

1

Quantum measurement theory

1.1 Classical measurement theory

1.1.1 Basic concepts

Although this chapter, and indeed this book, is concerned with quantum measurements, we begin with a discussion of some elementary notions of measurement for classical systems. These are systems that operate at a level where quantum effects are not apparent. The purpose of this discussion is to introduce some ideas, which carry over to quantum theory, concerning states, conditional and non-conditional distributions, and stochastic processes. It will also make the distinct features of quantum measurements plainer.

A classical system can be described by a set of *system variables*, which we will call the system *configuration*. For example, for a system of N interacting particles these could be the N position and momentum vectors of the particles. The possible values of these variables form the *configuration space* \mathbb{S} for the system. In the above dynamical example, the configuration space would be \mathbb{R}^{6N} , where \mathbb{R} is the real line.¹ Alternatively, to take the simplest possible example, there may be a single system variable X that takes just two values, $X = 0$ or $X = 1$, so that the configuration space would be $\{0, 1\}$. Physically, this binary variable could represent a coin on a table, with $X = 0$ and $X = 1$ corresponding to heads and tails, respectively.

We define the *state* of a classical system to be a probability distribution on configuration space. Say (as in the example of the coin) that there is a single system variable $X \in \mathbb{S}$ that is discrete. Then we write the probability that X has the value x as $\Pr[X = x]$. Here, in general, $\Pr[E]$ is the probability of an event E . When no confusion is likely to arise, we write $\Pr[X = x]$ simply as $\wp(x)$. Here we are following the convention of representing variables by upper-case letters and the corresponding arguments in probability distributions by the corresponding lower-case letters. If X is a continuous variable, then we define a probability density $\wp(x)$ by $\wp(x)dx = \Pr[X \in (x, x + dx)]$. In either case, the state of the system is represented by the function $\wp(x)$ for all values of x . When we choose to be more careful, we write this as $\{\wp(x): x \in \mathbb{S}\}$, or as $\{\wp(x): x\}$. We will use these conventions

¹ This space is often called ‘phase space’, with ‘configuration space’ referring only to the space of positions. We will not use ‘configuration space’ with this meaning.

conscientiously for the first two chapters, but in subsequent chapters we will become more relaxed about such issues in order to avoid undue notational complexity.

The system state, as we have defined it, represents an observer's knowledge about the system variables. Unless the probability distribution is non-zero only for a single configuration, we say that it represents a state of *uncertainty* or incomplete knowledge. That is, in this book we adopt the position that probabilities are *subjective*: they represent degrees of certainty rather than objective properties of the world. This point of view may be unfamiliar and lead to uncomfortable ideas. For example, different observers, with different knowledge about a system, would in general assign different states to the same system. This is not a problem for these observers, as long as the different states are *consistent*. This is the case as long as their supports on configuration space are not disjoint (that is, as long as they all assign a non-zero probability to at least one set of values for the system variables). This guarantees that there is at least one state of complete knowledge (that is, one configuration) that all observers agree is a possible state.

We now consider measurement of a classical system. With a perfect measurement of X , the observer would simply find out its value, say x' . The system state would then be a state of complete knowledge about this variable. For discrete variables this is represented by the Kronecker δ -function $\wp(x) = \delta_{x,x'}$, whereas for a continuous variable it is represented by the Dirac δ -function $\wp(x) = \delta(x - x')$. For comparison with the quantum case (in following sections), it is more enlightening to consider imperfect measurements. Suppose that one only has access to the values of the system variables indirectly, through an *apparatus* variable Y . The state of the apparatus is also specified by a probability distribution $\wp(y)$. By some physical process, the apparatus variable becomes statistically dependent on the system variable. That is, the configuration of the apparatus is correlated (perhaps imperfectly) with the configuration of the system. If the apparatus variable is observed, $\{\wp(y): y\}$ is simply the probability distribution of measurement outcomes.

One way of thinking about the system–apparatus correlation is illustrated in Fig. 1.1. The correlation is defined by a functional relationship among the readout variable, Y , the system variable, X , before the measurement, and a random variable, Ξ , which represents extra noise in the measurement outcome. We can specify this by a function

$$Y = G(X, \Xi), \quad (1.1)$$

together with a probability distribution $\wp(\xi)$ for the noise. Here, the noise is assumed to be independent of the system, and is assumed not to affect the system. That is, we restrict our consideration for the moment to *non-disturbing measurements*, for which X after the measurement is the same as X before the measurement.

1.1.2 Example: binary variables

To illustrate the above theory, consider the case of binary variables. As we will see, this is relevant in the quantum setting also. For this case, the state of the system $\wp(x)$ is completely specified by the probability $\wp(x := 0)$, since $\wp(x := 1) = 1 - \wp(x := 0)$. Here

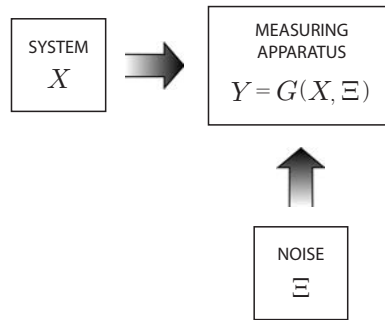


Fig. 1.1 System–apparatus correlation in a typical classical measurement.

we have introduced another abuse of notation, namely that $\wp(x := a)$ means $\wp(x)$ evaluated at $x = a$, where a is any number or variable. In other words, it is another way of writing $\Pr[X = a]$ (for the case of discrete variables). The convenience of this notation will become evident.

We assume that the apparatus and noise are also described by binary variables with values 0 and 1. We take the output variable Y to be the *binary addition* (that is, addition modulo 2) of the system variable X and the noise variable Ξ . In the language of binary logic, this is called the ‘exclusive or’ (XOR) of these two variables, and is written as

$$Y = X \oplus \Xi. \quad (1.2)$$

We specify the noise by $\wp(\xi := 0) = \mu$.

Equation 1.2 implies that the readout variable Y will reproduce the system variable X if $\Xi = 0$. If $\Xi = 1$, the readout variable is (in the language of logic) the negation of the system variable. That is to say, the readout has undergone a *bit-flip* error, so that $Y = 1$ when $X = 0$ and vice versa. We can thus interpret μ as the probability that the readout variable is ‘correct’. If no noise is added by the measurement apparatus, so that $\wp(\xi := 0) = 1$, we call the measurement *ideal*.

It should be intuitively clear that the apparatus state (i.e. the readout distribution) $\wp(y)$ is determined by the function G together with the noise probability $\wp(\xi)$ and the system state before the measurement, $\wp(x)$. This last state is called the a-priori state, or prior state. In the example above we find that

$$\wp(y := 1) = \mu \wp(x := 1) + (1 - \mu) \wp(x := 0), \quad (1.3)$$

$$\wp(y := 0) = \mu \wp(x := 0) + (1 - \mu) \wp(x := 1). \quad (1.4)$$

This may be written more succinctly by inverting Eq. (1.2) to obtain $\Xi = X \oplus Y$ and writing

$$\wp(y) = \sum_{x=0}^1 \wp(\xi := x \oplus y) \wp(x). \quad (1.5)$$

In the case of a binary variable X with distribution $\wp(x)$, it is easy to verify that the mean is given by $E[X] = \wp(x := 1)$. Here we are using E to represent ‘expectation of’. That is, in the general case,

$$E[X] = \sum_x x \Pr[X = x] = \sum_x x \wp(x). \quad (1.6)$$

More generally,

$$E[f(X)] = \sum_x f(x) \Pr[X = x] = \sum_x f(x) \wp(x). \quad (1.7)$$

Using this notation, we define the variance of a variable as

$$\text{Var}[X] \equiv E[X^2] - (E[X])^2. \quad (1.8)$$

Exercise 1.1 *Show that*

$$E[Y] = (1 - \mu) + (2\mu - 1)E[X], \quad (1.9)$$

$$\text{Var}[Y] = \mu(1 - \mu) + (2\mu - 1)^2 \text{Var}[X]. \quad (1.10)$$

Equation (1.9) shows that the average measurement result is the system variable mean, scaled by a factor of $2\mu - 1$, plus a constant off-set of $1 - \mu$. The scaling factor also appears in the variance equation (1.10), together with a constant (the first term) due to the noise added by the measurement process. When the measurement is ideal ($\mu = 1$), the mean and variance of the readout variable directly reflect the statistics of the measured system state.

1.1.3 Bayesian inference

We stated above that we are considering, at present, non-disturbing measurements, in which the system variable X is unaffected by the measurement. However, this does not mean that the system *state* is unaffected by the measurement. Recall that the state represents the observer’s incomplete knowledge of the system, and the point of making a measurement is (usually) to obtain more knowledge. Thus we *should* expect the state to change *given* that a certain readout is obtained.

The concept we are introducing here is the *conditional* state of the system, also known as the state conditioned on the readout. This state is sometimes called the a-posteriori state, or posterior state. The key to finding the conditioned state is to use Bayesian inference. Here, inference means that one infers information about the system from the readout, and Bayesian inference means doing this using Bayes’ theorem. This theorem is an elementary consequence of basic probability theory, via the double application of the conditional-probability definition

$$\Pr(A|B) = \Pr(A \cap B)/\Pr(B), \quad (1.11)$$

where A and B are events, $A \cap B$ is their intersection and $A|B$ is to be read as ‘ A given B ’. In an obvious generalization of this notation from events to the values of system variables,

Bayes' theorem says that the conditional system state may be written in terms of the a-priori (or prior) system state $\wp(x)$ as

$$\wp'(x|y) = \frac{\wp(y|x)\wp(x)}{\wp(y)}. \quad (1.12)$$

Here the prime emphasizes that this is an a-posteriori state, and (sticking to the discrete case as usual)

$$\wp(y) = \sum_x \wp(y|x)\wp(x), \quad (1.13)$$

as required for the conditional state to be normalized.

The crucial part of Bayesian inference is the known conditional probability $\wp(y|x)$, also known as the 'forward probability'. This is related to the measurement noise and the function G as follows:

$$\wp(y|x) = \sum_{\xi} \wp(y|x, \xi)\wp(\xi) = \sum_{\xi} \delta_{y, G(x, \xi)}\wp(\xi). \quad (1.14)$$

Here $\wp(y|x, \xi)$ means the state of y given the values of x and ξ . If the output function $Y = G(X, \Xi)$ is invertible in the sense that there is a function G^{-1} such that $\Xi = G^{-1}(X, Y)$, then we can further simplify this as

$$\wp(y|x) = \sum_{\xi} \delta_{\xi, G^{-1}(x, y)}\wp(\xi) = \wp(\xi := G^{-1}(x, y)). \quad (1.15)$$

Thus we obtain finally for the conditional system state

$$\wp'(x|y) = \frac{\wp(\xi := G^{-1}(x, y))\wp(x)}{\wp(y)}. \quad (1.16)$$

Exercise 1.2 *If you are unfamiliar with probability theory, derive the first equality in Eq. (1.14).*

As well as defining the conditional post-measurement system state, we can define an unconditional posterior state by averaging over the possible measurement results:

$$\wp'(x) = \sum_y \wp'(x|y)\wp(y) = \sum_y \wp(\xi := G^{-1}(x, y))\wp(x). \quad (1.17)$$

The terms conditional and unconditional are sometimes replaced by the terms selective and non-selective, respectively. In this case of a non-disturbing measurement, it is clear that

$$\wp'(x) = \wp(x). \quad (1.18)$$

That is, the unconditional posterior state is always the same as the prior state. This is the counterpart of the statement that the system variable X is unaffected by the measurement.

Exercise 1.3 *Determine the posterior conditional states $\wp'(x|y)$ in the above binary example for the two cases $y = 0$ and $y = 1$. Show that, in the limit $\mu \rightarrow 1$, $\wp'(x|y) \rightarrow \delta_{x, y}$ (assuming that $\wp(x := y) \neq 0$), whereas, in the case $\mu = 1/2$, $\wp'(x|y) = \wp(x)$. Interpret these results.*

1.1.4 Example: continuous variables

We now turn to the case of continuous state variables. Suppose one is interested in determining the position of a particle on the real line. Let the a-priori state of this system be the probability density $\wp(x)$. As explained earlier, this means that, if an ideal measurement of the position X is made, then the probability that a value between x and $x + dx$ will be obtained is $\wp(x)dx$.

As in the binary case, we introduce an apparatus with configuration Y and a noise variable Ξ , both real numbers. To define the measurement we specify the output function, $G(X, \Xi)$, for example

$$Y = X + \Xi, \quad (1.19)$$

so that $\Xi = G^{-1}(X, Y) = Y - X$. We must also specify the probability density for the noise variable ξ , and a common choice is a zero-mean Gaussian with a variance Δ^2 :

$$\wp(\xi) = (2\pi\Delta^2)^{-1/2} e^{-\xi^2/(2\Delta^2)}. \quad (1.20)$$

The post-measurement apparatus state is given by the continuous analogue of Eq. (1.13),

$$\wp(y) = \int_{-\infty}^{\infty} \wp(y|x)\wp(x)dx. \quad (1.21)$$

Exercise 1.4 Show that the mean and variance of the state $\wp(y)$ are $E[X]$ and $\text{Var}[X] + \Delta^2$, respectively. This clearly shows the effect of the noise.

Finding the conditional states in this case is difficult in general. However, it is greatly simplified if the a-priori system state is Gaussian:

$$\wp(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right), \quad (1.22)$$

because then the conditional states are still Gaussian.

Exercise 1.5 Verify this, and show that the conditional mean and variance given a result y are, respectively,

$$\bar{x}' = \frac{\sigma^2 y + \Delta^2 \bar{x}}{\Delta^2 + \sigma^2}, \quad (\sigma')^2 = \frac{\sigma^2 \Delta^2}{\Delta^2 + \sigma^2}. \quad (1.23)$$

Hence show that, in the limit $\Delta \rightarrow 0$, the conditional state $\wp'(x|y)$ converges to $\delta(x - y)$, and an ideal measurement is recovered.

1.1.5 Most general formulation of classical measurements

As stated above, so far we have considered only non-disturbing classical measurements; that is, measurements with no *back-action* on the system. However, it is easy to consider classical measurements that do have a back-action on the system. For example, one could measure whether or not a can has petrol fumes in it by dropping a lit match inside. The

result (nothing, or flames) will certainly reveal whether or not there *was* petrol inside the can, but the final state of the system after the measurement will have no petrol fumes inside in either case.

We can generalize Bayes' theorem to deal with this case by allowing a state-changing operation to act upon the state after applying Bayes' theorem. Say the system state is $\wp(x)$. For simplicity we will take X to be a discrete random variable, with the configuration space being $\{0, 1, \dots, n-1\}$. Say Y is the result of the measurement as usual. Then this state-changing operation is described by an $n \times n$ matrix \mathcal{B}_y , whose element $\mathcal{B}_y(x|x')$ is the probability that the measurement will cause the system to make a transition, from the state in which $X = x'$ to the state in which $X = x$, given that the result $Y = y$ was obtained. Thus, for all x' and all y ,

$$\mathcal{B}_y(x|x') \geq 0, \quad \sum_x \mathcal{B}_y(x|x') = 1. \quad (1.24)$$

The posterior system state is then given by

$$\wp'(x|y) = \frac{\sum_{x'} \mathcal{B}_y(x|x') \wp(y|x') \wp(x')}{\wp(y)}, \quad (1.25)$$

where the expression for $\wp(y)$ is unchanged from before.

We can unify the Bayesian part and the back-action part of the above expression by defining a new $n \times n$ matrix \mathcal{O}_y with elements

$$\mathcal{O}_y(x|x') = \mathcal{B}_y(x|x') \wp(y|x'), \quad (1.26)$$

which maps a normalized probability distribution $\wp(x)$ onto an unnormalized probability distribution:

$$\tilde{\wp}'(x|y) = \sum_{x'} \mathcal{O}_y(x|x') \wp(x'). \quad (1.27)$$

Here we are introducing the convention of using a tilde to indicate an unnormalized state, with a norm of less than unity. This norm is equal to

$$\wp(y) = \sum_x \sum_{x'} \mathcal{O}_y(x|x') \wp(x'), \quad (1.28)$$

the probability of obtaining the result $Y = y$. Maps that take states to (possibly unnormalized) states are known as *positive maps*. The normalized conditional system state is

$$\wp'(x|y) = \sum_{x'} \mathcal{O}_y(x|x') \wp(x') / \wp(y). \quad (1.29)$$

From the properties of \mathcal{O}_y , it follows that it is possible to find an n -vector E_y with positive elements $E_y(x)$, such that the probability formula simplifies:

$$\sum_x \sum_{x'} \mathcal{O}_y(x|x') \wp(x') = \sum_x E_y(x) \wp(x). \quad (1.30)$$

Specifically, in terms of Eq. (1.26),

$$E_y(x) = \wp(y|x). \quad (1.31)$$

This satisfies the completeness condition

$$\forall x, \sum_y E_y(x) = 1. \quad (1.32)$$

This is the only mathematical restriction on $\{\mathcal{O}_y: y\}$ (apart from requiring that it be a positive map).

Exercise 1.6 *Formulate the match-in-the-tin measurement technique described above. Let the states $X = 1$ and $X = 0$ correspond to petrol fumes and no petrol fumes, respectively. Let the results $Y = 1$ and $Y = 0$ correspond to flames and no flames, respectively. Determine the two matrices $\mathcal{O}_{y:=0}$ and $\mathcal{O}_{y:=1}$ (each of which is a 2×2 matrix).*

The unconditional system state after the measurement is

$$\wp'(x) = \sum_y \sum_{x'} \mathcal{O}_y(x|x') \wp(x') = \sum_{x'} \mathcal{O}(x|x') \wp(x'). \quad (1.33)$$

Here the unconditional evolution map \mathcal{O} is

$$\mathcal{O} = \sum_y \mathcal{O}_y. \quad (1.34)$$

Exercise 1.7 *Show that \mathcal{O} is the identity if and only if there is no back-action.*

1.2 Quantum measurement theory

1.2.1 Probability and quantum mechanics

As we have discussed, with a classical system an ideal measurement can determine with certainty the values of all of the system variables. In this situation of complete knowledge, all subsequent ideal measurement results are determined with certainty. In consequence, measurement and probability do not play a significant role in the foundation of classical mechanics (although they do play a very significant role in practical applications of classical mechanics, where noise is inevitable [Whi96]).

The situation is very different in quantum mechanics. Here, for any sort of measurement, there are systems about which one has maximal knowledge, but for which the result of the measurement is not determined. The best one can do is to give the probability distributions for measurement outcomes. From this it might be inferred that a state of maximal knowledge about a quantum system is not a state of complete knowledge. That is, that there are ‘hidden’ variables about which one has incomplete knowledge, even when one has maximal knowledge, and these hidden variables determine the measurement outcomes.

Although it is possible to build a perfectly consistent interpretation of quantum mechanics based on this idea (see for example Ref. [BH93]), most physicists reject the idea. Probably the chief reason for this rejection is that in 1964 John Bell showed that any such deterministic hidden-variables theory must be nonlocal (that is, it must violate local causality) [Bel64]. That is, in such a theory, an agent with control over some local macroscopic parameters (such as the orientation of a magnet) can, under particular circumstances, instantaneously affect the hidden variables at an arbitrarily distant point in space. It can be shown that this effect cannot allow faster-than-light signalling, and hence does not lead to causal paradoxes within Einstein's theory of relativity. Nevertheless, it is clearly against the spirit of relativity theory. It should be noted, however, that this result is not restricted to hidden-variable interpretations; any interpretation of quantum mechanics that allows the concept of local causality to be formulated will be found to violate it – see Refs. [Bel87] (p. 172) and [Wis06].

Another, perhaps better, justification for ignoring hidden-variables theories is that there are infinitely many of them (see Ref. [GW04] and references therein). While some are more natural than others [Wis07], there is, at this stage, no compelling reason to choose one over all of the others. Thus one would be forced to make a somewhat arbitrary choice as to which hidden-variables interpretation to adopt, and each interpretation would have its own unique explanation as to the nature of quantum-mechanical uncertainty.

Rather than grappling with these difficulties, in this book we take an *operational* approach. That is, we treat quantum mechanics as simply an algorithm for calculating what one expects to happen when one performs a measurement. We treat uncertainty about future measurement outcomes as a primitive in the theory, rather than ascribing it to lack of knowledge about existing hidden variables.

We will still talk of a quantum state as representing our knowledge about a system, even though strictly it is our knowledge about the outcomes of our future measurements on that system. Also it is still useful, in many cases, to think of a quantum state of maximal knowledge as being like a classical state of incomplete knowledge about a system. Very crudely, this is the idea of 'quantum noise'. The reader who is not familiar with basic quantum mechanics (pure states, mixed states, time-evolution, entanglement etc.) should consult Appendix A for a summary of this material. However, before moving to quantum measurements, we note an important point of terminology. The matrix representation ρ of a mixed quantum state is usually called (for historical reasons) the density operator, or density matrix. We will call it the *state matrix*, because it generalizes the *state vector* for pure states.

Finally, just as in the classical case, observers with different knowledge may assign different states simultaneously to a single system. The most natural way to extend the concept of consistency to the quantum case is to replace the common state of maximal knowledge with a common pure state. That is, the condition for the consistency of a collection of states $\{\rho_j\}$ from different observers is that there exists a positive ϵ and a ket $|\psi\rangle$ such that, for all j , $\rho_j - \epsilon|\psi\rangle\langle\psi|$ is a positive operator. In other words, each observer's state ρ_j can be written as a mixture of the pure state $|\psi\rangle\langle\psi|$ and some other states.

Exercise 1.8 Show, from this definition, that two different pure states $\hat{\pi}_1$ and $\hat{\pi}_2$ cannot be consistent states for any system.

Hint: Consider the operator $\hat{\sigma}_j = \hat{\pi}_j - \epsilon |\psi\rangle\langle\psi|$. Assuming $\hat{\pi}_j \neq |\psi\rangle\langle\psi|$, show that $\text{Tr}[\sigma_j^2] > (\text{Tr}[\sigma_j])^2$ and hence deduce the result.

1.2.2 Projective measurements

The traditional description of measurement in quantum mechanics is in terms of projective measurements, as follows. Consider a measurement of the physical quantity Λ . First we note that the associated operator $\hat{\Lambda}$ (often called an *observable*) can be diagonalized as

$$\hat{\Lambda} = \sum_{\lambda} \lambda \hat{\Pi}_{\lambda}, \quad (1.35)$$

where $\{\lambda\}$ are the eigenvalues of $\hat{\Lambda}$ which are real and which we have assumed for convenience are discrete. $\hat{\Pi}_{\lambda}$ is called the projection operator, or projector, onto the subspace of eigenstates of $\hat{\Lambda}$ with eigenvalue λ . If the spectrum (set of eigenvalues $\{\lambda\}$) is non-degenerate, then the projector would simply be the rank-1 projector $\hat{\pi}_{\lambda} = |\lambda\rangle\langle\lambda|$. We will call this special case *von Neumann measurements*.

In the more general case, where the eigenvalues of $\hat{\Lambda}$ are N_{λ} -fold degenerate, $\hat{\Pi}_{\lambda}$ is a rank- N_{λ} projector, and can be written as $\sum_{j=1}^{N_{\lambda}} |\lambda, j\rangle\langle\lambda, j|$. For example, in the simplest model of the hydrogen atom, if Λ is the energy then λ would be the principal quantum number n and j would code for the angular-momentum and spin quantum numbers l, m and s of states with the same energy. The projectors are orthonormal, obeying

$$\hat{\Pi}_{\lambda} \hat{\Pi}_{\lambda'} = \delta_{\lambda, \lambda'} \hat{\Pi}_{\lambda}. \quad (1.36)$$

The existence of this orthonormal basis is a consequence of the spectral theorem (see Box 1.1).

When one measures Λ , the result one obtains is one of the eigenvalues λ . Say the measurement begins at time t and takes a time T . Assuming that the system does not evolve significantly from other causes during the measurement, the probability for obtaining that particular eigenvalue is

$$\text{Pr}[\Lambda(t) = \lambda] = \wp_{\lambda} = \text{Tr}[\rho(t) \hat{\Pi}_{\lambda}]. \quad (1.37)$$

After the measurement, the conditional (a-posteriori) state of the system given the result λ is

$$\rho_{\lambda}(t + T) = \frac{\hat{\Pi}_{\lambda} \rho(t) \hat{\Pi}_{\lambda}}{\text{Pr}[\Lambda(t) = \lambda]}. \quad (1.38)$$

That is to say, the final state has been projected by $\hat{\Pi}_{\lambda}$ into the corresponding subspace of the total Hilbert space. This is known as the *projection postulate*, or sometimes as *state collapse*, or *state reduction*. The last term is best avoided, since it invites confusion with the reduced state of a bipartite system as discussed in Section A.2.2 of Appendix A. This process

Box 1.1 Spectral theorem

The spectral theorem states that any *normal operator* \hat{N} has a complete set of eigenstates that are orthonormal. A normal operator is an operator such that $[\hat{N}, \hat{N}^\dagger] = 0$. That is, for every such \hat{N} there is a basis $\{|v, j\rangle: v, j\}$ for the Hilbert space such that, with $v \in \mathbb{C}$,

$$\hat{N}|v, j\rangle = v|v, j\rangle.$$

Here the extra index j is necessary because the eigenvalues v may be degenerate. The general diagonal form of \hat{N} is

$$\hat{N} = \sum_v v \hat{\Pi}_v,$$

where the $\hat{\Pi}_v = \sum_j |v, j\rangle\langle v, j|$ form a set of orthogonal projectors obeying

$$\hat{\Pi}_v \hat{\Pi}_{v'} = \delta_{v,v'} \hat{\Pi}_v.$$

Hermitian operators (for which $\hat{N} = \hat{N}^\dagger$) are a special class of normal operators. It can be shown that a normal operator that is not Hermitian can be written in the form $\hat{N} = \hat{R} + i\hat{H}$, where \hat{R} and \hat{H} are commuting Hermitian operators. For any two operators that commute, there is some complete basis comprising states that are eigenstates of both operators, which in this case will be a basis $|v, j\rangle$ diagonalizing \hat{N} . As will be discussed in Section 1.2.2, operators that share eigenstates are simultaneously measurable.

Thus it is apparent that a non-Hermitian normal operator is really just a compact way to represent two simultaneously observable quantities by having eigenvalues v in the complex plane rather than the real line. By considering vectors or other multi-component objects, any number of commuting operators can be combined to represent the corresponding simultaneously observable quantities. This demonstrates that, for projective quantum measurement theory, the important thing is not an operator representing the observables, but rather the projector $\hat{\Pi}_r$ corresponding to a result r .

should be compared to the classical Bayesian update rule, Eq. (1.12). A consequence of this postulate is that, if the measurement is immediately repeated, then

$$\Pr[\Lambda(t+T) = \lambda' | \Lambda(t) = \lambda] = \text{Tr}[\rho_\lambda(t+T) \hat{\Pi}_{\lambda'}] = \delta_{\lambda', \lambda}. \quad (1.39)$$

That is to say, the same result is guaranteed. Moreover, the system state will not be changed by the second measurement. For a deeper understanding of the above theory, see Box 1.2.

For pure states, $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$, the formulae (1.37) and (1.38) can be more simply expressed as

$$\Pr[\Lambda(t) = \lambda] = \wp_\lambda = \langle\psi(t)|\hat{\Pi}_\lambda|\psi(t)\rangle \quad (1.40)$$

Box 1.2 Gleason's theorem

It is interesting to ask how much of quantum measurement theory one can derive from assuming that quantum measurements are described by a complete set of projectors, one for each result r . Obviously there must be some rule for obtaining a probability \wp_r from a projector $\hat{\Pi}_r$, such that $\sum_r \wp_r = 1$. Gleason [Gle57] proved that, if one considers measurements with at least three outcomes (requiring a Hilbert-space dimension of at least three), then it follows that there exists a non-negative operator ρ of unit trace such that

$$\wp_r = \text{Tr}[\hat{\Pi}_r \rho].$$

That is, the probability rule (1.37) can be derived, not assumed.

It must be noted, however, that Gleason required an additional assumption: *non-contextuality*. This means that \wp_r depends only upon $\hat{\Pi}_r$, being independent of the other projectors which complete the set. That is, if two measurements each have one outcome represented by the same projector $\hat{\Pi}_r$, the probabilities for those outcomes are necessarily the same, even if the measurements cannot be performed simultaneously.

Gleason's theorem shows that the state matrix ρ is a consequence of the structure of Hilbert space, if we require probabilities to be assigned to projection operators. It suggests that, rather than introducing pure states and then generalizing to mixed states, the state matrix ρ can be taken as fundamental.

and

$$|\psi_\lambda(t + T)\rangle = \hat{\Pi}_\lambda |\psi(t)\rangle / \sqrt{\wp_\lambda}. \quad (1.41)$$

However, if one wishes to describe the unconditional state of the system (that is, the state *if one makes the measurement, but ignores the result*) then one must use the state matrix:

$$\rho(t + T) = \sum_\lambda \text{Pr}[\Lambda(t) = \lambda] \rho_\lambda(t + T) = \sum_\lambda \hat{\Pi}_\lambda \rho(t) \hat{\Pi}_\lambda. \quad (1.42)$$

Thus, if the state were pure at time t , and we make a measurement, but ignore the result, then in general the state at time $t + T$ will be mixed. That is, projective measurement, unlike unitary evolution,² is generally an entropy-increasing process unless one keeps track of the measurement results. This is in contrast to non-disturbing measurements in classical mechanics, where (as we have seen) the unconditional a-posteriori state is identical to the a-priori state (1.17).

Exercise 1.9 Show that a projective measurement of Λ decreases the purity $\text{Tr}[\rho^2]$ of the unconditional state unless the a-priori state $\rho(t)$ can be diagonalized in the same basis as can $\hat{\Lambda}$.

² Of course unitary evolution *can* change the entropy of a *subsystem*, as we will discuss in Chapter 3.

Hint: Let $p_{\lambda\lambda'} = \text{Tr}[\hat{\Pi}_\lambda \rho(t) \hat{\Pi}_{\lambda'} \rho(t)]$ and show that $\forall \lambda, \lambda', p_{\lambda\lambda'} \geq 0$. Then express $\text{Tr}[\rho(t)^2]$ and $\text{Tr}[\rho(t+T)^2]$ in terms of these $p_{\lambda\lambda'}$.

From the above measurement theory, it is simple to show that the mean value for the result Λ is

$$\begin{aligned} \langle \Lambda \rangle &= \sum_{\lambda} \text{Pr}[\Lambda = \lambda] \lambda \\ &= \sum_{\lambda} \text{Tr}[\rho \hat{\Pi}_\lambda] \lambda \\ &= \text{Tr} \left[\rho \left(\sum_{\lambda} \lambda \hat{\Pi}_\lambda \right) \right] = \text{Tr}[\rho \hat{\Lambda}]. \end{aligned} \quad (1.43)$$

Here we are using angle brackets as an alternative notation for expectation value when dealing with quantum observables.

Exercise 1.10 Using the same technique, show that

$$\langle \Lambda^2 \rangle = \sum_{\lambda} \lambda^2 \text{Pr}[\Lambda = \lambda] = \text{Tr}[\rho \hat{\Lambda}^2]. \quad (1.44)$$

Thus the mean value (A.6) and variance (A.8) can be derived rather than postulated, provided that they are interpreted in terms of the moments of the results of a projective measurement of Λ .

Continuous spectra. The above results can easily be generalized to treat physical quantities with a continuous spectrum, such as the position \hat{X} of a particle on a line. Considering this non-degenerate case for simplicity, the spectral theorem becomes

$$\hat{X} = \int_{-\infty}^{\infty} x \hat{\Pi}(x) dx = \int_{-\infty}^{\infty} x |x\rangle \langle x| dx. \quad (1.45)$$

Note that $\hat{\Pi}(x)$ is not strictly a projector, but a projector density, since the orthogonality conditions are

$$\hat{\Pi}(x) \hat{\Pi}(x') = \delta(x - x') \hat{\Pi}(x), \quad (1.46)$$

or, in terms of the unnormalizable *improper states* $|x\rangle$,

$$\langle x | x' \rangle = \delta(x - x'). \quad (1.47)$$

These states are discussed in more detail in Appendix A.

The measurement outcomes are likewise described by probability densities. For example, if the system is in a pure state $|\psi(t)\rangle$, the probability that an ideal measurement of position gives a result between x and $x + dx$ is

$$\wp(x) dx = \text{Tr}[|\psi(t)\rangle \langle \psi(t)| \hat{\Pi}(x)] dx = |\psi(x, t)|^2 dx. \quad (1.48)$$

Here we have defined the wavefunction $\psi(x, t) = \langle x | \psi(t) \rangle$. Unfortunately it is not possible to assign a proper a-posteriori state to the system, since $|x\rangle$ is unnormalizable. This problem

can be avoided by considering an approximate measurement of position with finite accuracy Δ , as will always be the case in practice. This can still be described as a projective measurement, for example by using the (now discrete) set of projectors

$$\hat{\Pi}_j = \int_{x_j}^{x_{j+1}} \hat{\Pi}(x) dx, \quad (1.49)$$

where, for all j , $x_{j+1} = x_j + \Delta$.

Exercise 1.11 Show that the $\hat{\Pi}_j$ defined here form an orthonormal set.

Simultaneous measurements. Heisenberg's uncertainty relation (see Exercise A.3) shows that it is impossible for both the position and the momentum of a particle to be known exactly (have zero variance). This is often used as an argument for saying that it is impossible simultaneously to measure position and momentum. We will see in the following sections that this is not strictly true. Nevertheless, it is the case that it is impossible to carry out a simultaneous *projective* measurement of position and momentum, and this is the case of interest here.

For two quantities A and B to be measurable simultaneously, it is sufficient (and necessary) for them to be measurable consecutively, such that the joint probability of the results a and b does not depend on the order of the measurement. Considering a system in a pure state for simplicity, we thus require for all a and b and all $|\psi\rangle$ that

$$\hat{\Pi}_a \hat{\Pi}_b |\psi\rangle = e^{i\theta} \hat{\Pi}_b \hat{\Pi}_a |\psi\rangle, \quad (1.50)$$

for some θ . By considering the norm of these two vectors (which must be equal) it can be seen that $e^{i\theta}$ must equal unity.

Exercise 1.12 Prove this, and hence that

$$\forall a, b, [\hat{\Pi}_a, \hat{\Pi}_b] = 0. \quad (1.51)$$

This is equivalent to the condition that $[\hat{A}, \hat{B}] = 0$, and means that there is a basis, say $\{|\phi_k\rangle\}$, in which both \hat{A} and \hat{B} are diagonal. That is,

$$\hat{A} = \sum_k a_k |\phi_k\rangle \langle \phi_k|, \quad \hat{B} = \sum_k b_k |\phi_k\rangle \langle \phi_k|, \quad (1.52)$$

where the eigenvalues a_k may be degenerate (that is, there may exist k and k' such that $a_k = a_{k'}$) and similarly for b_k . Thus, one way of making a simultaneous measurement of A and B is to make a measurement of $\hat{K} = \sum_k k |\phi_k\rangle \langle \phi_k|$, and from the result k determine the appropriate values a_k and b_k for A and B .

1.2.3 Systems and meters

The standard (projective) presentation of quantum measurements is inadequate for a number of reasons. A prosaic, but very practical, reason is that very few measurements can be

made in such a way that the apparatus adds no classical noise to the measurement result. A more interesting reason is that there are many measurements in which the a-posteriori conditional system state is clearly not left in the eigenstate of the measured quantity corresponding to the measurement result. For example, in photon counting by a photodetector, at the end of the measurements all photons have been absorbed, so that the system (e.g. the cavity that originally contained the photons) is left in the vacuum state, not a state containing the number n of photons counted. Another interesting reason is that non-projective measurements allow far greater flexibility than do projective measurements. For example, the simultaneous measurement of position and momentum is a perfectly acceptable idea, so long as the respective accuracies do not violate the Heisenberg uncertainty principle, as we will discuss below.

The fundamental reason why projective measurements are inadequate for describing real measurements is that experimenters never directly measure the system of interest. Rather, the system of interest (such as an atom) interacts with its environment (the continuum of electromagnetic field modes), and the experimenter observes the effect of the system on the environment (the radiated field). Of course, one could argue that the experimenter does not observe the radiated field, but rather that the field interacts with a photodetector, which triggers a current in a circuit, which is coupled to a display panel, which radiates more photons, which interact with the experimenter's retina, and so on. Such a chain of systems is known as a von Neumann chain [vN32]. The point is that, at some stage before reaching the mind of the observer, one has to cut the chain by applying the projection postulate. This cut, known as Heisenberg's cut [Hei30], is the point at which one considers the measurement as having been made.

If one were to apply a projection postulate directly to the atom, one would obtain wrong predictions. However, assuming a projective measurement of the field will yield results negligibly different from those obtained assuming a projective measurement at any later stage. This is because of the rapid decoherence of macroscopic material objects such as photodetectors (see Chapter 3). For this reason, it is sufficient to consider the field to be measured projectively. Because the field has interacted with the system, their quantum states are correlated (indeed, they are entangled, provided that their initial states are pure enough). The projective measurement of the field is then effectively a measurement of the atom. The latter measurement, however, is not projective, and we need a more general formalism to describe it.

Let the initial system state vector be $|\psi(t)\rangle$, and say that there is a second quantum system, which we will call the *meter*, or *apparatus*, with the initial state $|\theta(t)\rangle$. Thus the initial (unentangled) combined state is

$$|\Psi(t)\rangle = |\theta(t)\rangle|\psi(t)\rangle. \quad (1.53)$$

Let these two systems be coupled together for a time T_1 by a unitary evolution operator $\hat{U}(t + T_1, t)$, which we will write as $\hat{U}(T_1)$. Thus the combined system-meter state after

this coupling is

$$|\Psi(t + T_1)\rangle = \hat{U}(T_1)|\theta(t)\rangle|\psi(t)\rangle. \quad (1.54)$$

This cannot in general be written in the factorized form of Eq. (1.53).

Now let the meter be measured projectively over a time interval T_2 , and say $T = T_1 + T_2$. We assume that the evolution of the system and meter over the time T_2 is negligible (this could be either because $T_2 \ll T_1$, or because the coupling Hamiltonian is time-dependent). Let the projection operators for the meter be rank-1 operators, so that $\hat{\Pi}_r = \hat{\pi}_r \otimes \hat{1}$. The order of the tensor product is meter then system, as in Eq. (1.53), and $\hat{\pi}_r = |r\rangle\langle r|$. Here r denotes the value of the observed quantity R . The set $\{|r\rangle\}$ forms an orthonormal basis for the meter Hilbert space. Then the final combined state is

$$|\Psi_r(t + T)\rangle = \frac{|r\rangle\langle r|\hat{U}(T_1)|\theta(t)\rangle|\psi(t)\rangle}{\sqrt{\wp_r}}, \quad (1.55)$$

where the probability of obtaining the value r for the result R is

$$\Pr[R = r] = \wp_r = \langle\psi(t)|\langle\theta(t)|\hat{U}^\dagger(T_1)[|r\rangle\langle r| \otimes \hat{1}]\hat{U}(T_1)|\theta(t)\rangle|\psi(t)\rangle. \quad (1.56)$$

The measurement on the meter disentangles the system and the meter, so that the final state (1.55) can be written as

$$|\Psi_r(t + T)\rangle = |r\rangle \frac{\hat{M}_r|\psi(t)\rangle}{\sqrt{\wp_r}}, \quad (1.57)$$

where \hat{M}_r is an operator that acts only in the system Hilbert space, defined by

$$\hat{M}_r = \langle r|\hat{U}(T_1)|\theta(t)\rangle. \quad (1.58)$$

We call it a *measurement operator*. The probability distribution (1.56) for R can similarly be written as

$$\wp_r = \langle\psi(t)|\hat{M}_r^\dagger \hat{M}_r|\psi(t)\rangle. \quad (1.59)$$

1.2.4 Example: binary measurement

To understand the ideas just introduced, it is helpful to consider a specific example. We choose one analogous to the classical discrete binary measurement discussed in Section 1.1.2. The quantum analogue of a system with a single binary system variable is a quantum system in a two-dimensional Hilbert space. Let $\{|x\rangle: x = 0, 1\}$ be an orthonormal basis for this Hilbert space. An obvious physical realization is a spin-half particle. The spin in any direction is restricted to one of two possible values, $\pm\hbar/2$. These correspond to the spin being up (+) or down (−) with respect to the given direction. Choosing a particular direction (z is conventional), we label these states as $|0\rangle$ and $|1\rangle$, respectively. Other physical realizations include an atom with only two relevant levels, or a single electromagnetic cavity mode containing no photon or one photon. The latter two examples will be discussed in detail in Section 1.5.

Now consider a measured system S and a measurement apparatus A , both described by two-dimensional Hilbert spaces. We will use x for the system states and y for the apparatus states. Following the above formalism, we assume that initially both systems are in pure states, so that the joint state of the system at time t is

$$|\Psi(t)\rangle = |\theta(t)\rangle|\psi(t)\rangle = \sum_{x,\xi} s_x a_\xi |y := \xi\rangle|x\rangle. \quad (1.60)$$

Note that we have used an analogous notation to the classical case, so that $|y := \xi\rangle$ is the apparatus state $|y\rangle$ with y taking the value ξ . To make a measurement, the system and apparatus states must become correlated. We will discuss how this may take place physically in Section 1.5. For now we simply postulate that, as a result of the unitary interaction between the system and the apparatus, we have

$$|\Psi(t + T_1)\rangle = \hat{G}|\Psi(t)\rangle = \sum_{x,\xi} a_\xi s_x |y := G(x, \xi)\rangle|x\rangle, \quad (1.61)$$

where \hat{G} is a unitary operator defined by

$$\hat{G}|y := \xi\rangle|x\rangle = |y := G(x, \xi)\rangle|x\rangle. \quad (1.62)$$

Note that the interaction between the system and the apparatus has been specified by reference to a particular basis for the system and apparatus, $\{|y\rangle|x\rangle\}$. We will refer to this (for the system, or apparatus, or both together) as the *measurement basis*.

Exercise 1.13 Show that \hat{G} as defined is unitary if there exists an inverse function G^{-1} in the sense that, for all y , $y = G(x, G^{-1}(x, y))$.

Hint: Show that $\hat{G}^\dagger \hat{G} = \hat{1} = \hat{G} \hat{G}^\dagger$ using the matrix representation in the measurement basis.

The invertibility condition is the same as we used in Section 1.1.3 for the classical binary measurement model.

As an example, consider $G(x, \xi) = x \oplus \xi$, as in the classical case, where again this indicates binary addition. In this case $\hat{G} = \hat{G}^{-1}$. The system state is unknown and is thus arbitrary. However, the apparatus is assumed to be under our control and can be prepared in a *fiducial* state. This means a standard state for the purpose of measurement. Often the fiducial state is a particular state in the measurement basis, and we will assume that it is $|y := 0\rangle$, so that $a_\xi = \delta_{\xi,0}$. In this case the state after the interaction is

$$|\Psi(t + T_1)\rangle = \sum_x s_x |y := x\rangle|x\rangle \quad (1.63)$$

and there is a perfect correlation between the system and the apparatus. Let us say a projective measurement (of duration T_2) of the apparatus state in the measurement basis is made. This will give the result y with probability $|s_y|^2$, that is, with exactly the probability that a projective measurement directly on the system in the measurement basis would have given. Moreover, the conditioned system state at time $t + T$ (where $T = T_1 + T_2$ as above), given the result y , is

$$|\psi_y(t + T)\rangle = |x := y\rangle. \quad (1.64)$$

Again, this is as would have occurred with the appropriate projective measurement of duration T on the system, as in Eq. (1.41).

This example is a special case of a model introduced by von Neumann. It would appear to be simply a more complicated version of the description of standard projective measurements. However, as we now show, it enables us to describe a more general class of measurements in which extra noise appears in the result due to the measurement apparatus.

Suppose that for some reason it is not possible to prepare the apparatus in one of the measurement basis states. In that case we must use the general result given in Eq. (1.61). Using Eq. (1.58), we find

$$\begin{aligned}\hat{M}_y|\psi(t)\rangle &= \langle y|\Psi(t+T_1)\rangle \\ &= \sum_{x,\xi} \delta_{y,G(x,\xi)} a_\xi s_x |x\rangle \\ &= \sum_x a_{G^{-1}(x,y)} s_x |x\rangle \\ &= \sum_{x'} a_{G^{-1}(x',y)} |x'\rangle \langle x'| \sum_x s_x |x\rangle.\end{aligned}\quad (1.65)$$

Thus we have the measurement operator

$$\hat{M}_y = \sum_x a_{G^{-1}(x,y)} |x\rangle \langle x|. \quad (1.66)$$

For the particular case $G^{-1}(x, y) = x \oplus y$, this simplifies to

$$\hat{M}_y = \sum_\xi a_\xi |x := y \oplus \xi\rangle \langle x := y \oplus \xi|. \quad (1.67)$$

Returning to the more general form of Eq. (1.65), we find that the probability for the result y is

$$\wp(y) = \langle \psi(t) | \hat{M}_y^\dagger \hat{M}_y | \psi(t) \rangle = \sum_x |s_x|^2 |a_{G^{-1}(x,y)}|^2. \quad (1.68)$$

If we define

$$\wp(\xi) = |a_\xi|^2, \quad (1.69)$$

$$\wp(x) = |s_x|^2 = \text{Tr}[\rho(t)|x\rangle\langle x|], \quad (1.70)$$

where $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$ is the system state matrix, then the probability distribution for measurement results may then be written as

$$\wp(y) = \sum_x \wp(\xi := G^{-1}(x, y)) \wp(x). \quad (1.71)$$

This is the same form as for the classical binary measurement scheme; see Eq. (1.13) and Eq. (1.15). Here the noise distribution arises from *quantum noise* associated with the fiducial (purposefully prepared) apparatus state. It is quantum noise because the initial

apparatus state is still a pure state. The noise arises from the fact that it is not prepared in one of the measurement basis states. Of course, the apparatus may be prepared in a mixed state, in which case the noise added to the measurement result may have a classical origin. This is discussed below in Section 1.4.

The system state conditioned on the result y is

$$|\psi_y(t+T)\rangle = \hat{M}_y |\psi(t)\rangle / \sqrt{\wp(y)} = \sum_x a_{G^{-1}(x,y)} s_x |x\rangle / \sqrt{\wp(y)}. \quad (1.72)$$

If, from this, we calculate the probability $|\langle x | \psi_y(t+T) \rangle|^2$ for the system to have $X = x$ after the measurement giving the result y , we find this probability to be given by

$$\wp'(x|y) = \frac{\wp(y|x)\wp(x)}{\wp(y)}. \quad (1.73)$$

Again, this is the same as the classical result derived using Bayes' theorem. The interesting point is that the projection postulate does that work for us in the quantum case. Moreover, it gives us the full a-posteriori conditional state, from which the expectation value of any observable (not just X) can be calculated. The quantum measurement here is thus more than simply a reproduction of the classical measurement, since the conditional state (1.72) cannot be derived from Bayes' theorem.

Exercise 1.14 Consider two infinite-dimensional Hilbert spaces describing a system and a meter. Show that the operator \hat{G} , defined in the joint position basis $|y\rangle|x\rangle$ by

$$\hat{G}|y := \xi\rangle|x\rangle = |y := \xi + x\rangle|x\rangle, \quad (1.74)$$

is unitary. Let the fiducial apparatus state be

$$|\theta\rangle = \int_{-\infty}^{\infty} d\xi [(2\pi\Delta^2)^{-1/2} \exp(-\xi^2/(2\Delta^2))]^{1/2} |y := \xi\rangle. \quad (1.75)$$

Following the example of this subsection, show that, insofar as the statistics of \hat{X} are concerned, this measurement is equivalent to the classical measurement analysed in Section 1.1.4.

1.2.5 Measurement operators and effects

As discussed in Section 1.2.3, the system and apparatus are no longer entangled at the end of the measurement. Thus it is not necessary to continue to include the meter in our description of the measurement. Rather we can specify the measurement completely in terms of the measurement operators \hat{M}_r . The conditional state of the system, given that the result R has the value r , after a measurement of duration T , is

$$|\psi_r(t+T)\rangle = \frac{\hat{M}_r |\psi(t)\rangle}{\sqrt{\wp_r}}. \quad (1.76)$$

As seen above, the probabilities are given by the expectation of another operator, defined in terms of the measurement operators by

$$\hat{E}_r = \hat{M}_r^\dagger \hat{M}_r. \quad (1.77)$$

These operators are known as *probability operators*, or *effects*. The fact that $\sum_r \wp_r$ must equal unity for all initial states gives a *completeness condition* on the measurement operators:

$$\sum_r \hat{E}_r = \hat{1}_S. \quad (1.78)$$

This restriction, that $\{\hat{E}_r; r\}$ be a *resolution of the identity* for the system Hilbert space, is the only restriction on the set of measurement operators (apart from the fact that they must be positive, of course).

The set of all effects $\{\hat{E}_r; r\}$ constitutes an *effect-valued measure* more commonly known as a *probability-operator-valued measure* (POM³) on the space of results r . This simply means that, rather than a probability distribution (or probability-valued measure) over the space of results, we have a probability-operator-valued measure. Note that we have left behind the notion of ‘observables’ in this formulation of measurement. The possible measurement results r are not the eigenvalues of an Hermitian operator representing an observable; they are simply labels representing possible results. Depending on the circumstances, it might be convenient to represent the result R by an integer, a real number, a complex number, or an even more exotic quantity.

If one were making only a single measurement, then the conditioned state $|\psi_r\rangle$ would be irrelevant. However, one often wishes to consider a sequence of measurements, in which case the conditioned system state is vital. In terms of the state matrix ρ , which allows the possibility of mixed initial states, the conditioned state is

$$\rho_r(t+T) = \frac{\mathcal{J}[\hat{M}_r]\rho(t)}{\wp_r}, \quad (1.79)$$

where $\wp_r = \text{Tr}[\rho(t)\hat{E}_r]$ and, for arbitrary operators A and B ,

$$\mathcal{J}[\hat{A}]\hat{B} \equiv \hat{A}\hat{B}\hat{A}^\dagger. \quad (1.80)$$

The *superoperator*

$$\mathcal{O}_r = \mathcal{J}[\hat{M}_r] \quad (1.81)$$

is known as the *operation* for r . It is called a superoperator because it takes an operator (here ρ) to another operator. Operations can be identified with the class of superoperators that take physical states to physical states. (See Box 1.3.) This very important class is also known as *completely positive maps*.

³ The abbreviation POVM is used also, and, in both cases, PO is sometimes understood to denote ‘positive operator’ rather than ‘probability operator’.

Box 1.3 Superoperators and operations

A superoperator \mathcal{S} is an operator on the space of Hilbert-space operators:

$$\hat{A} \rightarrow \hat{A}' = \mathcal{S}\hat{A}. \quad (1.82)$$

A superoperator \mathcal{S} must satisfy three conditions in order to correspond to a physical process (such as measurement or dynamics).

1. \mathcal{S} is trace-preserving or decreasing. That is, $0 \leq \text{Tr}[\mathcal{S}\rho] \leq 1$ for any state ρ . Moreover, $\text{Tr}[\mathcal{S}\rho]$ is the probability that the process occurs.
2. \mathcal{S} is a convex linear map on operators. That is, for probabilities \wp_j , we have that $\mathcal{S} \sum_j \wp_j \rho_j = \sum_j \wp_j \mathcal{S}\rho_j$.
3. \mathcal{S} is *completely positive*. That is, not only does \mathcal{S} map positive operators to positive operators for the system of interest S, but so does $(\mathcal{I} \otimes \mathcal{S})$. Here \mathcal{I} is the identity superoperator for an arbitrary second system R.

The final property deserves some comment. It might have been thought that positivity of a superoperator would be sufficient to represent a physical process. However, it is always possible that a system S is entangled with another system R before the physical process represented by \mathcal{S} acts on system S. It must still be the case that the total state of both systems remains a physical state with a positive state matrix. This gives condition 3.

If a superoperator satisfies these three properties then it is called an *operation*, and has the Kraus representation [Kra83], or *operator sum representation*,

$$\mathcal{S}(\rho) = \sum_j \hat{K}_j \rho \hat{K}_j^\dagger \quad (1.83)$$

for some set of operators \hat{K}_j satisfying

$$\hat{1} - \sum_j \hat{K}_j^\dagger \hat{K}_j \geq 0. \quad (1.84)$$

There is another important representation theorem for operations, which follows from the Gelfand–Naimark–Segal theorem [Con90]. Consider, as above, an apparatus or ancilla system A in addition to the quantum system of interest S. Then there is a pure state $|\theta\rangle_A$ of A and some unitary evolution, \hat{U}_{SA} , describing the coupling of system S to system A, such that

$$\mathcal{S}\rho_S = \text{Tr}_A \left[(\hat{1}_S \otimes \hat{\Pi}_A) \hat{U}_{SA} (\rho_S \otimes |\theta\rangle_A \langle \theta|) \hat{U}_{SA}^\dagger \right], \quad (1.85)$$

where $\hat{\Pi}_A$ is some projector for the ancilla system A. This is essentially the converse of the construction of operations for measurements from a system–apparatus coupling in Section 1.2.3.

If the measurement were performed but the result R ignored, the final state of the system would be

$$\rho(t+T) = \sum_r \mathcal{O}_r \rho(t+T) = \sum_r \mathcal{J}[\hat{M}_r] \rho(t) \equiv \mathcal{O} \rho(t). \quad (1.86)$$

Here \mathcal{O} is also an operation, and is trace-preserving.

For non-projective measurements, there is no guarantee that repeating the measurement will yield the same result. In fact, the final state of the system may be completely unrelated to either the initial state of the system or the result obtained. This is best illustrated by an example.

Example 1. Consider the set of measurement operators $\{\hat{M}_r\}$ defined by $\hat{M}_r = |0\rangle\langle r|$, where $r \in \{0, 1, 2, \dots\}$ and $\{|r\rangle\}$ is a complete basis for the system Hilbert space. Then the effects for the measurements are projectors $\hat{E}_r = \hat{\Pi}_r = |r\rangle\langle r|$, which obviously obey the completeness condition (1.78). The probability of obtaining $R = r$ is just $\langle r|\rho(t)|r\rangle$. However, the final state of the system, regardless of the result r , is $\rho_r(t+T) = |0\rangle\langle 0|$. Lest it be thought that this is an artificial example, it in fact arises very naturally from counting photons. There R is the number of photons, and, because photons are typically absorbed in order to be counted, the number of photons left after the measurement has finished is zero.

In the above example, the effects are still projection operators. However, there are other measurements in which this is not the case.

Example 2. Consider a two-dimensional Hilbert space with the basis $|0\rangle, |1\rangle$. Consider a *continuous* measurement result ϕ that can take values between 0 and 2π . We define the measurement operators $\hat{M}_\phi = |\phi\rangle\langle\phi|/\sqrt{\pi}$, where $|\phi\rangle$ is defined by

$$|\phi\rangle = [|0\rangle + \exp(i\phi)|1\rangle]/\sqrt{2}. \quad (1.87)$$

In this case the effects are

$$\hat{E}_\phi = \frac{1}{\pi} |\phi\rangle\langle\phi|, \quad (1.88)$$

and the completeness condition, which is easy to verify, is

$$\int_0^{2\pi} d\phi \hat{E}_\phi = |0\rangle\langle 0| + |1\rangle\langle 1| = \hat{1}. \quad (1.89)$$

Although \hat{E}_ϕ is proportional to a projection operator it is not equal to one. It does not square to itself: $(\hat{E}_\phi d\phi)^2 = \hat{E}_\phi d\phi(d\phi/\pi)$. Neither are different effects orthogonal in general: $\hat{E}_\phi \hat{E}_{\phi'} \neq 0$ unless $\phi' = \phi + \pi$. Thus, even if the system is initially in the state $|\phi\rangle$, there is a finite probability for any result to be obtained except $\phi + \pi$.

The effects \hat{E}_r need not even be proportional to projectors, as the next example shows.

Example 3. Consider again an infinite-dimensional Hilbert space, but now use the continuous basis $|x\rangle$ (see Section 1.2.2 and Appendix A), for which $\langle x|x'\rangle = \delta(x - x')$. Define an effect

$$\hat{E}_y = \int_{-\infty}^{\infty} dx (2\pi\Delta^2)^{-1/2} \exp[-(y-x)^2/(2\Delta^2)] |x\rangle\langle x|. \quad (1.90)$$

This describes an imprecise measurement of position. It is easy to verify that the effects are not proportional to projectors by showing that \hat{E}_y^2 is not proportional to \hat{E}_y . Nevertheless, they are positive operators and obey the completeness relation

$$\int_{-\infty}^{\infty} dy \hat{E}_y = \hat{1}. \quad (1.91)$$

Exercise 1.15 Verify Eq. (1.91). Also show that these effects can be derived from the measurement model introduced in Exercise 1.14.

The previous examples indicate some of the flexibility that arises from not requiring the effects to be projectors. As mentioned above, another example of the power offered by generalized measurements is the simultaneous measurement of position \hat{X} and momentum \hat{P} . This is possible provided that the two measurement results have a certain amount of error. A simple model for this was first described by Arthurs and Kelly [AK65]. A more abstract description directly in terms of the resulting projection valued measure was given by Holevo [Hol82]. The description given below is based on the discussion in [SM01].

Example 4. The model of Arthurs and Kelly consists of two meters that are allowed to interact instantaneously with the system. The interaction couples one of the meters to position and the other to momentum, encoding the results of the measurement in the final states of the meters. Projective measurements are then made on each of the meter states separately. These measurements can be carried out simultaneously since operators for distinct meters commute. For appropriate meter states, this measurement forces the conditional state of the system into a Gaussian state (defined below). We assume some appropriate length scale such that the positions and momenta for the system are dimensionless, and satisfy $[\hat{X}, \hat{P}] = i$.

The appropriate unitary interaction is

$$\hat{U} = \exp[-i(\hat{X}\hat{P}_1 + \hat{P}\hat{P}_2)]. \quad (1.92)$$

Here the subscripts refer to the two detectors, which are initially in minimum-uncertainty states (see Appendix A) $|d_1\rangle$ and $|d_2\rangle$, respectively. Specifically, we choose the wavefunctions in the position representation to be

$$\langle x_j|d_j\rangle = (\pi/2)^{-1/4} e^{-x_j^2}. \quad (1.93)$$

After the interaction, the detectors are measured in the position basis. The measurement result is thus the pair of numbers (X_1, X_2) . Following the theory given above, the

measurement operator for this result is

$$\hat{M}(x_1, x_2) = \langle x_1 | \langle x_2 | \hat{U} | d_2 \rangle | d_1 \rangle. \quad (1.94)$$

With a little effort it is possible to show that $\hat{M}(x_1, x_2)$ is proportional to a projection operator:

$$\hat{M}(x_1, x_2) = \frac{1}{\sqrt{2\pi}} |(x_1, x_2)\rangle \langle (x_1, x_2)|. \quad (1.95)$$

Here the state $|(x_1, x_2)\rangle$ is a minimum-uncertainty state for the system, with a position probability amplitude distribution

$$\langle x | (x_1, x_2) \rangle = (\pi)^{-1/4} \exp \left[ix_2 - \frac{1}{2}(x - x_1)^2 \right]. \quad (1.96)$$

From Appendix A, this is a state with mean position and momentum given by x_1 and x_2 , respectively, and with the variances in position and momentum equal to $1/2$.

Exercise 1.16 Verify Eq. (1.95).

The corresponding probability density for the observed values, (x_1, x_2) , is found from the effect density

$$\hat{E}(x_1, x_2) dx_1 dx_2 = \frac{1}{2\pi} |(x_1, x_2)\rangle \langle (x_1, x_2)| dx_1 dx_2. \quad (1.97)$$

Exercise 1.17 Show that

$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \hat{E}(x_1, x_2) = \hat{1}. \quad (1.98)$$

From this POM we can show that

$$E[X_1] = \langle \hat{X} \rangle, \quad E[X_1^2] = \langle \hat{X}^2 \rangle + \frac{1}{2}, \quad (1.99)$$

$$E[X_2] = \langle \hat{P} \rangle, \quad E[X_2^2] = \langle \hat{P}^2 \rangle + \frac{1}{2}, \quad (1.100)$$

where $\langle \hat{A} \rangle = \text{Tr}[\hat{A}\rho]$ is the quantum expectation, while E is a classical average computed by evaluating an integral over the probability density $\wp(x_1, x_2)$. Thus the readout variables X_1 and X_2 give, respectively, the position and momentum of the system with additional noise.

It is more conventional to denote the state $|(x_1, x_2)\rangle$ by $|\alpha\rangle$, where the single complex parameter α is given by $\alpha = (x_1 + ix_2)/\sqrt{2}$. In this form the states are known as coherent states (see Section A.4). The corresponding effect density is $\hat{F}(\alpha) = |\alpha\rangle\langle\alpha|/\pi$ and the resulting probability density $\wp(\alpha)d^2\alpha = \text{Tr}[\hat{F}(\alpha)\rho]d^2\alpha$. This is known as the Q -function in quantum optics – see Section A.5. For a general choice of initial pure states for the detectors, the probability density for observed results is known as the Husimi function [Hus40].

1.2.6 Non-selective evolution and choice of basis

Recall that in the analysis above, using system and meter states, the combined state prior to the measurement of the meter was

$$|\Psi(t + T_1)\rangle = \hat{U}(T_1)|\theta(t)\rangle|\psi(t)\rangle. \quad (1.101)$$

As explained there, it is not possible to assign a state vector to the system at time $t + T_1$, because it is entangled with the meter. However, it is possible to assign a state matrix to the system. This state matrix is found by taking the partial trace over the meter:

$$\begin{aligned} \rho(t + T_1) &= \text{Tr}_A[|\Psi(t + T_1)\rangle\langle\Psi(t + T_1)|] \\ &\equiv \sum_j {}_A\langle\phi_j|\Psi(t + T_1)\rangle\langle\Psi(t + T_1)|\phi_j\rangle_A, \end{aligned} \quad (1.102)$$

where $\{|\phi_j\rangle_A: j\}$ is an arbitrary set of basis states for the meter. But this basis can of course be the basis $\{|r\rangle: r\}$ appropriate for a measurement of R on the meter. Thus the reduced system state $\rho(t + T_1)$ is the same as the average system state $\rho(t + T)$ (for $T \geq T_1$) of Eq. (1.86), which is obtained by averaging over the measurement results. That is, the non-selective system state after the measurement does not depend on the basis in which the meter is measured.

Different measurement bases for the meter can be related by a unitary transformation thus:

$$|r\rangle = \sum_s U_{r,s}^* |s\rangle, \quad (1.103)$$

where U is a c-number matrix satisfying $\sum_r U_{r,s} U_{r,q}^* = \delta_{s,q}$. In terms of the measurement operators \hat{M}_s , this amounts to a unitary rearrangement to \hat{M}_r defined by

$$\hat{M}_r = \sum_s U_{r,s} \hat{M}_s. \quad (1.104)$$

Exercise 1.18 Verify that the unconditional final state under the new measurement operators $\{\hat{M}_s\}$ is the same as that under the old measurement operators $\{\hat{M}_r\}$.

The binary example. Although the unconditional system state is the same regardless of how the meter is measured, the conditional system states are quite different. This can be illustrated using the binary measurement example of Section 1.2.4. Consider the simple case in which the fiducial apparatus state is the measurement basis state $|0\rangle_A = |y := 0\rangle$. The measurement basis states are eigenstates of the apparatus operator

$$\hat{Y} = \sum_{y=0}^1 y |y\rangle\langle y|. \quad (1.105)$$

Then, if the apparatus is measured in the measurement basis, the measurement operators are

$$\hat{M}_y = {}_A\langle y|\hat{G}|0\rangle_A = |x := y\rangle\langle x := y|. \quad (1.106)$$

As stated before, these simply project or ‘collapse’ the system into its measurement basis, the eigenstates of

$$\hat{X} = \sum_{x=0}^1 x|x\rangle\langle x|. \quad (1.107)$$

Now consider an alternative orthonormal basis for the apparatus, namely the eigenstates of the complementary operator

$$\hat{P}_A = \sum_{p=0}^1 p|p\rangle_A\langle p|. \quad (1.108)$$

Here the eigenstates for the apparatus are

$$|p\rangle_A = 2^{-1/2}(|y := 0\rangle + e^{i\pi p}|y := 1\rangle), \quad (1.109)$$

and \hat{X} and \hat{P} are complementary in the sense that \hat{X} is maximally uncertain for a system in a \hat{P} -eigenstate, and vice versa. In this case the measurement operators are, in the measurement (x) basis,

$$\hat{M}_p = 2^{-1/2}(|0\rangle\langle 0| + e^{-i\pi p}|1\rangle\langle 1|). \quad (1.110)$$

Exercise 1.19 *Verify that the non-selective evolution is the same under these two different measurements, and that it always turns the system into a mixture diagonal in the measurement basis.*

Clearly, measurement of the apparatus in the complementary basis does not collapse the system into a pure state in the measurement basis. In fact, it does not change the occupation probabilities for the measurement basis states at all. This is because the measurement yields no information about the system, since the probabilities for the two results are independent of the system:

$$\Pr[P_A = p] = \langle\psi(t)|\hat{M}_p^\dagger\hat{M}_p|\psi(t)\rangle = 1/2. \quad (1.111)$$

This ‘measurement’ merely changes the relative phase of these states by π if and only if $p = 1$:

$$\frac{\hat{M}_p \sum_x s_x |x\rangle}{\sqrt{\Pr[P_A = p]}} = \sum_x s_x e^{-i\pi p x} |x\rangle. \quad (1.112)$$

That is to say, with probability 1/2, the relative phase of the system states is flipped. In this guise, the interaction between the system and the apparatus is seen not to collapse the system into a measurement eigenstate, but to introduce noise into a complementary system property: the relative phase.

This dual interpretation of an interaction between a system and another system (the meter) is very common. The non-selective evolution reduces the system to a mixture diagonal in some basis. One interpretation (realized by measuring the meter in an appropriate way) is that the system is collapsed into a particular state in that basis, but an equally valid

interpretation (realized by measuring the meter in a complementary way) is that the meter is merely adding noise into the relative phases of the system components in this basis. In the following section, we will see how both of these interpretations can be seen simultaneously in the Heisenberg picture.

Exercise 1.20 Consider the quantum position-measurement model introduced in Exercise 1.14. Show that, if the apparatus is measured in the momentum basis (see Appendix A), then the measurement operators are

$$\hat{M}_p = \left(\frac{2\Delta^2}{\pi} \right)^{1/4} \exp(-p^2 \Delta^2) \exp(-ip\hat{X}). \quad (1.113)$$

Show also that the non-selective evolution is the same as in Exercise 1.14, and that the selective evolution in this exercise can be described in terms of random momentum kicks on the system.

1.3 Representing outcomes as operators

1.3.1 ‘Correlations without correlata’

We have already met the idea that an operator can represent an outcome in Sections A.1 and 1.2.2, where it was shown that

$$\langle f(\Lambda) \rangle = \text{Tr}[f(\hat{\Lambda})\rho]. \quad (1.114)$$

That is, if an operator $\hat{\Lambda}$ represents an observable Λ , then any function of the result of a measurement of Λ is represented by that function of the operator $\hat{\Lambda}$. Here ρ is the state of the system at the time of the measurements. Clearly, if ρ evolves after the measurement has finished, then the formula (1.114) using this new ρ might no longer give the correct expectation values for the results that had been obtained.

This problem can be circumvented by using the system–meter model of measurement we have presented. Let us assume an entangled system–meter state of the form

$$|\Psi(t + T_1)\rangle = \sum_{\lambda} |\lambda\rangle_A \hat{\Pi}_{\lambda} |\psi(t)\rangle_S, \quad (1.115)$$

where $\{|\lambda\rangle_A: \lambda\}$ is an orthonormal set of apparatus states and $\{\hat{\Pi}_{\lambda}: \lambda\}$ is the set of eigenprojectors of the system observable Λ_S . This is the ideal correlation for the apparatus to ‘measure’ Λ_S . The *apparatus* observable represented by

$$\hat{\Lambda}_A = \sum_{\lambda} \lambda |\lambda\rangle_A \langle \lambda| \quad (1.116)$$

has identical moments to the system observable Λ_S for the original system state $|\psi(t)\rangle$, or indeed for the (mixed) system state at time $t + T_1$ derived from Eq. (1.115).

Exercise 1.21 Show this.

What has been gained by introducing the meter is that $\hat{\Lambda}_A$ will continue to represent the result Λ of the measurement made at time t , for all times in the future, regardless of the system evolution. We require only that the statistics of $\hat{\Lambda}_A$ do not change after the measurement; that is, that Λ_A be a so-called QND (*quantum non-demolition*) observable. Since meter operators by definition commute with system operators, the meter operator is a classical quantity insofar as the system is concerned – a c-number rather than a q-number. For instance, one could consider a Hamiltonian, acting some time after the measurement, of the form

$$\hat{H} = \hat{\Lambda}_A \otimes \hat{F}_S, \quad (1.117)$$

where \hat{F}_S is an Hermitian system operator, and not have to worry about the operator ordering. In fact, insofar as the system is concerned, this Hamiltonian is equivalent to the Hamiltonian

$$\hat{H} = \Lambda \hat{F}_S, \quad (1.118)$$

where here Λ is the measurement result (a random variable) obtained in the projective measurement of the system at time t .

Exercise 1.22 *Convince yourself of this.*

The action of Hamiltonians such as these (a form of feedback) will be considered in greater detail in later chapters.

This idea of representing measurement results by meter operators is not limited to projective measurements of the system. Say one has the entangled state between system and meter

$$|\Psi(t + T_1)\rangle = \hat{U}(T_1)|\theta\rangle_A |\psi(t)\rangle_S, \quad (1.119)$$

and one measures the meter in the (assumed non-degenerate) eigenbasis $\{|r\rangle_A\}$ of the operator

$$\hat{R}_A = \sum_r r |r\rangle_A \langle r|. \quad (1.120)$$

Then the operator \hat{R}_A represents the outcome of the measurement that, for the system, is described using the measurement operators $\hat{M}_r = \langle r | \hat{U}(T_1) | \theta \rangle$. Recall that the results r are just labels, which need not be real numbers, so \hat{R}_A is not necessarily an Hermitian operator. If the result R is a complex number, then \hat{R}_A is a *normal* operator (see Box 1.1). If R is a real vector, then \hat{R}_A is a vector of commuting Hermitian operators.

It is important to note that \hat{R}_A represents the measurement outcome *whether or not* the projective measurement of the apparatus is made. That is, it is possible to represent a measurement outcome simply by modelling the apparatus, without including the extra step of apparatus state collapse. In this sense, the von Neumann chain can be avoided, not by placing the Heisenberg cut between apparatus and higher links (towards the observer's consciousness), but by ignoring these higher links altogether. The price to be paid for this

parsimony is a high one: the loss of any notion of *actual* outcomes. The measurement result R remains a random variable (represented by the operator \hat{R}_A) that never takes any particular one of its ‘possible’ values r . Within this philosophical viewpoint one denies the existence of events, but nevertheless calculates their statistics; in other words, ‘correlations without correlata’ [Mer98].

1.3.2 Measurement in the Heisenberg picture

The ‘measurement without collapse’ formulation outlined above is obviously the ideal one for working in the Heisenberg picture. Recall from Section A.1.3 that, in the Heisenberg picture, the state vector or state matrix is constant, while operators evolve in time. However, this was formulated only for unitary evolution; if one wishes to describe a measurement for which some particular result is obtained, this can be done only by invoking state collapse. That is, one must still allow the state to change, even though one is working in the Heisenberg picture. But if one is content to describe a measurement simply as the coupling of the system to the meter, with the result being represented by a meter operator, then state collapse never occurs. Consequently, it is possible to describe all evolution, including measurement, in terms of changing operators. Of course, to do this, one needs to consider system and apparatus operators, not just system operators.

The necessity of using apparatus operators might not be obvious to the reader. After all, when considering unitary evolution of the system alone, we can use essentially the same transformation in the Schrödinger and Heisenberg pictures: $\rho \rightarrow \hat{U}\rho\hat{U}^\dagger$ and $\hat{O} \rightarrow \hat{U}^\dagger\hat{O}\hat{U}$, respectively. This suggests that for measurement the analogue of $\rho \rightarrow \rho'_r = \hat{M}_r\rho\hat{M}_r^\dagger$ would be $\hat{O} \rightarrow \hat{O}'_r = \hat{M}_r^\dagger\hat{O}\hat{M}_r$. However, this construction does not work when one considers operator products. The correct post-measurement expectation for $\hat{A}\hat{B}$, weighted by the probability for outcome r , is

$$\text{Tr}[\hat{A}\hat{B}\rho'_r] = \text{Tr}[\hat{M}_r^\dagger\hat{A}\hat{B}\hat{M}_r\rho]. \quad (1.121)$$

In general, this is quite different from

$$\text{Tr}[\hat{A}'_r\hat{B}'_r\rho] = \text{Tr}[\hat{M}_r^\dagger\hat{A}\hat{M}_r\hat{M}_r^\dagger\hat{B}\hat{M}_r\rho], \quad (1.122)$$

because, in general, \hat{M}_r is not unitary.

The correct Heisenberg formulation of measurement is as follows. The total state (of system plus apparatus) remains equal to the initial state, which is usually taken to factorize as

$$\rho_{\text{total}}(t) = \rho_S \otimes \rho_A. \quad (1.123)$$

The measurement outcome is described, as above, by the apparatus operator

$$\hat{R}_A(t) = \sum_r r|r\rangle_A\langle r|, \quad (1.124)$$

which here is for time t , before the measurement interaction between system and apparatus. This interaction, of duration T_1 , changes \hat{R}_A to

$$\hat{R}_A(t + T) = \hat{U}^\dagger(T_1)[\hat{R}_A(t) \otimes \hat{I}_S] \hat{U}(T_1) \quad (1.125)$$

$$= \sum_r \hat{U}^\dagger(T_1)(|r\rangle_A \langle r| \otimes \hat{I}_S) \hat{U}(T_1). \quad (1.126)$$

Here T is any time greater than or equal to T_1 , since we are assuming that the measurement interaction ceases at time $t + T_1$ and that \hat{R}_A is a QND observable for all subsequent evolution of the meter.

It follows trivially from the analysis of Section A.1.3 that the Heisenberg-picture operator $\hat{R}_A(t + T)$ with respect to $\rho_{\text{total}}(t)$ has the same statistics as does the Schrödinger-picture operator with respect to $\rho_{\text{total}}(t + T)$, evolved according to the measurement interaction. Hence, if the initial apparatus state is pure,

$$\rho_A = |\theta\rangle_A \langle \theta|, \quad (1.127)$$

as we assumed, then these statistics are identical to those of the random variable R_A , the result of a measurement on the system with measurement operators $\{\hat{M}_r\}$.

Being an apparatus operator, $\hat{R}_A(s)$ commutes with system operators at all times s . For $s \leq t$ (that is, before the system and apparatus interact), it is also uncorrelated with all system operators. That is, for $s \leq t$, expectation values factorize:

$$\langle \hat{O}_S(t) f(\hat{R}_A(t)) \rangle = \langle \hat{O}_S(t) \rangle \langle f(\hat{R}_A(t)) \rangle. \quad (1.128)$$

Here \hat{O}_S is an arbitrary system operator and f is an arbitrary function. For $s > t$, this is no longer true. In particular, for $s = t + T$ the correlation with the system is the same as one would calculate using state collapse, namely

$$\langle \hat{O}_S(t + T) f(\hat{R}_A(t + T)) \rangle = \sum_r \wp_r f(r) \text{Tr}[\hat{O}_S \rho_r(t + T)], \quad (1.129)$$

where $\rho_r(t + T)$ is the a-posteriori conditioned system state.

Exercise 1.23 *Convince yourself of this.*

It should be noted that these two descriptions of measurement, in terms of changing operators or changing states, have classical equivalents. They are descriptions in terms of changing system variables or changing probability distributions for these variables. We have already used these two descriptions in Section 1.1. Specifically, we began with the ‘Heisenberg’ description, with correlations arising between system and apparatus variables, and then moved to the complementary ‘Schrödinger’ description with system state collapse derived using Bayes’ theorem.

In the Heisenberg picture, an important difference between quantum and classical measurement stands out. The back-action of the measurement on the system is seen in changes in the system operators, rather than changes in the system state. Classically, a non-disturbing measurement does not introduce any noise into the system. Hence classically there may be

no change in the system variables, but in the quantum case any measurement will necessarily cause changes to the system operators. This quantum back-action is best illustrated by example, as we will do in the next subsection. The same distinction between quantum and classical mechanics is also present in the Schrödinger picture, but only in the weaker form given in Exercise 1.9.

1.3.3 The binary example

To illustrate the description of measurement in the Heisenberg picture, we use again the example of a binary measurement. Rather than using \hat{X} for the system and \hat{Y} for the apparatus, we use \hat{X}_S and \hat{X}_A . Similarly, we use \hat{P}_S and \hat{P}_A for the complementary operators. These are defined by the relation between the eigenstates $|x\rangle$ and $|p\rangle$ defined in Section 1.2.6. The operators \hat{X} and \hat{P} each act as a displacement in the complementary basis, by which we mean that, for binary variables k and n ,

$$\exp(i\pi k \hat{X})|p\rangle = |p \oplus k\rangle, \quad (1.130)$$

$$\exp(i\pi n \hat{P})|x\rangle = |x \oplus n\rangle. \quad (1.131)$$

It is now easy to see that the measurement interaction between the system and the apparatus may be realized by

$$\hat{G} = \exp(i\pi \hat{X}_S \otimes \hat{P}_A). \quad (1.132)$$

Exercise 1.24 Show that this does produce Eq. (1.63).

In the Heisenberg picture, this unitary operator transforms the operators according to $\hat{O}(t + T_1) = \hat{G}^\dagger \hat{O}(t) \hat{G}$, where \hat{O} is an arbitrary operator. Thus we find

$$\hat{X}_S(t + T_1) = \hat{X}_S(t), \quad (1.133)$$

$$\hat{P}_S(t + T_1) = \hat{P}_S(t) \oplus \hat{P}_A(t), \quad (1.134)$$

$$\hat{X}_A(t + T_1) = \hat{X}_S(t) \oplus \hat{X}_A(t), \quad (1.135)$$

$$\hat{P}_A(t + T_1) = \hat{P}_A(t). \quad (1.136)$$

The binary addition \oplus is defined for operators by, for example,

$$\hat{X}_S \oplus \hat{X}_A = \sum_{x,y} (x \oplus y) |x\rangle_S \langle x| \otimes |y\rangle_A \langle y|. \quad (1.137)$$

If we make the identifications

$$X = X_S, \quad Y = X_A(t + T_1), \quad \Xi = X_A(t), \quad (1.138)$$

then Eq. (1.135) is identical in form and content to the classical Eq. (1.2). The noise term is seen to arise from the initial apparatus state. Note that \hat{X}_S is unchanged by the interaction. This quantity is a QND variable and the measurement interaction realizes a QND measurement of \hat{X}_S . However, unlike in the classical case, the system is affected by

the measurement. This quantum back-action is seen in the change in the complementary system quantity, \hat{P}_S , in Eq. (1.134). The ‘quantum noise’ added to the system here is \hat{P}_A , which is another QND variable. Clearly, if one were to measure \hat{P}_A , one would gain no information about the system. (Indeed, one gains most information about the system by measuring the apparatus in the \hat{X}_A basis, which is a basis complementary to the \hat{P}_A basis). However, by measuring \hat{P}_A , one directly finds out the noise that has affected the system, as discussed in Section 1.2.6. We see now that, in the Heisenberg picture, both interpretations of the interaction, namely in terms of gaining information about the system and in terms of adding noise to the system, can be seen simultaneously.

Exercise 1.25 Analyse the case of a generalized position measurement from Exercise 1.14 in the same manner as the binary example of this subsection.

Hint: First show that \hat{P} generates displacements of position \hat{X} and vice versa. That is, $e^{iq\hat{P}}|x\rangle = |x+q\rangle$ and $e^{-ik\hat{X}}|p\rangle = |p+k\rangle$ (see Section A.3). Note that, unlike in the binary example, there is no π in the exponential.

1.4 Most general formulation of quantum measurements

1.4.1 Operations and effects

The theory of measurements we have presented thus far is not quite the most general, but can easily be generalized to make it so. This generalization is necessary to deal with some cases in which the initial meter state is not pure, or the measurement on the meter is not a von Neumann measurement. In such cases the conditioned system state may be impure, even if the initial system state was pure. We call these *inefficient measurements*.

To give the most general formulation⁴ we must dispense with the measurement operators \hat{M}_r and use only operations and effects. The operation \mathcal{O}_r for the result r is a completely positive superoperator (see Box 1.3), not restricted to the form of Eq. (1.81). It can nevertheless be shown that an operation can always be written as

$$\mathcal{O}_r = \sum_j \mathcal{J}[\hat{\Omega}_{r,j}], \quad (1.139)$$

for some set of operators $\{\hat{\Omega}_{r,j}: j\}$.

For a given operation \mathcal{O}_r , the set $\{\hat{\Omega}_{r,j}: j\}$ is not unique. For this reason it would be wrong to think of the operators $\hat{\Omega}_{r,j}$ as measurement operators. Rather, the operation is the basic element in this theory, which takes the a-priori system state to the conditioned a-posteriori state:

$$\tilde{\rho}_r(t+T) = \mathcal{O}_r \rho(t). \quad (1.140)$$

⁴ It is possible to be even more general by allowing the apparatus to be initially correlated with the system. We do not consider this situation because it removes an essential distinction between apparatus and system, namely that the former is in a fiducial state known to the experimenter, while the latter can be in an arbitrary state (perhaps known to a different experimenter). If the two are initially correlated they should be considered jointly as the system.

The state in Eq. (1.140) is unnormalized. Its norm is the probability \wp_r for obtaining the result $R = r$,

$$\wp_r = \text{Tr}[\mathcal{O}_r \rho(t)], \quad (1.141)$$

so that the normalized state is

$$\rho_r(t+T) = \mathcal{O}_r \rho(t) / \wp_r. \quad (1.142)$$

As for efficient measurements in Section 1.2.5, it is possible to define a *probability operator*, or *effect*, \hat{E}_r , such that, for all ρ ,

$$\text{Tr}[\mathcal{O}_r \rho] = \text{Tr}[\rho \hat{E}_r]. \quad (1.143)$$

It is easy to verify that

$$\hat{E}_r = \sum_j \hat{\Omega}_{r,j}^\dagger \hat{\Omega}_{r,j}, \quad (1.144)$$

which is obviously Hermitian and positive. The completeness condition

$$\sum_r \hat{E}_r = \hat{1} \quad (1.145)$$

is the only mathematical restriction on the set of operations \mathcal{O}_r .

The unconditional system state after the measurement is

$$\rho(t+T) = \sum_r \mathcal{O}_r \rho(t) = \mathcal{O} \rho(t). \quad (1.146)$$

Here the non-selective operation can be written

$$\mathcal{O} = \sum_{r,j} \mathcal{J}[\hat{\Omega}_{r,j}]. \quad (1.147)$$

In terms of the unitary operator $\hat{U}(T_1)$ coupling system to apparatus, this operation can also be defined by

$$\mathcal{O} \rho \equiv \text{Tr}_A[\hat{U}(T_1)(\rho \otimes \rho_A)\hat{U}^\dagger(T_1)], \quad (1.148)$$

where ρ_A is the initial apparatus state matrix.

Exercise 1.26 By decomposing ρ_A into an ensemble of pure states, and considering an apparatus basis $\{|r\rangle\}$, derive an expression for $\hat{\Omega}_{r,j}$. Also show the non-uniqueness of the set $\{\hat{\Omega}_{r,j}; j\}$.

This completes our formal description of quantum measurement theory. Note that the above formulae, from Eq. (1.139) to Eq. (1.147), are exact analogues of the classical formulae from Eq. (1.26) to Eq. (1.34). The most general formulation of classical measurement was achieved simply by adding back-action to Bayes' theorem. The most general formulation of quantum measurement should thus be regarded as the quantum generalization of

Table 1.1. *Quantum measurement theory as generalized Bayesian analysis*

Concept	Quantum formula	Bayesian formula
Initial state	$\rho(t)$, a positive operator	$\wp(t)$, a positive vector
such that	$\text{Tr}[\rho(t)] = 1$	$\sum_x \wp(x; t) = 1$
Measurement result	R , a random variable	R , a random variable
For each r define	an operation \mathcal{O}_r	a matrix \mathcal{O}_r
such that	$\tilde{\rho}_r(t+T) = \mathcal{O}_r \rho(t) \geq 0$	$\tilde{\wp}_r(t+T) = \mathcal{O}_r \wp(t) \geq 0$
$\text{Pr}[R=r]$	$\wp(r) = \text{Tr}[\tilde{\rho}_r(t+T)]$	$\wp(r) = \sum_x \tilde{\wp}_r(x; t+T)$
can be written as	$\wp(r) = \text{Tr}[\rho(t) \hat{E}_r]$	$\wp(r) = \sum_x \wp(x; t) E_r(x)$
where	$\sum_r \hat{E}_r = \hat{1}$	$\forall x, \sum_r E_r(x) = 1$
Conditioned state	$\rho_r(t+T) = \tilde{\rho}_r(t+T)/\wp(r)$	$\wp_r(t+T) = \tilde{\wp}_r(t+T)/\wp(r)$
Interpretation	a matter of debate!	Bayes' rule: $E_r(x) = \wp(r x)$

Bayes' theorem, in which back-action is an inseparable part of the measurement. This difference arises simply from the fact that a quantum state is represented by a positive matrix, whereas a classical state is represented by a positive vector (i.e. a vector of probabilities). This analogy is summarized in Table 1.1.

We now give a final example to show how generalized measurements such as these arise in practice, and why the terminology *inefficient* is appropriate for those measurements for which measurement operators cannot be employed. It is based on Example 1 in Section 1.2.5, which is a description of *efficient* photon counting if $|n\rangle$ is interpreted as the state with n photons.

Say one has an inefficient photon detector, which has only a probability η of detecting each photon. If the perfect detector would detect n photons, then, from the binomial expansion, the imperfect detector would detect r photons with probability

$$\wp(r|n) = \eta^r (1 - \eta)^{n-r} \binom{n}{r}. \quad (1.149)$$

Thus, if r photons are counted at the end of the measurement, the probability that n photons 'would have been' counted by the perfect detector is, by Bayes' theorem,

$$\wp(n|r) = \frac{\wp(r|n) \langle n | \rho(t) | n \rangle}{\sum_m \wp(r|m) \langle m | \rho(t) | m \rangle}. \quad (1.150)$$

Hence, the conditioned system state is the mixture

$$\rho_r(t+T) = \sum_n \wp(n|r) \frac{\mathcal{J}[|0\rangle \langle n|] \rho(t)}{\langle n | \rho(t) | n \rangle} \quad (1.151)$$

$$= \sum_n \frac{\wp(r|n) \mathcal{J}[|0\rangle \langle n|] \rho(t)}{\sum_m \wp(r|m) \langle m | \rho(t) | m \rangle} \quad (1.152)$$

$$= \frac{\mathcal{O}_r \rho(t)}{\text{Tr}[\rho(t) \hat{E}_r]}, \quad (1.153)$$

Table 1.2. *Eight useful or interesting classes of quantum measurements*

Symbol	Name	Definition
E	Efficient	$\forall r, \exists \hat{M}_r, \mathcal{O}_r = \mathcal{J}[\hat{M}_r]$
C	Complete	$\forall \rho, \forall r, \mathcal{O}_r \rho \propto \mathcal{O}_r \hat{1}$
S	Sharp	$\forall r, \text{rank}(\hat{E}_r) = 1$
O	Of an observable X	$\forall r, \hat{E}_r = E_r(\hat{X})$
BAE	Back-action-evading	O with $\forall \rho, \forall x \in \lambda(\hat{X}), \text{Tr}[\hat{\Pi}_x \rho] = \text{Tr}[\hat{\Pi}_x \mathcal{O} \rho]$
MD	Minimally disturbing	E with $\forall r, \hat{M}_r = \hat{M}_r^\dagger$
P	Projective	MD and O
VN	von Neumann	P and S

where the operations and effects are, respectively,

$$\mathcal{O}_r = \sum_n \eta^r (1 - \eta)^{n-r} \binom{n}{r} \mathcal{J}[|0\rangle\langle n|], \quad (1.154)$$

$$\hat{E}_r = \sum_n \eta^r (1 - \eta)^{n-r} \binom{n}{r} |n\rangle\langle n|. \quad (1.155)$$

1.4.2 Classification of measurements

The formalism of operations and effects encompasses an enormous, even bewildering, variety of measurements. By placing restrictions on the operations, different classes of measurements may be defined. In this section, we review some of these classes and their relation to one another. We restrict our consideration to eight classes, identified and defined in Table 1.2. Their complicated inter-relations are defined graphically by the Venn diagram in Fig. 1.2.

Some classes of measurement are characterized by the disturbance imposed on the system by the measurement ('efficient', 'complete', and 'minimally disturbing'). Others are characterized by the sort of information the measurement yields ('sharp' and 'of an observable X '), and so can be defined using the effects only. The remainder are characterized by both the sort of information obtained and the disturbance of the system ('back-action-evading', 'projective', and 'von Neumann').

Some of these classes are well known (such as back-action-evading measurements) while others are not (such as complete measurements). Below, we briefly discuss each of the eight. This also allows us to discuss various concepts relevant to quantum measurement theory.

[E]: Efficient measurements. As already discussed, efficient measurements are ones for which each operation is defined in terms of a measurement operator: $\mathcal{O}_r = \mathcal{J}[\hat{M}_r]$. These measurements take pure states to pure states. Any noise in efficient measurements can be interpreted as quantum noise. The complementary set is that of *inefficient* measurements, which introduce classical noise or uncertainty into the measurement.

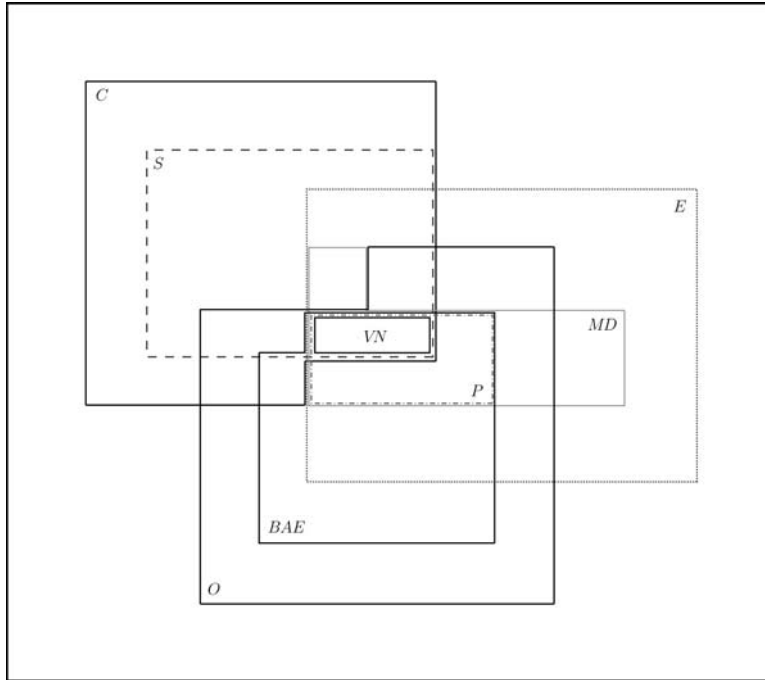


Fig. 1.2 A Venn diagram for the eight classes of quantum measurements described in Table 1.2.

It is only for the class of efficient measurements that one can derive the following powerful theorem [Nie01, FJ01]:

$$H[\rho(t)] \geq \sum_r \wp_r H[\rho_r(t+T)]. \quad (1.156)$$

Here, $H[\rho]$ is any measure of the mixedness of ρ that is invariant under unitary transformations of ρ and satisfies

$$H[w_1\rho_1 + w_2\rho_2] \geq w_1 H[\rho_1] + w_2 H[\rho_2] \quad (1.157)$$

for arbitrary state matrices ρ_j and positive weights w_j summing to unity. Examples of such measures are the entropy $-\text{Tr}[\rho \log \rho]$ and the ‘linear entropy’ $1 - \text{Tr}[\rho^2]$.⁵ The interpretation of this theorem is that, as long as no classical noise is introduced in the measurement, the a-posteriori conditional state is *on average* less mixed than (or just as mixed as) the a-priori state. That is, the measurement refines one’s knowledge of the system, as one would hope. Note that it is *not* true that the conditional a-posteriori state is *always* less mixed than the a-priori state.

⁵ An even stronger version of this theorem, using majorization to classify the relative mixedness of two states, has also been proven [Nie01, FJ01].

Exercise 1.27 Prove the foregoing statement, by finding an example for a binary system.

Hint: This is a classical phenomenon, so the measurement operators and state matrix in the example can all be diagonal in the same basis.

[C]: *Complete measurements.* The definition of complete measurements in Table 1.2 implies that, for all results r , the conditioned a-posteriori state

$$\rho_r(t+T) = \mathcal{O}_r \rho(t) / \text{Tr}[\hat{E}_r \rho(t)] \quad (1.158)$$

is independent of $\rho(t)$. In other words, at the end of the measurement, no information remains in the system about its initial state. This is the sense in which the measurement is *complete*: no further measurements could yield any more information about the initial system state.

The definition of complete measurements implies that the operations must be of the form

$$\mathcal{O}_r = \sum_{j,k} \mathcal{J}[|\theta_{rk}\rangle\langle\phi_{rj}|], \quad (1.159)$$

where θ and ϕ denote (possibly unnormalized) system states. From this, it is easy to see that the conditioned state, independently of $\rho(t)$, is

$$\rho_r(t+T) = \frac{\sum_k |\theta_{rk}\rangle\langle\theta_{rk}|}{\sum_k \langle\theta_{rk}|\theta_{rk}\rangle}. \quad (1.160)$$

The concept of complete measurements (or, more particularly, ‘incomplete measurements’) will be seen to be very useful when discussing *adaptive measurements* in Section 2.5.

[S]: *Sharp measurements.* The definition of sharp measurements in Table 1.2 implies that the effects are rank-1 positive operators. That is to say, each effect is of the form $\hat{E}_r = |\phi_r\rangle\langle\phi_r|$, for some (possibly unnormalized) state $|\phi_r\rangle$. This implies that the operations must be of the form

$$\mathcal{O}_r = \sum_k \mathcal{J}[|\theta_{rk}\rangle\langle\phi_r|]. \quad (1.161)$$

From this it is apparent that sharp measurements are a subclass of complete measurements. Also, it is apparent that, for efficient measurements, sharpness and completeness are identical properties.

The significance of sharpness is that a sharp measurement cannot be an unsharp version of a different measurement [MdM90a, MdM90b]. That is, the results of a sharp measurement cannot be generated by making a different measurement and then rendering it ‘unsharp’ by classically processing the results. Mathematically, a sharp measurement $\{\hat{E}_r\}$ is one for which there is no other measurement $\{\hat{E}'_s\}$ such that

$$\hat{E}_r = \sum_s w_{r|s} \hat{E}'_s, \quad (1.162)$$

where $w_{r|s}$ is the probability that r is reported as the measurement result when the second measurement result is s . The object $\{w_{r|s}\}$ is sometimes called a stochastic map from $\{s\}$ to $\{r\}$. We also require that this stochastic map $\{w_{r|s}\}$ be nontrivial. A trivial stochastic map is a deterministic one for which $w_{r|s}^2 = w_{r|s}$ for all r and s , which simply relabels measurement results.

Another fact about sharp measurements is that it is always possible to prepare the system in a state such that a given result r *cannot* be obtained. Note, however, that there is no requirement that the effects be orthogonal, so it is not necessarily possible to prepare the system such that a given result r is guaranteed.

[O]: *Measurements of an observable.* If an effect \hat{E}_r is a function of an Hermitian operator \hat{X} , then the probability of obtaining the result r is given by

$$\wp_r = \text{Tr}[E_r(\hat{X})\rho(t)] = \sum_x E_r(x) \text{Tr}[\hat{\Pi}_x \rho(t)], \quad (1.163)$$

where $\{x\}$ are the (assumed discrete for simplicity) eigenvalues of \hat{X} and $\hat{\Pi}_x$ the corresponding projectors. If all of the effects are functions of the same operator \hat{X} , then it is evident that the measurement is equivalent to a (possibly unsharp) measurement of the observable X . That is, the result R could be obtained by making a projective measurement of X and then processing the result. Note that this definition places no restriction on the state of the system after the measurement.

The class labelled O in Fig. 1.2 should be understood to be the class of measurements that are measurements of some observable X . Note that, by virtue of the definition here, a measurement in this class may be a measurement of more than one observable. For example, it is obvious from the above definition that any measurement of X^2 is also a measurement of X . However, if \hat{X} has eigenvalues of equal magnitude but opposite sign, then the converse is not true. This is because, for example, it is not possible to write the effects for a projective measurement of \hat{X} , which are

$$\hat{E}_x = |x\rangle\langle x| = \delta_{\hat{X},x}, \quad (1.164)$$

as a function of $\hat{S} = \hat{X}^2$. This is the case even though the projectors for the latter *are* functions of \hat{X} :

$$\hat{E}_s = \sum_x \delta_{x^2,s} |x\rangle\langle x| = \delta_{\hat{X}^2,s}. \quad (1.165)$$

By binning results (corresponding to values of X with the same magnitude), one can convert the measurement of X into a measurement of X^2 . However, it is not permissible to allow such binning in the above definition, because then every measurement would be a measurement of any observable; simply binning all the results together gives a single $\hat{E} = \hat{1}$, which can be written as a (trivial) function of any observable.

[BAE]: *Back-action-evading measurements.* Consider a measurement of an observable X according to the above definition. A hypothetical projective measurement of X before this measurement will not affect the results of this measurement, because the effects are

a function of \hat{X} . However, the converse is not necessarily true. Because the definition of ‘measurement of an observable X ’ is formulated in terms of the effects alone, it takes no account of the disturbance or *back-action* of the measurement on the system. A back-action-evading (BAE) measurement of X is one for which a projective measurement of X *after* the measurement will have the same statistics as one *before*. If the total (i.e. non-selective) operation for the measurement in question is $\mathcal{O} = \sum_r \mathcal{O}_r$, then the requirement is that, for all ρ and all eigenvalues x of \hat{X} ,

$$\text{Tr}[\hat{\Pi}_x \rho] = \text{Tr}[\hat{\Pi}_x \mathcal{O} \rho]. \quad (1.166)$$

This is the condition in Table 1.2, where we use $\lambda(\hat{X})$ to denote the set of eigenvalues of \hat{X} .

A concept closely related to BAE measurement is QND measurement. Recall from Section 1.3 that X is a QND (quantum non-demolition) observable if the operator \hat{X} is a constant of motion (in the Heisenberg picture). Thus, we can talk of a QND measurement of \hat{X} if the effects are functions of \hat{X} and

$$\hat{X} = \hat{U}^\dagger(T_1) \hat{X} \hat{U}(T_1), \quad (1.167)$$

where $\hat{U}(T_1)$ is the unitary operator describing the coupling of the system to the meter, as in Section 1.3.2, so that \hat{X} is to be understood as $\hat{X}_S \otimes \hat{I}_A$.

The condition for a back-action-evading measurement (1.166) is implied by (and hence is weaker than) that for a quantum non-demolition measurement. To see this, first note that a unitary transformation preserves eigenvalues, so that Eq. (1.167) implies that, for all x ,

$$\hat{\Pi}_x \otimes \hat{I}_A = \hat{U}^\dagger(T_1) (\hat{\Pi}_x \otimes \hat{I}_A) \hat{U}(T_1). \quad (1.168)$$

Now post-multiply both sides of Eq. (1.168) by $\rho \otimes \rho_A$, where ρ_A is the initial apparatus state. This gives

$$(\hat{\Pi}_x \rho) \otimes \rho_A = \hat{U}^\dagger(T_1) (\hat{\Pi}_x \otimes \hat{I}_A) \hat{U}(T_1) (\rho \otimes \rho_A). \quad (1.169)$$

Now pre- and post-multiply by $\hat{U}(T_1)$ and $\hat{U}^\dagger(T_1)$, respectively. This gives

$$\hat{U}(T_1) [(\hat{\Pi}_x \rho) \otimes \rho_A] \hat{U}^\dagger(T_1) = (\hat{\Pi}_x \otimes \hat{I}_A) \hat{U}(T_1) (\rho \otimes \rho_A) \hat{U}^\dagger(T_1). \quad (1.170)$$

Taking the total trace of both sides then yields Eq. (1.166), from the result in Eq. (1.148).

Often the terms back-action-evading (BAE) measurement and quantum non-demolition (QND) measurement are used interchangeably, and indeed the authors are not aware of any proposal for a BAE measurement that is not also a QND measurement. The advantage of the BAE definition given above is that it is formulated in terms of the operations and effects, as we required.

It is important not to confuse the non-selective and selective a-posteriori states. The motivating definition (1.166) is formulated in terms of the non-selective total operation \mathcal{O} . The definition would be silly if we were to replace this by the selective operation \mathcal{O}_r (even if an appropriate normalizing factor were included). That is because, if the system were prepared in a state with a non-zero variance in X , then the measurement would in general collapse the state of the system into a new state with a smaller variance for X .

That is, the statistics of X would not remain the same. The actual definition ensures that *on average* (that is, ignoring the measurement results) the statistics for X are the same after the measurement as before.

[MD]: *Minimally disturbing measurements*. Minimally disturbing measurements are a subclass of efficient measurements. The *polar decomposition theorem* says that an arbitrary operator, such as the measurement operator \hat{M}_r , can be decomposed as

$$\hat{M}_r = \hat{U}_r \hat{V}_r, \quad (1.171)$$

where \hat{U}_r is unitary and $\hat{V}_r = \sqrt{\hat{E}_r}$ is Hermitian and positive. We can interpret these two operators as follows. The Hermitian \hat{V}_r is responsible for generating the necessary back-action (the ‘state collapse’) associated with the information gained in obtaining the result r (since the statistics of the results are determined solely by \hat{E}_r , and hence solely by \hat{V}_r). The unitary \hat{U}_r represents surplus back-action: an extra unitary transformation independent of the state.

A minimally disturbing measurement is one for which \hat{U}_r is (up to an irrelevant phase factor) the identity. That is,

$$\hat{M}_r = \sqrt{\hat{E}_r}, \quad (1.172)$$

so that the only disturbance of the system is the necessary back-action determined by the probability operators \hat{E}_r . The name ‘minimally disturbing’ can be justified rigorously as follows. The fidelity between an a-priori state of maximal knowledge $|\psi\rangle$ and the a-posteriori state $\tilde{\rho}_r = \mathcal{O}_r |\psi\rangle\langle\psi|$, averaged over r and ψ , is

$$F_{\text{average}} = \int d\mu_{\text{Haar}}(\psi) \sum_r \langle\psi|\tilde{\rho}_r|\psi\rangle. \quad (1.173)$$

Here $d\mu_{\text{Haar}}(\psi)$ is the Haar measure over pure states, the unique measure which is invariant under unitary transformations. For a given POM $\{\hat{E}_r\}$, this is maximized for efficient measurements with measurement operators given by Eq. (1.172) [Ban01].

Exercise 1.28 Show that, for a given POM and a particular initial state ψ , a minimally disturbing measurement (as defined here) is in general not the one which maximizes the fidelity between a-priori and a-posteriori states.

Hint: Consider a QND measurement of $\hat{\sigma}_z$ on a state $|\psi\rangle = \alpha|\sigma_z := -1\rangle + \beta|\sigma_z := 1\rangle$. Compare this with the non-QND measurement of $\hat{\sigma}_z$ with measurement operators $|\psi\rangle\langle\sigma_z := -1|$ and $|\psi\rangle\langle\sigma_z := 1|$.

For minimally disturbing measurements, it is possible to complement the relation (1.156) by the following equally powerful theorem:

$$H[\rho(t+T)] \geq H[\rho(t)], \quad (1.174)$$

where $\rho(t+T) = \sum_r \wp_r \rho_r(t+T)$. That is, the unconditional a-posteriori state is at least as mixed as the a-priori state – if one does not take note of the measurement result,

one's information about the system can only decrease. This does not hold for measurements in general; for the measurement in Example 1, the a-posteriori state is the pure state $|0\rangle$ regardless of the a-priori state. However, it does hold for a slightly broader class than minimally disturbing measurements, namely measurements in which the surplus back-action U_r in Eq. (1.171) is the same for all r . These can be thought of as minimally disturbing measurements followed by a period of unitary evolution.

A minimally disturbing measurement of an observable X is a BAE measurement of that observable, but, of course, minimally disturbing measurements are not restricted to measurements of observables. Finally, it is an interesting fact that the class of minimally disturbing measurements does not have the property of closure. Closure of a class means that, if an arbitrary measurement in a class is followed by another measurement from the same class, the 'total' measurement (with a two-fold result) is guaranteed to be still a member of that class.

Exercise 1.29 Find an example that illustrates the lack of closure for the MD class.

[P]: *Projective measurements*. These are the measurements with which we began our discussion of quantum measurements in Section 1.2.2. They are sometimes referred to as orthodox measurements, and as Type I measurements (all other measurements being Type II) [Pau80]. From the definition that they are minimally disturbing and a measurement of an observable, it follows that the measurement operators \hat{M}_r and effects \hat{E}_r are identical and equal to projectors $\hat{\Pi}_r$.

[VN]: *Von Neumann measurements*. Sometimes the term 'von Neumann measurement' is used synonymously with the term 'projective measurements'. We reserve the term for sharp projective measurements (that is, those with rank-1 projectors). This is because von Neumann actually got the projection postulate wrong for projectors of rank greater than 1, as was pointed out (and corrected) by Lüders [Lüd51]. Von Neumann measurements are the only measurements which are members of all of the above classes.

1.4.3 Classification exercise

Appreciating the relations among the above classes of measurements requires a careful study of Fig. 1.2. To assist the reader in this study, we here provide a prolonged exercise. The Venn diagram in Fig. 1.2 has 17 disjoint regions. If there were no relations among the eight classes, there would be 2^8 , that is 256, regions. Thus the fact that there are only 17 testifies to the many inter-relationships among classes.

Below, we have listed 17 different measurements, defined by their set of operations $\{\mathcal{O}_r\}$. Each measurement belongs in a distinct region of 'measurement space' in Fig. 1.2. The object of the exercise is to number the 17 regions in this figure with the number (from 1 to 17) corresponding to the appropriate measurement in the list below.

All of the measurements are on an infinite-dimensional system, with basis states $|n\rangle$: $n = 0, 1, 2, \dots$, called number states. Any ket containing n or m indicates a number

state. Any ket containing a complex number $\pm\alpha$, β or γ indicates a *coherent state*, defined as

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \quad (1.175)$$

See also Appendix A. It is also useful to define sets \mathbb{E} and \mathbb{O} , the even and odd counting numbers, respectively. If the result r is denoted n , then the resolution of the identity is $\sum_{n=0}^{\infty} \hat{E}_n$. If it is denoted α then it is $\int d^2\alpha \hat{E}_\alpha$. If denoted \mathbb{E} , \mathbb{O} then it is $\hat{E}_{\mathbb{E}} + \hat{E}_{\mathbb{O}}$.

We also use the following operators in the list below. The operator \hat{D}_β denotes a *displacement operator* defined by how it affects a coherent state:

$$\hat{D}_\beta |\alpha\rangle = |\alpha + \beta\rangle, \quad (1.176)$$

for some non-zero complex number β . The number operator \hat{N} has the number states as its eigenstates. The two operators $\hat{\Pi}_{\mathbb{E}}$ and $\hat{\Pi}_{\mathbb{O}}$ are defined by

$$\hat{\Pi}_{\mathbb{E},\mathbb{O}} = \sum_{n \in \mathbb{E},\mathbb{O}} |n\rangle\langle n|. \quad (1.177)$$

Finally, $\wp_\eta(n|m)$ is as defined in Eq. (1.149) for some $0 < \eta < 1$.

Here is the list.

1. $\mathcal{O}_\alpha = \pi^{-1} \mathcal{J}[|\alpha\rangle\langle\alpha|]$
2. $\mathcal{O}_\alpha = \int d^2\beta \pi^{-2} e^{-|\alpha-\beta|^2} \mathcal{J}[|\beta\rangle\langle\beta|]$
3. $\mathcal{O}_\alpha = \pi^{-1} \mathcal{J}[|0\rangle\langle\alpha|]$
4. $\mathcal{O}_\alpha = \int d^2\gamma \pi^{-2} e^{-|\gamma|^2} \mathcal{J}[|\gamma\rangle\langle\alpha|]$
5. $\mathcal{O}_\alpha = \int d^2\gamma \pi^{-1} e^{-|\gamma|^2} \int d^2\beta \pi^{-2} e^{-|\alpha-\beta|^2} \mathcal{J}[|\gamma\rangle\langle\beta|]$
6. $\mathcal{O}_\alpha = \mathcal{J}[\hat{E}_\alpha^{1/2}]$, $\hat{E}_\alpha = (2\pi)^{-1}(|\alpha\rangle\langle\alpha| + |-\alpha\rangle\langle-\alpha|)$
7. $\mathcal{O}_\alpha = \mathcal{J}[\hat{D}_\beta \hat{E}_\alpha^{1/2}]$, $\hat{E}_\alpha = (2\pi)^{-1}(|\alpha\rangle\langle\alpha| + |-\alpha\rangle\langle-\alpha|)$
8. $\mathcal{O}_n = \mathcal{J}[|n\rangle\langle n|]$
9. $\mathcal{O}_n = \sum_{m=0}^{\infty} \wp_\eta(n|m) \mathcal{J}[|m\rangle\langle m|]$
10. $\mathcal{O}_n = \sum_{m=0}^{\infty} \wp_\eta(n|m) \mathcal{J}[\hat{D}_\beta |m\rangle\langle m|]$
11. $\mathcal{O}_n = \mathcal{J}[|0\rangle\langle n|]$
12. $\mathcal{O}_n = \sum_{m=0}^{\infty} 2^{-(m+1)} \mathcal{J}[|m\rangle\langle n|]$
13. $\mathcal{O}_{\mathbb{E},\mathbb{O}} = \mathcal{J}[\hat{\Pi}_{\mathbb{E},\mathbb{O}}]$
14. $\mathcal{O}_{\mathbb{E},\mathbb{O}} = \mathcal{J}[\exp(i\pi \hat{N}) \hat{\Pi}_{\mathbb{E},\mathbb{O}}]$
15. $\mathcal{O}_{\mathbb{E}} = \sum_{n \in \mathbb{E}} \mathcal{J}[|0\rangle\langle n|]$, $\mathcal{O}_{\mathbb{O}} = \sum_{n \in \mathbb{O}} \mathcal{J}[|1\rangle\langle n|]$
16. $\mathcal{O}_{\mathbb{E},\mathbb{O}} = \mathcal{J}[\hat{D}_\beta \hat{\Pi}_{\mathbb{E},\mathbb{O}}]$
17. $\mathcal{O}_{\mathbb{E},\mathbb{O}} = \sum_{n \in \mathbb{E},\mathbb{O}} \mathcal{J}[|0\rangle\langle n|]$

1.5 Measuring a single photon

In this section we give an experimental example of the quantum measurement of a binary variable, as introduced in Section 1.2.4. This experiment was realized as a ‘cavity QED’ system, a term used to denote the interaction between a discrete-level atomic system and a small number of electromagnetic field modes, which are also treated as quantum

systems. In the experiment performed by the Haroche group in Paris in 1999 [NRO⁺99], the measured system was the state of an electromagnetic field in a microwave cavity. Apart from small imperfections, the preparation procedure produced a pure state containing no more than a single photon. Thus the state of the cavity field may be written as $|\psi\rangle = c_0|0\rangle + c_1|1\rangle$. The measured variable is the photon number with result 0 or 1. The apparatus was an atom with three levels: ground state $|g\rangle$, excited state $|e\rangle$, and an auxiliary state $|i\rangle$. The final readout on the apparatus determines whether the atom is in state $|g\rangle$ by a selective ionization process, which we will describe below. This final readout is not ideal and thus we will need to add an extra classical noise to the description of the measurement.

We begin with a brief description of the interaction between the cavity field and a single two-level atom in order to specify how the correlation between the system and the apparatus is established. If, through frequency or polarization mismatching, the cavity mode does not couple to the auxiliary level $|i\rangle$, then we can define the atomic lowering operator by $\hat{\sigma} = |g\rangle\langle e|$. The field annihilation operator is \hat{a} (see Section A.4). The relevant parts of the total Hamiltonian are

$$\hat{H} = \omega_c \hat{a}^\dagger \hat{a} + \omega_g |g\rangle\langle g| + \omega_e |e\rangle\langle e| + \omega_i |i\rangle\langle i| + (i\Omega/2)(\hat{\sigma}^\dagger + \hat{\sigma})(\hat{a} - \hat{a}^\dagger), \quad (1.178)$$

where Ω is known as the single-photon Rabi frequency and is proportional to the dipole moment of the atom and inversely proportional to the square root of the volume of the cavity mode. We work in the interaction frame (see Section A.1.3) with the free Hamiltonian

$$\hat{H}_0 = \omega_c \hat{a}^\dagger \hat{a} + \omega_g |g\rangle\langle g| + (\omega_g + \omega_c) |e\rangle\langle e| + (\omega_g + \omega_d) |i\rangle\langle i|, \quad (1.179)$$

where ω_d is the frequency of a ‘driving field’, a classical microwave field (to be discussed later). The ‘interaction Hamiltonian’ $\hat{V} = \hat{H} - \hat{H}_0$ becomes the time-dependent Hamiltonian $\hat{V}_{\text{IF}}(t)$ in the interaction frame. However, the evolution it generates is well approximated by the time-independent Hamiltonian

$$\hat{V}_{\text{IF}} = \Omega(i\hat{\sigma}^\dagger \hat{a} - i\hat{\sigma} \hat{a}^\dagger)/2 + \Delta \hat{\sigma}^\dagger \hat{\sigma} + \delta |i\rangle\langle i|, \quad (1.180)$$

where Δ is the detuning $\omega_e - \omega_g - \omega_c$ of the $|e\rangle \leftrightarrow |g\rangle$ transition from the cavity resonance, and $\delta = \omega_i - \omega_g - \omega_d$ is that of the $|i\rangle \leftrightarrow |g\rangle$ transition from the classical driving field. The necessary approximation (called the *rotating-wave approximation*) is to drop terms rotating (in the complex plane) at high frequencies $\sim \omega_c \gg \Delta, \delta, \Omega$. This is justified because they average to zero over the time-scale on which evolution occurs in the interaction frame.

Exercise 1.30 Derive Eq. (1.180) and convince yourself of the validity of the rotating-wave approximation.

Hint: Finding $\hat{V}_{\text{IF}}(t)$ is the same as solving for this operator in the Heisenberg picture with the Hamiltonian \hat{H}_0 . Since \hat{H}_0 splits into a part acting on the atom and a part acting on the field, the Heisenberg equations of motion for the atom and field operators in \hat{V}_{IF} can be solved separately.

Let us now assume that the atom is resonant with the cavity ($\Delta = 0$), in which case the Hamiltonian (1.180) (apart from the final term) is known as the Jaynes-Cummings Hamiltonian. If this Hamiltonian acts for a time τ on an initial state $|1, g\rangle$, the final state is

$$\exp(-i\hat{V}_{\text{IF}}\tau)|1, g\rangle = \cos(\Omega\tau/2)|1, g\rangle + \sin(\Omega\tau/2)|0, e\rangle, \quad (1.181)$$

where $|n, g\rangle \equiv |n\rangle|g\rangle$ and $|n, e\rangle \equiv |n\rangle|e\rangle$.

Exercise 1.31 Show this and the analogous result for the initial state $|0, e\rangle$.

Hint: Derive the equations of motion for the coefficients of the state vector in the above basis.

If the total interaction time $\tau = 2\pi/\Omega$, then the probability that the atom is in the ground state again is unity, but the quantum state has acquired an overall phase. That is to say, for this interaction time, the state changes as $|1, g\rangle \rightarrow -|1, g\rangle$. However, if the field is initially in the vacuum state, there is no change: $|0, g\rangle \rightarrow |0, g\rangle$.

Exercise 1.32 Show this by verifying that $|0, g\rangle$ is an eigenstate of \hat{V}_{IF} with eigenvalue zero.

This sign difference in the evolution of states $|0\rangle$ and $|1\rangle$ provides the essential correlation between the system and the apparatus that is used to build a measurement. If the field is in a superposition of vacuum and one photon, the interaction with the atom produces the ‘conditional’ transformation

$$(c_0|0\rangle + c_1|1\rangle) \otimes |g\rangle \xrightarrow{C} (c_0|0\rangle - c_1|1\rangle) \otimes |g\rangle. \quad (1.182)$$

It is called conditional because the sign of the state is flipped if and only if there is one photon present. Note that we are not using the term here in the context of a measurement occurring.

As it stands this is not of the form of a binary quantum measurement discussed in Section 1.2.4 since the meter state (the atom) does not change at all. In order to configure this interaction as a measurement, we need to find a way to measure the relative phase shift introduced by the interaction between the field and the atom. This is done using the ‘auxiliary’ electronic level, $|i\rangle$, which does not interact with the cavity mode and cannot undergo a conditional phase shift. We begin by using a classical microwave pulse R_1 of frequency ω_d , to prepare the atom in a superposition of the auxiliary state and the ground state: $|g\rangle \rightarrow (|g\rangle + |i\rangle)/\sqrt{2}$. For the moment, we assume that this is resonant, so that $\delta = 0$ in Eq. (1.180). After the conditional interaction, C , between the atom and the cavity field, another microwave pulse R_2 of frequency ω_d again mixes the states $|g\rangle$ and $|i\rangle$. It reverses the action of R_1 , taking $|g\rangle \rightarrow (|g\rangle - |i\rangle)/\sqrt{2}$ and $|i\rangle \rightarrow (|g\rangle + |i\rangle)/\sqrt{2}$.

Exercise 1.33 Show that this transformation is unitary and reverses R_1 .

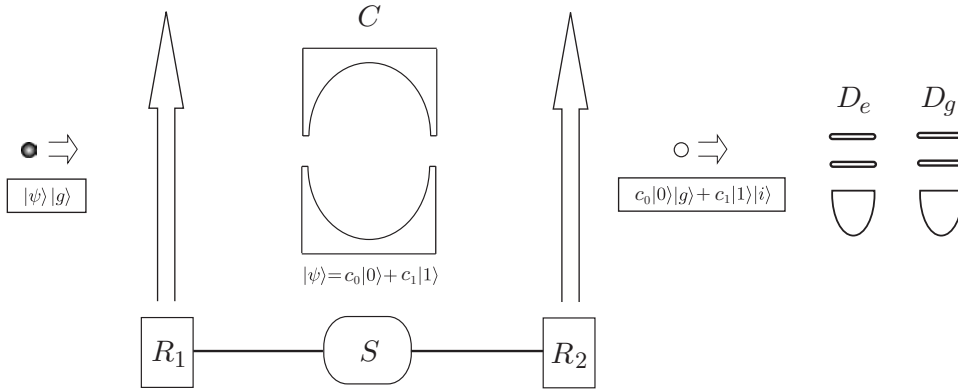


Fig. 1.3 A schematic diagram of the Haroche single-photon measurement [NRO⁺99]. A single atom traverses three microwave fields R_1 , C and R_2 , the middle one being described by a single-mode cavity field. It then encounters two ionization detectors, D_e and D_g , which detect whether the atom is in the excited state or ground state, respectively. The driving fields R_1 and R_2 are produced by the same microwave source, which locks their relative phase. Adapted by permission from Macmillan Publishers Ltd: G. Nogues *et al.*, *Nature* **400**, 239–242 (1999), (Fig. 1), copyright Macmillan Magazines Ltd 1999.

Finally, a projective readout of the ground state $|g\rangle$ is made, as shown in Fig. 1.3. The full measurement protocol can now be described:

$$\begin{aligned}
 (c_0|0\rangle + c_1|1\rangle)|g\rangle &\xrightarrow{R_1} (c_0|0\rangle + c_1|1\rangle)\frac{1}{\sqrt{2}}(|i\rangle + |g\rangle) \\
 &\xrightarrow{C} \frac{1}{\sqrt{2}}(c_0|0\rangle(|i\rangle + |g\rangle) + c_1|1\rangle(|i\rangle - |g\rangle)) \\
 &\xrightarrow{R_2} c_0|0\rangle|g\rangle + c_1|1\rangle|i\rangle.
 \end{aligned} \tag{1.183}$$

An ideal measurement of the ground state of the atom gives a yes (no) result with probability $|c_0|^2$ ($|c_1|^2$), and a measurement of the photon number has been made without absorbing the photon.

To compare this with the binary measurement discussed in Section 1.2.4, we use the apparatus state encoding $|g\rangle \leftrightarrow |0\rangle_A$, $|i\rangle \leftrightarrow |1\rangle_A$. The overall interaction ($R_2 \circ C \circ R_1$) between the system and the apparatus is then defined by Eq. (1.62). We can then specify the apparatus operators \hat{X}_A and \hat{P}_A used in Section 1.2.6,

$$\hat{X}_A = |i\rangle\langle i|, \tag{1.184}$$

$$\hat{P}_A = \frac{1}{2}(|g\rangle - |i\rangle)(\langle g| - \langle i|). \tag{1.185}$$

Likewise the equivalent operators for the system can be defined in the photon-number basis, $\hat{X}_S = |1\rangle\langle 1|$, $\hat{P}_S = (|0\rangle - |1\rangle)(\langle 0| - \langle 1|)/2$. Provided that the atom is initially restricted to the subspace spanned by $\{|g\rangle, |i\rangle\}$, the action of $R_2 \circ C \circ R_1$ can be represented in terms

of these operators by the unitary operator

$$\hat{U}_{R_2 \circ C \circ R_1} = \exp[i\pi \hat{X}_S \hat{P}_A]. \quad (1.186)$$

Certain aspects of the Paris experiment highlight the kinds of considerations that distinguish an actual measurement from simple theoretical models. To begin, it is necessary to prepare the states of the apparatus (the atoms) appropriately. Rubidium atoms from a thermal beam are first prepared by laser-induced optical pumping into the circular Rydberg states with principal quantum numbers 50 (for level *g*) or 51 (for level *e*). The $e \rightarrow g$ transition is resonant with a cavity field at 51.1 GHz. The auxiliary level, *i*, corresponds to a principal quantum number of 49 and the $i \rightarrow g$ transition is resonant at 54.3 GHz.

Next it is necessary to control the duration of the interaction between the system and the apparatus in order to establish the appropriate correlation. To do this, the atoms transiting the cavity field must have a velocity carefully matched to the cavity length. The optical-pumping lasers controlling the circular states are pulsed, generating at a preset time an atomic sample with on average 0.3–0.6 atoms. Together with velocity selection, this determines the atomic position at any time within ± 1 mm. The single-photon Rabi frequency at the cavity centre is $\Omega/(2\pi) = 47$ kHz. The selected atomic velocity is 503 m s^{-1} and the beam waist inside the cavity is 6 mm, giving an effective interaction time τ such that $\Omega\tau = 2\pi$. Finally, a small external electric field Stark-shifts the atomic frequency out of resonance with the cavity. This gives rise to an adjustable detuning Δ in Eq. (1.180), which allows fine control of the effective interaction.

The experiment is designed to detect the presence or absence of a single photon. Thus it is necessary to prepare the cavity field in such a way as to ensure that such a state is typical. The cavity is cooled to below 1.2 K, at which temperature the average thermal excitation of photon number \bar{n} in the cavity mode is 0.15. The thermal state of a cavity field is a mixed state of the form

$$\rho_c = (1 + \bar{n})^{-1} \sum_{n=0}^{\infty} e^{-n\beta} |n\rangle \langle n|, \quad (1.187)$$

where $\beta = \hbar\omega_c/(k_B T)$. At these temperatures, $\beta \ll 1$ and we can assume that the cavity field is essentially in the vacuum state $|0\rangle$. The small components of higher photon number lead to experimental errors.

In order to generate an average photon number large enough for one to see a single-photon signal, it is necessary to excite a small field coherently. This is done by injecting a ‘preparatory’ atom in the excited state, $|e\rangle$, and arranging the interaction time so that the atom-plus-cavity state is $|0, e\rangle + |1, g\rangle$. The state of this atom is then measured after the interaction. If it is found to be $|g\rangle$, then a single photon has been injected into the cavity field mode. If it is found to be $|e\rangle$, the cavity field mode is still the vacuum. Thus each run consists of randomly preparing either a zero- or a one-photon state and measuring it. Over many runs the results are accumulated, and binned according to what initial field state was prepared. The statistics over many runs are then used to generate the

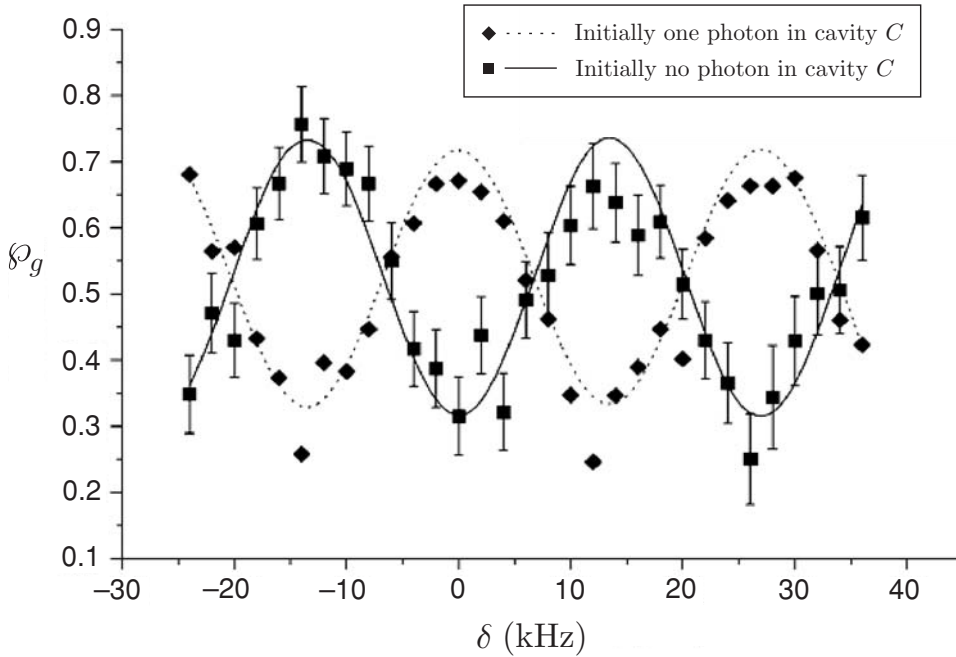


Fig. 1.4 The experimental results of the Paris single-photon experiment, showing the probability of measuring the atom in the ground state versus detuning of the cavity field. The dashed line corresponds to an initial field with a single photon, whereas the solid line is for an initial vacuum field state. Reprinted by permission from Macmillan Publishers Ltd: G. Nogues *et al.*, *Nature* **400**, 239–242 (1999), (Fig. 2), copyright Macmillan Magazines Ltd 1999.

conditional probability of finding the atom in the state $|g\rangle$ when there is one photon in the cavity.

Another refinement of the experiment is to use the detuning δ of the fields R_1 and R_2 to vary the quality of the measurement. This is a standard technique in atomic physics known as *Ramsey fringe interferometry*, or just Ramsey interferometry. This is explained in Box 1.4, where $|e\rangle$ plays the role of $|i\rangle$ in the present discussion. The extra Hamiltonian $\delta|i\rangle\langle i|$ causes free evolution of the atomic dipole. Its net effect is to introduce an extra phase factor δT , proportional to the time T between applications of each of these fields. The probability of finding the atom in state $|g\rangle$ at the end of measurement is then given by

$$\wp_g = \wp_0\mu + \wp_1(1 - \mu), \quad (1.188)$$

where \wp_0 and \wp_1 are the probabilities that the cavity contains no or one photon, respectively, and $\mu = \cos^2(\delta T)$. If $\wp_0 = 1$ or 0 at the start of the measurement, then \wp_g is an oscillatory function of the detuning δ , and the phase of the oscillation distinguishes the two cases.

In Fig. 1.4 we show the experimental results from the Paris experiment. Two cases are shown: in one case (dashed line) the initial state of the field was prepared in a one-photon state (the preparatory atom exited in the ground state), whereas in the second case (solid

Box 1.4 Ramsey fringe interferometry

Ramsey interferometry was developed to measure accurately the frequency ω_{eg} of an atomic transition. It works by producing a signal that depends on the difference δ between the unknown frequency and a known frequency ω_d . It is a standard technique for atomic frequency standards with application to time standards [Maj98]. We here give a simplified treatment of the essential physics behind the technique.

Consider a two-level atom, with ground and excited states $|g\rangle$ and $|e\rangle$, described in the interaction frame with respect to

$$\hat{H}_0 = \omega_g |g\rangle\langle g| + (\omega_g + \omega_d) |e\rangle\langle e|. \quad (1.189)$$

The atom is prepared in the ground state and injected through a classical field R_1 with frequency ω_d that differs from the atomic resonance frequency ω_{eg} by a small detuning δ . The atomic velocity is chosen so that the atom interacts with the field for a precise time τ . The interaction induces a superposition between the ground and excited states of the form

$$|g\rangle \rightarrow \alpha |g\rangle + \beta |e\rangle, \quad (1.190)$$

where the coefficients depend on τ and δ and the Rabi frequency for the transition. (The Rabi frequency is roughly the dot product of the classical electric field with the electric dipole moment of the atomic transition, divided by \hbar . It also equals the single-photon Rabi frequency times the square root of the mean number of photons in the field. For a classical field in a mode with a large mode volume (as here), the former is very small and the latter very large, giving a finite product.) If the detuning δ is small enough, one can arrange to obtain $\alpha = \beta = 1/\sqrt{2}$.

The atom then evolves freely for a time T during which the Hamiltonian in the interaction frame is $\hat{V}_{\text{IF}} = \delta |e\rangle\langle e|$. This changes β to $\beta e^{-i\delta T}$. After this it interacts with another classical field, R_2 , of the same frequency, which undoes the transformation R_1 . This means that we have to adjust T and/or the phase of R_2 so that, if $\delta = 0$, all atoms emerge in the ground state. Then the state of the atom after the second field is

$$\cos(\delta T/2) |g\rangle - i \sin(\delta T/2) |e\rangle. \quad (1.191)$$

The probability that an atom will emerge in the excited state when $\delta \neq 0$ is thus

$$\wp_e(\delta) = \sin^2(\delta T/2). \quad (1.192)$$

By varying the frequency ω_d of the driving fields R_1 and R_2 , and sampling this probability by repeated measurement, we produce interference fringes with a spacing proportional to T^{-1} . A complicating effect is that, for large detuning, the coefficients α and β are not exactly $1/\sqrt{2}$, but also depend on the detuning δ in both amplitude and phase, and this causes the interference-fringe visibility to decrease.

line) the field was prepared in a zero-photon state (the preparatory atom exited in the excited state). In both cases the probability of finding the apparatus atom in the ground state, $|g\rangle$, is plotted as a function of the detuning of the R_1 and R_2 fields. Note that the two cases are π out of phase, as expected.

It is quite apparent from the data that the measurement is far from perfect. The probabilities do not vary from zero to unity, so the contrast or visibility, defined as $(\wp_{\max} - \wp_{\min})/(\wp_{\max} + \wp_{\min})$, is not unity. A primary source of error is the efficiency of the ionization detectors, which is as low as 30%. Also, the interaction that correlates the field and the apparatus is not perfect, and there is a 20% residual probability for the apparatus atom to absorb the photon, rather than induce a π conditional phase shift. Other sources of error are imperfections of the $\pi/2$ Ramsey pulses, samples containing two atoms in the cavity, the residual thermal field in the cavity and the possibility that the injected photon will escape from the cavity before the detection atom enters.

Exercise 1.34 *What is the effect of imperfect ionization detection on the readout? Calculate the mean and variance of the readout variable Y in terms of the system (X) mean and variance for the binary asymmetric measurement defined by the conditioned probabilities $\wp(y|x)$ below:*

$$\begin{aligned}\wp(1|1) &= \eta = 1 - \wp(0|1), \\ \wp(1|0) &= \epsilon = 1 - \wp(0|0).\end{aligned}$$

Here η is the detection efficiency, while ϵ is related to the rate of so-called dark-counts.

1.6 Further reading

Bayesian inference is named after its discoverer, the Reverend Thomas Bayes, who published the idea in 1764. For a modern treatment, see *Bayesian Theory* by Bernardo and Smith [BS94]. It is also covered in *Introduction to Control Theory* by Jacobs [Jac93], and *Optimal Control Theory* by Whittle [Whi96], which are useful references for later chapters as well.

For good introductions to quantum mechanics, see the textbooks by Peres [Per95] and Jauch [Jau68]. Both are notable for discussion of the interpretation of Gleason's theorem. An on-line (in 2009) text that contains a good introduction to the quantum formalism using the Dirac notation is that of Preskill [Pre97].

The theory of generalized measurement has a long history. Early work by Davies and Kraus is summarized in *Quantum Theory of Open Systems* [Dav76], and *States, Effects, Operations: Fundamental Notions of Quantum Theory* [Kra83], respectively. However, both these sources are rather mathematical in style. A more readable book (for physicists) is *Quantum Measurement* by Braginsky and Khalili [BK92]. This is particularly good at helping to build intuition based on simple ideas such as the Heisenberg uncertainty principle. It also contains a detailed discussion of quantum non-demolition measurements.

A different approach to obtaining a formal analogy between classical measurement theory and quantum measurement theory has been developed by Hardy in Ref. [Har02]. In this he derives, for systems with a finite number of distinguishable states, both quantum theory and classical theory from a small number of plausible axioms. The extra axiom that leads to quantum theory rather than classical theory is that there is a *continuous* reversible transformation between any two pure states (pure states having been operationally defined). As suggested by our presentation, the essential difference that follows is one of dimensionality. In the classical case, K , the number of probabilities needed to specify the state, is equal to N , the maximum number of states that can be reliably distinguished in a single shot. In the quantum case, K is equal to N^2 , reflecting the fact that quantum states are represented most naturally as matrices. Note, however, that, in the quantum case, Hardy's map from the K probabilities defining the state to the probabilities of a particular measurement outcome is *not* a positive map – another difference from the classical case.

The consistency of states, both classical and quantum, is much more involved than the treatment we have given (which is equivalent to that in Ref. [BFM02]). The definitions we have given can be viewed as simply one definition in a hierarchy of operational definitions. Interestingly, the hierarchy has a different structure in the quantum and classical cases. This is discussed in detail by Caves, Fuchs and Schack in Ref. [CFS02a]. In this and other papers [CFS02b], these authors strongly argue for interpreting quantum theory using the analogy with classical Bayesian inference. A related question is that of how two or more observers can pool their states of knowledge. For a review of work in this area, and some special results in the quantum case, see Ref. [SW07].

Finally, we note that the work on QND photon detection has advanced far beyond that of Ref. [NRO⁺99] described above. See for example Refs. [GKG⁺07, GBD⁺07], also from the laboratory of Haroche, in which back-action-evading measurements of photon number are performed. This enabled the experimentalists to distinguish several different photon numbers with great accuracy, and to observe quantum jumps between them.