of the system. If H has eigenvalues E_n and eigenstates $|E_n\rangle$, then we can write

$$|\psi(t)\rangle \;=\; \sum_n a_n \exp\Big(\frac{1}{i\hbar}E_n t\Big)\,|E_n\rangle \tag{2.12}$$

where the coefficients a_n can be found by expanding the initial state vector in terms of the energy basis, $a_n = \langle E_n|\psi(0)\rangle$. A system which is initially in an eigenstate of H remains always in that state; in general, systems evolve as *superpositions* of energy eigenstates, where just the phases of the components of the superposition oscillate with rates corresponding to the energy eigenvalues.

In terms of density matrices, we can write Schrödinger's equation as

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t}\rho(t) \;=\; [H,\rho], \qquad \rho(0) \text{ given.} \tag{2.13}$$

where the *commutator* $[X,Y]$ of two operators X and Y is defined by

$$[X,Y] \;=\; XY - YX. \tag{2.14}$$

The solution is of course

$$\rho(t) \;=\; U(t)\rho(0)U^*(t) \tag{2.15}$$

where $U(t)$ is defined in Equation (2.11).

Example 2.3 (The Qubit: unitary transformations)**.** Since every 2×2 self-adjoint matrix is of the form $a\mathbf{1} + \vec{c} \cdot \vec{\sigma}$, an arbitrary 2×2 unitary matrix is of the form $\exp(i(a\mathbf{1} + \vec{c} \cdot \vec{\sigma})) = \exp(ia)\exp(i\gamma\vec{v} \cdot \vec{\sigma})$ where $\vec{v} = \vec{v}(\vartheta, \varphi) = \vec{c}/\|\vec{c}\|$ is a unit 3-vector with polar coordinates (ϑ, φ), and $\gamma = \|\vec{c}\|$ is real. The phase $\exp(ia)$ disappears when we compute the effect of the unitary U on a state ρ: the state is transformed into $U\rho U^*$. Taking ρ to be the pure state $|\vec{u}\rangle\langle\vec{u}|$ one may check that the unitary transformation rotates the Bloch vector $\vec{u}$ of the state by an angle γ about the direction $\vec{v}$. It follows by linearity that the effect of U on any state is to rotate its Bloch vector in the same way.

2.10 Welcome in The Church of the Larger Hilbert Space

Both simple measurement and unitary evolution have elegant descriptions in terms of pure state vectors. This leads to a picture of quantum mechanics in which a state vector, in isolation from the rest of the world, evolves in continuous time, according to a unitary evolution of the Schrödinger type (2.10). The evolution is deterministic. The system may get measured at discrete time points. At these time points, the state vector makes a random jump into one of a collection of subspaces associated with the measurement. Which subspace was chosen, is transmitted as classical information to the outside world. The density matrix is used as a convenient packaging of a probabilistic mixture of pure states: in order to make predictions about future outcomes of future measurements, we only need to store the present density matrix.

These two parts of quantum mechanics, Schrödinger evolution and von Neumann collapse of the wave function (with Born's law describing the probabilities of the outcomes) have lived together in uneasy cohabitation since the beginning of quantum mechanics. Physicists generally feel that Schrödinger's equation describes the real physics, whereas Born's law is some kind of add-on; however, without Born's law, there is no way to draw conclusions about the real world, from the model of quantum mechanics. Moreover, the random jumps during measurement are for real and can be observed in the laboratory too. Still, measuring a quantum system is also a physical process. Measurement apparatus is built of quantum systems. Surely, the process of measurement should be describable just in terms of a Schrödinger evolution applied to a composite system consisting of the quantum system of interest together with measurement apparatus.

These considerations lead to paradoxical questions such as what happens to Schrödinger's cat, which we will survey in Chapter 8. Here we want to briefly describe some mathematical results which go some way to reconciling measurement and evolution, and which also show that quantum instruments, as we have described them in most generality, actually have an alternative description involving

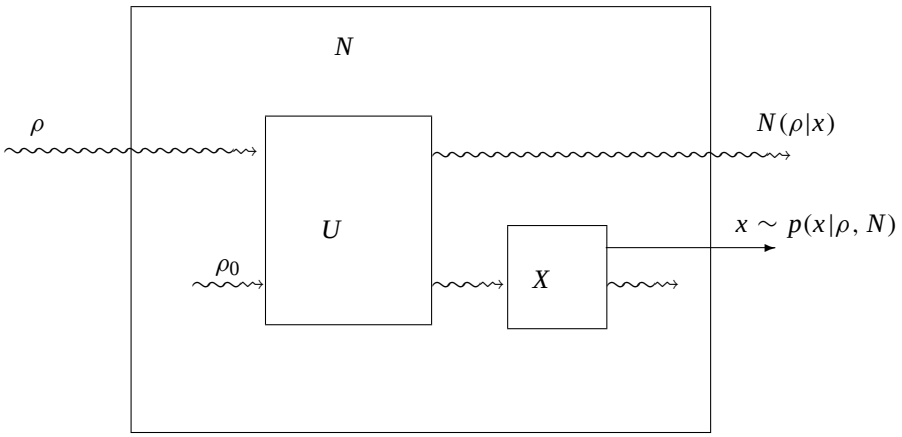

Fig. 2.8. Inside the blackbox: anatomy of an instrument.

only simple measurements, unitary evolution, and ancillary systems (the quantum system of the measurement apparatus?).

Measurements and operations are special cases of instruments, so we only need to deal with quantum instruments here. First of all, let us describe a particular type of quantum instrument N. After that, we will argue that this type is generic: every instrument whatsoever can be represented in this form.

Ingredients of our instrument are an auxiliary quantum system initially in the fixed state ρ_0 on the Hilbert space $\mathscr{K}$. An observable X is defined on $\mathscr{K}$. Further, we need a unitary operation U on the product system $\mathscr{H} \otimes \mathscr{K}$. Our instrument works as follows. The system to be measured, in state ρ on $\mathscr{H}$, is brought into interaction with the ancillary system ρ_0 forming initially the product state $\rho \otimes \rho_0$.

The composite system undergoes the unitary transformation U, converting it into $U(\rho \otimes \rho_0)U^*$. The observable X is measured on the ancilla, resulting in the outcome x and the final, joint, unnormalized state

$$(\mathbf{1} \otimes \Pi_{[X=x]})U(\rho \otimes \rho_0)U^*(\mathbf{1} \otimes \Pi_{[X=x]}) \tag{2.16}$$

of the joint system, with probability equal to the trace of that state,

$$p(x|\rho, N) \;=\; \text{trace}\,(\mathbf{1} \otimes \Pi_{[X=x]})U(\rho \otimes \rho_0)U^*. \tag{2.17}$$

The ancilla is discarded, leaving the system of interest in the (unnormalized) state

$$\sigma(x|\rho, N) \;=\; \text{trace}_{\mathscr{K}}\,(\mathbf{1} \otimes \Pi_{[X=x]})U(\rho \otimes \rho_0)U^*(\mathbf{1} \otimes \Pi_{[X=x]}). \tag{2.18}$$

What we have constructed here *is* a quantum instrument, of course. This can be seen in many ways. We can argue that each of the composing steps is an instrument, and hence the composition of the steps too. We may observe directly that our instrument does define a completely positive, normalized and linear map. Or we

may exhibit the Kraus matrices of our instrument, explicitly in terms of ρ_0, U and X.

According to theorems of Holevo (for measurements) and later Ozawa (for instruments), *every* instrument N can be represented in this way, for some choice of ancilla space $\mathscr{K}$, ancilla state ρ_0, joint unitary U and observable on the ancilla X. The proof relies on a deep result from functional analysis called the Naimark extension theorem (for measurements) and the Stinespring theorem (for the more general case of instruments).

These results make the circle round. One may start with density matrices; hypothesize linearity and complete positivity, and arrive at a certain universe of quantum probability theory, in which there is a special place for pure states, unitary evolution and simple measurement. Alternatively one may start with pure states, unitary evolution, simple measurement, and the possibility to form product systems. One arrives eventually at density matrices and completely positive instruments.

The representation theorem we have just given goes some way to resolve the possible conflict between measurement and evolution. At least, we see that these notions are consistent with one another, and moreover that one can place the boundary between quantum and classical at different levels, basically as a matter of convenience. The philosophical problems remain, and we postpone any discussion of them till Chapter 8.

2.11 Duality: Heisenberg vs. Schrödinger

TO BE DONE:
Heisenberg picture vs. Schrödinger picture.
Dual instruments etc.
States as expectations of observables.

2.12 Problems, extensions, and bibliographic notes

PROBLEMS:
Show that a normalized positive linear map from states to measurement outcome distributions is represented by a POVM.
Show that 'transpose' is represented in the Bloch sphere by reflection through the $y = 0$ plane; show that it does not have a Kraus decomposition.
Prove Naimark, Stinespring.
Check unitary on qubit is rotation.
no-cloning theorem
Bayes' rule to show why initial state $\rightarrow$ unnormalized state indexed by measurement outcome is linear
coarsening of an instrument, revisited

Remember characterization of a quantum operation by its effect on (one part of) an entangled state. This

gives an explicit interpretation of the Kraus matrices.