qitd412

Quantum Channels, Kraus Operators, POVMs

Robert B. Griffiths

Version of 22 March 2012

Contents

1 Introduction **1**

2 Kraus Operators **2**

3 Quantum Channels **4**

3.1 Introduction 4
3.2 Model quantum channel 4
3.3 Single qubit channel 5
3.4 Geometrical interpretation 6
3.5 Quantitative measures of noise 7
3.6 Types of quantum information 8

4 Quantum Operations and Superoperators **9**

4.1 Introduction 9
4.2 Quantum operations 11
4.3 One qubit 11
4.4 Kraus representation of quantum operations 14
4.5 Transition operator and dynamical operator 15

5 POVMs **15**

5.1 Definition 15
5.2 Example of a POVM 17
5.3 Naimark extension 18

References:

CQT = *Consistent Quantum Theory* by R. B. Griffiths (Cambridge, 2002)

QCQI = *Quantum Computation and Quantum Information* by M. A. Nielsen and I. L. Chuang (Cambridge, 2000).

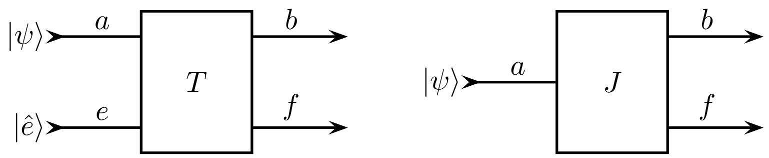
Peres = *Quantum Theory: Concepts and Methods* by A. Peres (Kluwer, 1995).

1 Introduction

★ The quantum circuit in the following figure will be central to our discussion.

• In Fig. 1(a) we have systems a and e, with Hilbert spaces $\mathcal{H}_a$, $\mathcal{H}_e$, of dimension d_a and d_e, initially in states $|\psi\rangle$ and $|\hat{e}\rangle$, respectively, that interact, and the time development from t_0 to t_1 is described by a unitary operator or "gate" T, resulting in a state $|\Psi\rangle$, see (3) below.

• We shall think of $|\Psi\rangle$ as a state on the combined system b and f, Hilbert space $\mathcal{H}_b \otimes \mathcal{H}_f$, at a later time. Simplest situation: b as the same as a, and f is the same as e.



(a) (b)

Figure 1: Quantum channel diagram as (a) unitary map T; (b) isometry J.

○ So why not use the same names? Because we want a formalism that allows the dimension d_b of $\mathcal{H}_b$ to be different from the dimension d_a of $\mathcal{H}_a$. This is consistent with a unitary T provided $d_b d_f = d_a d_e$. If $d_b = d_a$ and $d_f = d_e$ one can identify $\mathcal{H}_b$ with $\mathcal{H}_a$ and $\mathcal{H}_f$ with $\mathcal{H}_e$. The very simplest situation of interest is the one in which T maps two qubits to two qubits.

★ We suppose that e is always in the same initial normalized state $|\hat{e}\rangle$, whereas there are various possibilities for the initial normalized state $|\psi\rangle$ of a; think of $|\hat{e}\rangle$ as a constant and $|\psi\rangle$ as a variable.

• Because $|\hat{e}\rangle$ is held fixed we can take the alternative perspective indicated in Fig. 1(b) and define the operator

$$J|\psi\rangle = T\Big(|\psi\rangle \otimes |\hat{e}\rangle\Big) \tag{1}$$

as a map from $\mathcal{H}_a$ to $\mathcal{H}_b \otimes \mathcal{H}_f$. This map is an *isometry* in that if $|\Psi\rangle = J|\psi\rangle$ and $|\Phi\rangle = J|\phi\rangle$, then $\langle\Phi|\Psi\rangle = \langle\phi|\psi\rangle$, i.e., the J map preserves inner products, and therefore also the norm and the *metric* based upon the norm. (Isometry = preserves the metric.)

□ Exercise. Check that J as defined in (1) preserves norms, provided T is unitary and $|\hat{e}\rangle$ is normalized.

• The isometry J maps the whole Hilbert space $\mathcal{H}_a$ onto a subspace of dimension d_a of the tensor product $\mathcal{H}_b \otimes \mathcal{H}_f$. Consequently, we must have

$$d_a \leq d_b d_f. \tag{2}$$

• While relating J to the unitary T is important and useful for various physical applications, it is worth noting that much of what interests us depends only on the fact that J is an isometry from $\mathcal{H}_a$ to $\mathcal{H}_b \otimes \mathcal{H}_f$.

2 Kraus Operators

★ Choose some orthonormal basis (orbasis) $\{|f^k\rangle\}$ for $\mathcal{H}_f$, and expand

$$|\Psi\rangle = J|\psi\rangle = T\Big(|\psi\rangle \otimes |\hat{e}\rangle\Big) = \sum_k |\beta^k\rangle \otimes |f^k\rangle \tag{3}$$

in this basis, assuming $|\psi\rangle$ is normalized, where the "expansion coefficients," the kets $|\beta^k\rangle$, are in general neither normalized nor orthogonal.

○ One can think of carrying out a simple measurement on f in the $\{|f^k\rangle\}$ basis, and if the outcome (position of the pointer of the measuring apparatus) is k, then just before the measurement f was in the state $|f^k\rangle$. This measurement outcome, or the corresponding state itself just before the measurement, occurs with a probability

$$p_k = \|\beta^k\|^2 = \langle\beta^k|\beta^k\rangle, \tag{4}$$

and when it occurs we know that we should assign to $\mathcal{H}_b$ the normalized pre-probability

$$|\hat{\beta}^k\rangle = |\beta^k\rangle/\|\beta^k\| \tag{5}$$

if we want to calculate probabilities of properties of b.

★ It is useful to write the relationship between $|\beta^k\rangle$ and $|\psi\rangle$, assuming the isometry J is held fixed (which will be true if $|\hat{e}\rangle$ and T remain the same), and the basis $\{|f^k\rangle\}$ is also held fixed, in the form

$$|\beta^k\rangle = K_k|\psi\rangle \tag{6}$$

where the *Kraus* operator K_k is a linear map from $\mathcal{H}_a$ to $\mathcal{H}_b$ (hence from $\mathcal{H}_a$ to itself in the case in which b is the same as a). Then (3) takes the form

$$J|\psi\rangle = T\Big(|\psi\rangle \otimes |\hat{e}\rangle\Big) = \sum_k \Big(K_k|\psi\rangle\Big) \otimes |f^k\rangle. \tag{7}$$

○ The operators E_k used in QCQI Sec. 8.2.3 and later (pp. 360ff) are what we call Kraus operators; QCQI never uses the term "Kraus." Also, QCQI focuses on the situation $\mathcal{H}_b = \mathcal{H}_a$, but it is useful to also allow cases in which d_b is less than or larger than d_a.

○ Likewise, the operators M_m introduced in QCQI Sec. 2.2.3, which they call "measurement operators," are Kraus operators.

• That there is a linear relationship (6) between $|\beta^k\rangle$ and $|\psi\rangle$ may not be immediately evident. It follows from (3), and the fact that the expansion coefficient $|\beta^k\rangle$ is uniquely determined by $|\psi\rangle$ if everything else is held fixed. See the following exercise.

□ Exercise. Show that the relationship between $|\psi\rangle$ and $|\beta^k\rangle$ for a fixed k given by (3) is linear by looking at what happens when (i) $|\psi\rangle$ is multiplied by a complex number c, (ii)$|\psi\rangle$ is replaced by a sum $|\chi\rangle + |\omega\rangle$.

★ The fact that J is an isometry (implied by $|\hat{e}\rangle$ normalized and T unitary) means that

$$\langle\psi|\psi\rangle = \langle\Psi|\Psi\rangle = \sum_k \langle\beta^k|\beta^k\rangle = \sum_k \langle\psi|K_k^\dagger K_k|\psi\rangle = \langle\psi|\Big(\sum_k K_k^\dagger K_k\Big)|\psi\rangle, \tag{8}$$

will be true for every $|\psi\rangle$. This will be the case if and only if the *closure condition*

$$\sum_k K_k^\dagger K_k = I_a \tag{9}$$

is satisfied, where I_a is the identity on $\mathcal{H}_a$. See the exercise.

□ Exercise. Show that if for a fixed operator A it is the case that for *every* normalized $|\psi\rangle$ in $\mathcal{H}_a$ the quantity $\langle\psi|A|\psi\rangle$ is 1, then $A = I_a$. [Hint. Introduce an orbasis $\{|a^j\rangle\}$, expand $A|a^j\rangle$ in this basis, and show that $A|a^j\rangle = |a^j\rangle$.]

• Note that since K_k maps $\mathcal{H}_a$ to $\mathcal{H}_b$, its adjoint $K_k^\dagger$ maps $\mathcal{H}_b$ to $\mathcal{H}_a$, and consequently the product $K_k^\dagger K_k$ maps $\mathcal{H}_a$ to itself, which is why I_a appears on the right side of (9), and not I_b. This distinction is important when $\mathcal{H}_b$ is different from $\mathcal{H}_a$. When $\mathcal{H}_b = \mathcal{H}_a$ (as in QCQI) one does not need the subscript on I.

★ It is obvious from (7) that the Kraus operators *depend on the choice of orbasis* $\{|f^k\rangle\}$ for f. Do not misinterpret this as meaning that different types of measurements on f, corresponding to different choices of orbasis, will somehow "influence" $\mathcal{H}_b$. Instead, they reflect different frameworks for relating properties of b to those of f. Remember that the $\{|\beta^k\rangle = K_k|\psi\rangle\}$ are conditional *pre-probabilities* (up to normalization) used to calculate correlations; they are not physical properties created by some mysterious action-at-a-distance.

• Put in other words, what Fred, who measures f, can learn about the system b in Bob's possession depends on what he learns about f, which in turn depends on the type of measurement he carries out on f, say S_x as against S_z in the case of a qubit. Fred's measurement has absolutely no physical effect on b; to suppose otherwise is to fall prey to the nonlocality myth.

3 Quantum Channels

3.1 Introduction

★ A basic issue in both quantum computation and quantum cryptography is that one needs to get information from one point to another in a reliable way. Even storing quantum information at one particular point is a nontrivial issue, for it tends to decay or degrade.

• Hence the study of quantum channels, used to transmit or store quantum information, is a very important topic.

• Think of a quantum channel as a pipe through which one transmits a spin-half particle, thus (in its spin degree of freedom) a single qubit. Small magnetic fields inside the pipe may perturb the information carrier, or it may bump into something else in the pipe. This will produce noise.

• An alternative picture, which is easier to realize in practice, is an optical fiber in which a single photon "carries" one qubit in its polarization. Inhomogeneities in the fiber may disturb the polarization. Or the photon may simply disappear through some absorption process. (This last becomes a significant problem after a few kilometers.) Both of these produce undesired noise.

★ For classical channels (ordinary telephone lines) there are well-understood ways of overcoming noise in a channel through error correction. This is also in principle possible in quantum channels, though a lot more difficult.

3.2 Model quantum channel

★ A rather common model of a quantum channel as used in quantum information theory can be represented schematically by Fig. 1, where a is the *input* to the channel, b is the channel *output*, e is thought of as the *environment* which is initially in the state $|\hat{e}\rangle$, and f is again the *environment* at the later time when information emerges in the channel output.

• One thinks of a as, typically, some small system, a photon or an atom or a small number of such entities, and e as the "rest of the world," or at least enough of the world so that for the purposes of interest the two together can be treated as an isolated quantum system with time development governed by Schrödinger's equation, represented in Fig. 1(a) by the time transition operator T.

• The question one asks is then: how is the output system b related to or correlated with the channel input a if one ignores f, or knows nothing about f?

★ Imagine a spin-half particle which enters a pipe in the (spin) state $|\psi\rangle$. If the *same* $|\psi\rangle$ emerges at the other end, and this is true for *every* input $|\psi\rangle$, the channel is *perfect.* Anything else constitutes a *noisy* quantum channel.

• One source of noise could be a static magnetic field somewhere in the pipe, which causes the spin to precess. The result is that an input $|\psi\rangle$ emerges as $U|\psi\rangle$, where U is a unitary operator which depends on the magnetic field, but is independent of the input $|\psi\rangle$. I call this a *unitary* channel. It corresponds to the following circuit:

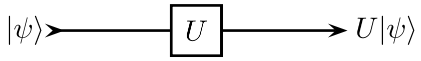


Figure 2: Unitary channel

• One can also call a unitary channel an *ideal* channel because performing the inverse operation $U^\dagger$ at the output changes it into a perfect channel. Thus any errors produced by the channel can be completely removed by this very simple form of error correction.

★ On the other hand, if the magnetic field varies randomly with time, so its value will be different each time the channel is used, of if the particle bumps into another particle which causes its spin to flip, one cannot describe the output of the channel in terms of a pure ket for a given input $|\psi\rangle$, but instead one will need a density operator ρ which depends on the input state.

★ It is customary to model such a channel as (what I call) a *linear* channel, in the manner indicated in Fig. 1, where a is the channel input, b, which may be the same as a, is the output, and noise is produced by interaction with an environment e through a unitary time development operator T.

• In this situation one is interested in how the output b is related to the input a if we know nothing about the final environment f. Thus we will need to describe the output by means of an ensemble, or more compactly (with less irrelevant information) by a density operator ρ:

$$|\Psi_0\rangle = |\psi\rangle \otimes |\hat{e}\rangle, \quad |\Psi_1\rangle = T|\Psi_0\rangle, \quad \rho = \mathrm{Tr}_f\big([\Psi_1]\big), \tag{10}$$

where, as usual, $[\Psi_1]$ is short for $|\Psi_1\rangle\langle\Psi_1|$.

∘ What if the environment is not initially in a pure state but in some mixed state? We can always take the corresponding environment density operator and purify it by introducing a fictitious reference system $\mathcal{R}$. Thinking of $\mathcal{R}$ as part of the environment, but as a part that interacts with nothing else, allows us to employ a pure state in working out the formalism.

3.3 Single qubit channel

★ A simple example is provided by the case in which channel and environment are both single qubits, and T is a controlled-not gate, as in Fig. 3

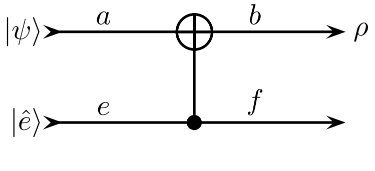


Figure 3: Simple example of a noisy channel

• The nature of the $|\psi\rangle$-to-ρ channel depends upon the initial state $|\hat{e}\rangle$. For $|\hat{e}\rangle = |0\rangle$ we have a perfect channel. For $|\hat{e}\rangle = |1\rangle$ the result is a unitary channel in which $|\psi\rangle = |0\rangle$ is mapped to $|1\rangle$ and vice versa, thus a unitary *bit flip* or (in the notation of QCQI) X gate; i.e., $|\psi\rangle$ is mapped to $\sigma_x|\psi\rangle$

• Suppose we look at something in between,

$$|\hat{e}\rangle = \sqrt{1-p}\,|0\rangle + \sqrt{p}\,|1\rangle. \tag{11}$$

□ Exercise. Find ρ using (11) and an initial state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ for qubit a by applying the controlled-not gate to $|\Psi_0\rangle = |\psi\rangle \otimes |\hat{e}\rangle$ to get $|\Psi_1\rangle$, and then a partial trace to $[\Psi_1]$.

★ The results can be worked out analytically, see the exercise. Here let us take an alternative approach which provides a geometrical representation of the operation of the channel. Imagine that instead of throwing away the environmental qubit we measure it in the standard basis after it has interacted with a. The result will be $D = 0$ or $D = 1$, with probabilities $1 - p$ and p, respectively. If $D = 0$ then we know that qubit a is unchanged, so described by the state $|\psi\rangle$. If $D = 1$ we know that it has been "flipped", and is described by the state $\sigma_x|\psi\rangle$.

• Thus we can describe qubit a using an *ensemble* of $|\psi\rangle$ with probability $1-p$ and $\sigma_x|\psi\rangle$ with probability p, which is to say:

$$\rho = (1-p)[\psi] + p[\sigma_x\psi]. \tag{12}$$

□ Exercise. Measuring f in the standard basis will give rise to a pair of Kraus operators, Sec. 3, K_0 and K_1. Find out what these are, check that they satisfy the closure condition (9), and relate them to the preceding discussion.

◦ Comment. Measuring or not measuring the environment qubit (f) after it has interacted with a cannot possibly have any influence on a (unless we were to use the result of the measurement to carry out some additional operation on a); avoid the idea that it produces a "collapse of the wave function" in some physical sense. Instead, we are using a hypothetical measurement as a means of calculating the density operator ρ, which will be exactly the same when computed in this way or by calculating the unitary time evolution of the initial ket and then taking a partial trace (see preceding exercise). Remember, ρ functions as a pre-probability.

□ Exercise. Suppose that instead of measuring f in the standard basis we measure it in the S_x or $|+\rangle$, $|-\rangle$ basis. Once again find the Kraus operators K_+ and K_-. Show that using them leads to the same result as in (12). Write the operators K_+ and K_- as linear combinations of the K_0 and K_1 found earlier; the relationship should be unitary. (See QCQI Sec. 8.2.4.)

3.4 Geometrical interpretation

★ One can give the following *geometrical interpretation* to (12). The σ_x or X transformation is a rotation of the Bloch sphere by 180° around the x axis. Therefore, if we start with a particular $|\psi\rangle$, represented by a point on the Bloch sphere, (12) is telling us that the final density operator is a *convex combination* of this original point with another point obtained by rotating the original point by 180° around the x axis, with weights $1-p$ and p, respectively.

• Density operators (not just of qubits) form a convex set. A convex combination of two points of a convex set, as in (12), can be thought of as a point on the straight line connecting them. This point is located at the *center of mass* when a mass of $(1-p)$ is placed at $[\psi]$ and a mass of p at $[\sigma_x\psi]$.

• Consequently, in the map which produces ρ from $|\psi\rangle$, every point on the (surface of the) Bloch sphere is mapped to a point lying inside the Bloch sphere (or possibly on the surface) in the same y, z plane, since the rotation takes place around the x axis.

◦ The net result is to shrink the original Bloch sphere into an ellipsoid which for $0 < p < 1/2$ touches the original sphere at only two points: where the original sphere intersects the x axis. See Fig. 8.8 in QCQI, p. 376.

□ Exercise. What happens for $1/2 < p < 1$, and in what way does this channel differ from $0 < p < 1/2$? Is there some analogy with a symmetrical ($\epsilon_0 = \epsilon_1 = p$) classical 1-bit channel?

□ Exercise. Suppose in place of (11) we use $|\hat{e}\rangle = \sqrt{1-p}\,|0\rangle + i\sqrt{p}\,|1\rangle$. How does this change the $|\psi\rangle$-to-ρ channel? What channel is produced by a general $|\hat{e}\rangle = \zeta|0\rangle + \eta|1\rangle$ with ζ and η two complex numbers with $|\zeta|^2 + |\eta|^2 = 1$?

3.5 Quantitative measures of noise

★ The simplest measure of noise in a classical or quantum channel of the sort we are considering is the *error rate*, which will be denoted by ϵ. It depends upon what enters the channel.

• For a classical channel with conditional probabilities $\Pr(b\,|\,a)$ of output in terms of input of

$$\Pr(0\,|\,0) = 1 - \epsilon_0, \quad \Pr(1\,|\,0) = \epsilon_0, \quad \Pr(0\,|\,0) = 1 - \epsilon_1, \quad \Pr(1\,|\,0) = \epsilon_1. \tag{13}$$

the error rate is ϵ_0 for input 0 and ϵ_1 for input 1. That is, every time a 0 is sent into the channel there is a probability ϵ_0 that it will be flipped to 1, thus making an error.

○ The error rate can be measured experimentally using the usual methods for estimating probabilities. If 0 is sent into the channel N times, and 1 emerges M times, while 0 emerges $N - M$ times, then M/N is an estimate for ϵ_0, one which should improve as N increases. Error rates obviously cannot be determined by using a channel just once.

★ The definition of error rate for a quantum channel is as follows, assuming the input and output Hilbert spaces are identical: $\mathcal{H}_a = \mathcal{H}_b$ in the notation of Fig. 1. Suppose one repeatedly sends in the state $|\psi\rangle$, and measures the output to determine whether it is in the same state $|\psi\rangle$ or in a state *orthogonal* to $|\psi\rangle$. That is, do a measurement corresponding to the decomposition of the identity

$$I = \big([\psi]\big) + \big(I - [\psi]\big). \tag{14}$$

○ Note that it generally doesn't make sense to ask whether a quantum system is in a physical state $|\psi\rangle$ or in a state $|\phi\rangle$ when these states are not orthogonal to each other. Suppose, for example, that a qubit is sent into a unitary channel in the state $|0\rangle$ and emerges in the state $\big(0.99|0\rangle + 0.1|1\rangle\big)$. The output state as a mathematical object (or a pre-probability) is obviously not identical to the input state. So was there an error? Not necessarily, according to the above definition. Instead, there is a nonzero probability of an error. Whether or not an error occurs on some particular occasion when the channel is used is not something quantum mechanics can predict, since quantum dynamics is inherently random or stochastic.

□ Exercise. Compute this nonzero probability.

★ Rather than the error rate ϵ one can use the *fidelity* F defined by

$$F = 1 - \epsilon. \tag{15}$$

• Warning! For reasons best known to them, Nielsen and Chuang define the fidelity to be $\sqrt{(1-\epsilon)}$. They then go on to change their definition in midstream, when they get to entanglement fidelity. replacing their earlier definition by the one used here. I shall always use the definition in (15).

★ For a general linear quantum channel and supposing that the input is a state $|\psi\rangle$, the error rate ϵ is the probability of the second projector on the right side of (14), and the fidelity is the probability of the first, thus

$$F(\psi) = \mathrm{Tr}(\rho[\psi]) = \langle\psi|\rho|\psi\rangle, \quad \epsilon = 1 - F = 1 - \langle\psi|\rho|\psi\rangle. \tag{16}$$

where ρ is the pre-probability for what emerges from the channel when $|\psi\rangle$ is fed in. Thus ρ depends on the initial $|\psi\rangle$, though that is not indicated explicitly in (16).

• The fidelity, just like the error rate, depends on which state one puts into the channel. In the classical case of a 1 bit channel this isn't too bad, for there are only two possibilities. For a one qubit quantum channel, things are a lot more complicated, as we shall see, but not infinitely so, at least not for the linear model introduced in Sec. 3.2 (and which underlies all the discussions in QCQI).

★ It is sometimes helpful to have a single number rather than a function of the initial $|\psi\rangle$ to characterize a quantum channel. One possibility is the *minimum fidelity*

$$F_{\min} := \min_{|\psi\rangle} F(\psi) \tag{17}$$

over all possible inputs. While this is a relatively crude measure of what is going on, one can see how it would be useful to an engineer trying to design a quantum computer and producing specifications for the manufacture of some channel. If $F_{\min}$ is close to 1, then the probability of any error is small, no matter what is sent into the channel.

• The engineer building a quantum computer might actually want to use something else: the *entanglement fidelity* defined in the following way. Imagine that the input to the channel is part of an entangled state $|\Psi_0\rangle$ on the tensor product $\mathcal{H}_a \otimes \mathcal{H}_r$ of the Hilbert space of the channel input a and that of another system r. Then imagine that the time development of a plus its environment e (which defines the channel, see Fig. 1) along with r is given by $T \otimes I$, where T acts on $\mathcal{H}_a \otimes \mathcal{H}_e$, see (10) and I on $\mathcal{H}_r$. The entanglement fidelity is then defined by using the density operator R:

$$F_{\text{ent}} = \langle\Psi_0|R|\Psi_0\rangle, \quad R := \text{Tr}_f\big[T([\Psi_0] \otimes [\hat{e}])T^\dagger\big]. \tag{18}$$

∘ Note that F_{ent} is nothing but the (ordinary) fidelity associated with a bigger channel whose input and output is $\mathcal{H}_a \otimes \mathcal{H}_r$, and which interacts with the environment only through the a part; the r part is a perfect channel.

• Once again the fidelity F_{ent} depends upon the initial $|\Psi_0\rangle$, though one can show (see QCQI) that it only depends upon $|\Psi_0\rangle$ through the initial reduced density operator on a defined by

$$\rho_0 := \text{Tr}_r([\Psi_0]). \tag{19}$$

∘ The entanglement fidelity in QCQI, p. 420, is written as $F(\rho, \mathcal{E})$, where ρ means the same thing as ρ_0 in (19)), and $\mathcal{E}$ is the superoperator for the channel, see below.

• The *minimum entanglement fidelity*, let us denote it by the somewhat awkward F_{minent}, is the minimum of F_{ent} over all possible $|\Psi_0\rangle$ (the same as all possible ρ_0). It is always less than or equal to $F_{\min}$; this is the significance of (9.137) in QCQI, which would look a little less odd had the authors not changed their definition of "fidelity".

∘ The reason the minimum entanglement fidelity could be useful to the quantum engineer is that he does not know in advance what will actually be going on inside the quantum computer, and some unitary operation could require that the channel in question be in an entangled state. Then F_{minent} limits how bad things can be in the worst-case scenario.

3.6 Types of quantum information

★ Associated with an orthonormal basis $\{|a^j\rangle\}$ of the channel input Hilbert space $\mathcal{H}_a$ is what we shall term a particular *type* or *species* of quantum information. More generally, a type is associated with a decomposition of the identity

$$I_a = \sum_j P^j. \tag{20}$$

• In the case of a qubit the Z or S_z type of information is associated with the orthonormal basis $\{|0\rangle, |1\rangle\}$ or the decomposition $\{[0], [1]\}$; the X or S_x type of information corresponds to the basis $\{|+\rangle, |-\rangle\}$, and so forth.

• Two *types* of information are *compatible* if and only if the projectors from one decomposition commute with those from the other; if some pairs fail to commute the types are *incompatible.*

★ A perfect quantum channel is one for which all types of information are transmitted to the output with zero error rate or a fidelity of 1. An ideal quantum channel becomes a perfect quantum channel if a single unitary gate applied to the output changes it into a perfect channel.

★ A noisy quantum channel will typically transmit different types of information with different fidelities. By applying a unitary correction at the end of the channel, or perhaps doing something more complicated, one may be able to improve the fidelity or reduce the error rate for one type of information, but this will, in general, increase the error rate for other types of information.

★ An *ideal* classical channel is a quantum channel which can, with suitable error correction if needed, transmit one particular type of quantum information without error.

• The geometrical interpretation of Sec. 3.4 is helpful in understanding this. Suppose that the original Bloch sphere is mapped into an ellipsoid that touches the original sphere at the points which intersect the x axis but nowhere else: $|+\rangle$ or $[+]$ is mapped to $[+]$, $|-\rangle$ or $[-]$ is mapped to $[-]$, but all other pure states on the surface of the Bloch sphere are mapped into points in the interior. Then so far as the X (S_x) type of information is concerned one has an ideal, or in fact a perfect classical channel since no corrections are needed. But any other type of information is transmitted with a certain amount of noise, so the channel will not be ideal for these other types.

• The extreme case is that in which the ellipsoid just discussed shrinks down to nothing but a line joining the $[+]$ and $[-]$ points on the Bloch sphere. One then has an example of what might be called a "classical" channel or a "perfectly decohered" quantum channel for this type of information.

◦ These names are not standard. In fact there is some confusion in the literature associated with various uses of the term "classical" in the context of quantum information.

□ Exercise. Can such a perfectly decohered, ideal quantum channel be produced using the circuit in Fig. 3 with a suitable choice of $|\hat{e}\rangle$?

4 Quantum Operations and Superoperators

4.1 Introduction

★ The case of a quantum channel corresponds to the situation in Fig. 1 in which we are interested in how the properties of system b are related to the initial state $|\psi\rangle$ without regard to any correlations it may have with system f. Quantum mechanics provides several alternative mathematical tools for describing a situation of this sort.

• We may compute the "big" wave function $|\Psi\rangle$ generated from $|\psi\rangle$ by the unitary time evolution T of the isometry J, and use it to calculate probabilities of projectors $\{P^j\}$ acting on $\mathcal{H}_b$. Or we may form the reduced density operator

$$\rho_b = \mathrm{Tr}_f\Big(|\Psi\rangle\langle\Psi|\Big) \tag{21}$$

and calculate the same probabilities using the formula $\mathrm{Tr}_a(P^j\rho_b)$. Or imagine an ensemble $\{p_k, |\hat{\beta}^k\rangle\}$, see (4) and (5), and calculate averages with respect to each element. Thus if P^j is a projector—or, for that matter, any operator whatsoever—on $\mathcal{H}_b$, the following all give the same result for what is often written in the form $\langle P^j\rangle$:

$$\langle\Psi|P^j|\Psi\rangle = \langle\Psi|P^j\otimes I_f|\Psi\rangle = \mathrm{Tr}_b(P^j\rho_b) = \sum_k p_k\langle\hat{\beta}^k|P^j|\hat{\beta}^k\rangle. \tag{22}$$

○ Note that one very often writes P^j in place of $P^j\otimes I_f$ when its meaning is clear from the context.

□ Exercise. Check the last two equalities in (22)

• In many ways the "cleanest" of the expressions in (22) is the one involving ρ_b, since this reduced density operator contains no information at all about the correlations between b and f. It is relatively compact in comparison with the big wave function $|\Psi\rangle$, and it does not depend upon the choice of a basis for f, unlike the ensemble $\{p_k, |\hat{\beta}^k\rangle\}$.

• A disadvantage of ρ_b is that it does not depend in a linear fashion on the original $|\psi\rangle$, and calculations are usually simpler if we employ linear relationships. This defect can be remedied by using

$$\rho_a := |\psi\rangle\langle\psi| \tag{23}$$

in place of $|\psi\rangle$, and introducing the *superoperator* $\mathcal{S}$ that provides a linear relationship

$$\rho_b = \mathcal{S}(\rho_a) \tag{24}$$

between the output and the input.

★ A superoperator $\mathcal{S}$ is any linear map from the space $\hat{\mathcal{H}}_a$ of *operators* on $\mathcal{H}_a$ to the space $\hat{\mathcal{H}}_b$ of *operators* on $\mathcal{H}_b$.

○ The dimension of $\mathcal{H}_b$ can be different from the dimension of $\mathcal{H}_a$.

• $\hat{\mathcal{H}}_a$ is a complex linear vector space since the sum $A + A'$ of any two operators is an operator, as is cA when c is a complex number. In the finite-dimensional case we can introduce an operator inner product, known as the Frobenius or Hilbert-Schmidt inner product, by writing

$$\langle A, A'\rangle := \mathrm{Tr}_a(A^\dagger A'). \tag{25}$$

□ Exercise. Check that (25) defines a proper inner product satisfying the usual conditions: linear in the second argument, antilinear in the first, and greater than 0 when $A' = A$ (unless $A = 0$).

• Since $\hat{\mathcal{H}}_a$ and $\hat{\mathcal{H}}_b$ are Hilbert spaces, the superoperator $\mathcal{S}$ is in this sense nothing more than a linear map from one Hilbert space to another Hilbert space.

★ A superoperator can be characterized by its *matrix elements* relative to bases of operators for the spaces $\hat{\mathcal{H}}_a$ and $\hat{\mathcal{H}}_b$.

• Let $\{A_j\}$ be a collection of linearly-independent operators spanning $\hat{\mathcal{H}}_a$ and $\{B_k\}$ another collection spanning $\hat{\mathcal{H}}_b$. We define the matrix S_{kj} of the superoperator $\mathcal{S}$ in the standard fashion

$$\mathcal{S}(A_j) = \sum_k S_{kj}B_k, \tag{26}$$

where note the order of the subscripts on the right side: kj and not jk. This is the standard definition for a matrix in linear algebra. Given $\mathcal{S}$ and the two bases, the matrix S_{kj} is unique; conversely, S_{kj} along with the two bases defines a unique superoperator.

∘ The operator bases are often chosen to be orthogonal using the inner product (25), and sometimes normalized, but don't have to be.

∘ Some authors reserve the term "superoperator" for the situation where it corresponds to a *quantum operation* as defined below.

4.2 Quantum operations

★ Superoperators which can be written in the form

$$\mathcal{S}(A) = \mathrm{Tr}_f\Big(T\big(A \otimes |\hat{e}\rangle\langle\hat{e}|\big)T^\dagger\Big), \tag{27}$$

where T is a unitary operator and $|\hat{e}\rangle$ is normalized have some rather special properties:
i) They map Hermitian operators to Hermitian operators.
ii) They map positive operators to positive operators.
iii) They preserve the trace:

$$\mathrm{Tr}_b\Big(\mathcal{S}(A)\Big) = \mathrm{Tr}_a(A) \tag{28}$$

iv) They are *completely positive*, a stronger condition than (ii) which will be defined in Sec. 4.5.

∘ Note that properties (ii) and (iii) guarantee than when $\mathcal{S}$ is applied to a density operator, the result is a density operator.

★ We shall use the term *quantum operation* for a superoperator which satisfies these four properties. Or in slightly different terms, such a superoperator *represents* (the effects of) a quantum operation.

• The term "completely positive trace preserving map" is often used in the literature.

∘ This is approximately the same terminology used in QCQI where, however, b is the same as a. Also QCQI ties the quantum operation to a collection of Kraus operators, see Sec. 4.4.

• It can be shown that any quantum operation is of the form (27); i.e., one can find additional Hilbert spaces $\mathcal{H}_e$ and $\mathcal{H}_f$, with the former initially in a normalized state $|\hat{e}\rangle$, and a unitary T, such that the superoperator representing the quantum operation has the form (27).

4.3 One qubit

★ In the case of a single qubit the four operators

$$\sigma_0 = I, \quad \sigma_1 = \sigma_x = X, \quad \sigma_2 = \sigma_y = Y, \quad \sigma_3 = \sigma_z = Z, \tag{29}$$

indicated using a variety of notations, are a basis for the operator Hilbert space $\hat{\mathcal{H}}$. Each of these operators is Hermitian, and consequently a general Hermitian operator on the space of 1 qubit can be written as a linear combination of these four operators with *real* coefficients.

□ Exercise. Show that (29) is actually an *orthonormal* basis of the operator space if one redefines the inner product (25) with a factor of 1/2 on the right side.

□ Exercise. Starting with (29) construct an orthonormal (when (25) has been suitably modified) basis for the operator space of the tensor product of 2 qubits. Is it obvious how to proceed to the general case of n qubits?

★ If we use the basis (29) for both $\hat{\mathcal{H}}_a$ and $\hat{\mathcal{H}}_b$, and write (26) in the form

$$\mathcal{S}(\sigma_{aj}) = \sum_k S_{kj}\sigma_{bk} \tag{30}$$

the 4×4 matrix S_{jk} with $0 \leq j, k \leq 3$ representing a quantum operation has the form

$$S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ b_1' & c_{11}' & c_{12}' & c_{13}' \\ b_2' & c_{21}' & c_{22}' & c_{23}' \\ b_3' & c_{31}' & c_{32}' & c_{33}' \end{pmatrix}, \tag{31}$$

where the $\{b_j'\}$ can be thought of as a real three-dimensional vector, and the $\{c_{jk}'\}$ form a real 3×3 matrix.

□ Exercise. Explain why the elements of the matrix S are real. [Hint. Property (i) of Sec. 4.2.]

□ Exercise. Show that the top row of (31) is a consequence of condition (iii).

□ Exercise. Show that $\mathcal{S}(\rho)$ maps a point $\mathbf{r} = (r_1, r_2, r_3)$ in the Bloch sphere to $\mathbf{r}'$, where

$$r_j' = b_j' + \sum_{k=1}^{3} c_{jk}' r_k. \tag{32}$$

• The twelve parameters in (31) are in some sense the 1 *qubit* channel counterparts of the two parameters ϵ_0 and ϵ_1 for the 1 *bit* classical channel of (13). In this sense the quantum channel is much more complicated than its classical counterpart. However, one can make it a bit simpler, as we shall now show.

★ Consider the unitary channel, Fig. 2. For any U except the identity this channel will be noisy for most input states. But there is a simple way, at least in principle, to get rid of this noise. By making lots of measurements one can determine U (up to an uninteresting overall phase), and if one's laboratory is well-equipped, the noise can be effectively removed by passing the carrier of information through the inverse transformation $U^\dagger = U^{-1}$ when it emerges from the channel. Or one can apply $U^\dagger$ at the sending end, before the carrier enters the channel.

◦ The classical counterpart: interchanging the inputs, or the outputs, of a 1 bit channel in which every bit is flipped.

• A similar simplification is possible for a general noisy channel if one imagines that unitary operations (in general different from each other) can be applied both at the beginning and at the end of the channel. Doing so helps us reduce the number of free parameters characterizing the channel if we suppose that two channels which become identical by applying unitaries at the beginning and end of one of them are in some sense interchangeable.

★ Thus we define two quantum channels characterized by superoperators $\mathcal{S}_1$ and $\mathcal{S}_2$ as *equivalent up to local unitaries* when it is the case that there are unitaries U and V such that

$$\mathcal{S}_2(A) = V\mathcal{S}_1(UAU^\dagger)V^\dagger \tag{33}$$

for any operator A (thus, in particular, for any density operator ρ).

□ Exercise. Rewrite (33) in the form $\mathcal{S}_1(A') = \ldots \mathcal{S}_2(\ldots A' \ldots) \ldots$.

◦ Note that this definition is not limited to qubit channels, but applies quite generally provided both channels have input spaces of the same dimension, and also output spaces of the same dimension. The dimensions of the input and output spaces can be different.

◦ The term "local" in "local unitaries" reflects the fact that one often imagines the input of the channel to be located in one place (Alice's laboratory), where U is applied, and the output in another (Bob's laboratory), where V is applied.

• Notice that error rates and fidelities, as we have defined them, are *not* invariant under local unitaries. In particular a unitary channel with very bad fidelity for at least some $|\psi\rangle$ can be turned into a perfect channel by employing a local unitary.

★ In the case of a single qubit, any unitary operation is equivalent to rotating the Bloch sphere in some manner; we noted this earlier in connection with σ_x. (Reflections and other improper rotations of the Bloch sphere, for which the determinant of the rotation matrix is -1, do *not* correspond to unitary transformations.)

• As a consequence it is always possible by means of local unitaries to transform the S matrix in (31) into the form

$$S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ b_1 & c_1 & 0 & 0 \\ b_2 & 0 & c_2 & 0 \\ b_3 & 0 & 0 & c_3 \end{pmatrix}. \tag{34}$$

The 3×3 c' matrix in (31) can be diagonalized by applying distinct rotation matrices R_l and R_r on the left and right, so that $c = R_l c' R_r$, and $b = R_l b'$ is the change in the 3 component vector in the first column of S. The rotations R_l and R_r correspond to unitary operators U and V, as in (33), applied before the qubit enters and after it emerges from the channel. The geometrical significance of these rotations in terms of the Bloch sphere can be worked out using (32).

○ One can think of the action of S in (34) on the Bloch sphere in the following way, see (32). The c_j act to shrink each axis of the sphere by a corresponding factor, with $c_j < 0$ inverting the axis at the same time as shrinking it by a factor of $|c_j|$. This results in all the points originally inside the unit sphere being mapped to an ellipsoid centered at the origin with principal axes are along x, y, and z. Next, the action of the b_j is to shift the center of this ellipsoid from the origin to the point $\vec{b} = (b_1, b_2, b_3)$.

• We are now down to 6 parameters in (34). To make further progress, let us arbitrarily set all the b_j equal to zero, and call the result a *Pauli* 1 qubit channel. It is not unlike the symmetric 1 bit classical channel in which $\epsilon_1 = \epsilon_0$. It is characterized by three parameters c_1, c_2, c_3.

• This channel maps the identity operator at the input to the identity operator at the output. Such a channel is often referred to as a *unital* channel.

□ Exercise. Find a set of Kraus operators, each K_j proportional to the Pauli σ_j, that corresponds to a Pauli channel.

○ Each c_j falls in the interval

$$-1 \leq c_j \leq 1. \tag{35}$$

In addition there are restrictions imposed by the complete positivity condition (iv) on various sums of the c_j. Let us ignore these for the moment, and ask what interpretation we can give to (34) with $b_j = 0$.

• Suppose that we start with a pure input state $[z^+] = [0]$, or $x_0 = y_0 = 0$, $z_0 = 1$ in the Bloch sphere. Multiplying the corresponding column vector by S, one finds the output state ρ_1 corresponds to $x_1 = y_1 = 0, z_1 = c_3$ when the particle emerges from the channel. This corresponds to an ensemble of

$$\rho_1 = \left(\frac{1 + c_3}{2}\right)[0] + \left(\frac{1 - c_3}{2}\right)[1], \tag{36}$$

so it is like the classical 1 bit channel in (13), with $\epsilon_0 = \frac{1}{2}(1 - c_3)$. Similarly, if we start with an input state $[z^-] = [1]$, the result will be (36) with the coefficients of [0] and [1] interchanged, thus like a 1 bit channel with $\epsilon_1 = \frac{1}{2}(1 - c_3)$.

• Of course the same considerations apply to starting states of $[x^+]$ and $[x^-]$, with $[0]$ and $[1]$ in (36) replaced with $[x^+]$ and $[x^-]$, and c_3 with c_1. And similarly for $[y^+]$ and $[y^-]$.

◦ The situation with other starting states, say $[w^+]$ and $[w^-]$ with w a direction other than x, y or z is, in general, more complicated: one cannot express the output density operators in terms of $[w^+]$ and $[w^-]$ alone. The exceptions to "in general" arise when two, or perhaps three, of the c_j are identical.

• Now that we understand (to some extent) the situation with $b_j = 0$, let us remove that restriction. The conclusion is left as an exercise.

□ Exercise. Consider the case $b_3 \neq 0$ (its value must like between -1 and $+1$), and find ρ_1 for initial states $[0]$ and $[1]$. Show that one can again use (36) with modified coefficients, and the result resembles a classical biased ($\epsilon_1 \neq \epsilon_0$) channel. Do all possible combinations of values for b_3 and c_3 in the interval -1 to $+1$ make sense in terms of probabilities? (There are restrictions on these quantities due to the requirement that the superoperator be completely positive.)

★ By setting two of the c_j along with the corresponding b_j equal to 0, say $c_1 = 0 = b_1$ and $c_2 = 0 = b_2$, one arrives at a one-qubit representation of a "classical" channel. To be sure, "classical" is not a precisely defined term in a quantum context, but a channel of this type comes close to exemplifying a situation in which "quantum" effects of nonorthogonal states play no role.

□ Exercise. Show that a "classical" channel of this type is produced by a circuit in which the a qubit representing the channel interacts with a one qubit environment, with initial state $|\hat{e}\rangle = |0\rangle$, through a controlled-not gate, with the a qubit the control. Show that S is of the form (34), and work out the b's and the c's.

4.4 Kraus representation of quantum operations

★ In QCQI the superoperator representing a quantum operation—i.e., satisfying conditions (i) to (iv) in Sec. 4.2— is discussed using Kraus operators $\{E_j\}$, denoted here by $\{K_k\}$, and takes the form

$$\mathcal{S}(A) = \sum_j K_k A K_k^\dagger. \tag{37}$$

• That (37) agrees with (27) can be seen by working out the latter using (7).

□ Exercise. Do it.

• The main advantage of using Kraus operators is that they ensure that $\mathcal{S}$ is completely positive, a condition which is hard to check in terms of the matrix S_{jk} of (26).

• The main disadvantage of Kraus operators is that they are not unique. A given superoperator can, in general, be represented by many different collections of Kraus operators. As pointed out in QCQI, any two such collections are related to each other by a unitary matrix, but this fact does not make it easy to check equivalence.

★ The reason the Kraus representation is not unique is that the choice of orthonormal basis $\{|f^k\rangle\}$ in (7) is not unique. Different bases give rise to different collections of Kraus operators. Since the choice of the basis of f has no influence on b, and we are only interested in how b is related to a in Fig. 1, the Kraus operators provide information on correlations between the channel output and the environment, information that is superfluous from the point of view of a quantum operation.

4.5 Transition operator and dynamical operator

★ A superoperator $\mathcal{S}$ mapping $\hat{\mathcal{H}}_a$ to $\hat{\mathcal{H}}_b$ can be represented in the form

$$\mathcal{S}(A) = \mathrm{Tr}_a\Big((A \otimes I_b)Q\Big). \tag{38}$$

where the operator Q on $\mathcal{H}_a \otimes \mathcal{H}_b$ is known as the *transition operator*.

○ There is a one-to-one correspondence between the transition operator Q and the superoperator $\mathcal{S}$, unlike the many-to-one correspondence between collections $\{K_k\}$ of Kraus operators and $\mathcal{S}$.

• If $\mathcal{S}$ maps Hermitian operators to Hermitian operators, Q is Hermitian. If $\mathcal{S}$ is trace preserving then

$$\mathrm{Tr}_b(Q) = I_a. \tag{39}$$

★ Complete positivity of $\mathcal{S}$ is equivalent to simple positivity for the operator R which is the *partial transpose* of Q relative to some orthonormal basis $\{|a^j\rangle\}$ of $\mathcal{H}_a$, in the sense that

$$\langle a^j b^p|R|a^k b^q\rangle = \langle a^k b^p|Q|a^j b^q\rangle, \tag{40}$$

where $\{|b^p\rangle\}$ is an arbitrary orthonormal basis of $\mathcal{H}_b$.

• The operator R obtained in this way depends upon the choice of the basis $\{|a^j\rangle\}$ (but not the choice of $\{|b^p\rangle\}$), which is a bit of an annoyance. However, whether or not the partial transpose is positive, for a given transition operator Q, does not depend upon the choice of $\{|a^j\rangle\}$. R is known as the *dynamical operator*.

• Transition operators have the advantage that the complete positivity of $\mathcal{S}$ is (fairly) easily checked, and there is a unique transition operator associated with any particular superoperator. The main disadvantage is that the transition operator does not provide as simple an intuitive picture as the matrix S_{jk} of (26) or (30). Also, transition and dynamical operators are not at present widely employed in the research literature.

□ Exercise. If $\mathcal{H}_a$ and $\mathcal{H}_b$ are both two-dimensional spaces, any operator on $\mathcal{H}_a \otimes \mathcal{H}_b$ can be written in the form

$$Q = \sum_{j=0}^{3}\sum_{k=0}^{3} C_{jk}\sigma_{aj} \otimes \sigma_{bk} \tag{41}$$

with suitable (in general complex) coefficients C_{jk}. Find these coefficients in the case where Q is the transition operator corresponding to $\mathcal{S}$ with matrix S given by (34).

5 POVMs

5.1 Definition

• Refer to Fig. 1, where $|\hat{e}\rangle$ and T, or J, are considered fixed, whereas $|\psi\rangle$ is variable, and ask the question: "what can we learn about $|\psi\rangle$ from a simple measurement on f if we ignore, or know nothing about b?" The answer is provided through the concept of a POVM, a "positive operator-valued measure." (The origin of this term in the study of infinite-dimensional Hilbert spaces need not concern us.)

★ We define a POVM as a collection $\{G_k\}$ of *positive* operators on a Hilbert space that sum to the identity:

$$\sum_k G_k = I. \tag{42}$$

∘ This definition is satisfactory for a finite-dimensional Hilbert space, and we shall not need to consider more complicated situations appropriate to infinite-dimensional Hilbert spaces.

∘ Recall that G_k positive, $G_k \geq 0$, means that G_k is Hermitian with no negative eigenvalues, or, equivalently, that $\langle\psi|G_k|\psi\rangle \geq 0$ for all $|\psi\rangle$.

∘ It is convenient to exclude the trivial case in which $G_k = 0$ for some k.

• If each G_k is a projector and (42) is satisfied, then one can show that the $\{G_k\}$ form a decomposition of the identity, i.e., they are mutually orthogonal to each other. Such a decomposition is thus an example of a POVM. But in general the POVM operators are not projectors.

∘ Because the operators in a POVM sum to the identity one can also refer to the collection as a "decomposition of the identity." To avoid confusion the term "projective decomposition" is sometimes used for the case in which all the G_k are projectors.

★ POVMs are used to generate probabilities using the formula

$$p_k = \mathrm{Tr}(\rho G_k) = \mathrm{Tr}(G_k \rho), \tag{43}$$

where ρ is the density operator (thought of as a pre-probability) for the system of interest to us. Note that if $\rho = |\psi\rangle\langle\psi|$, the right side of (43) is $\langle\psi|G_k|\psi\rangle$.

• In standard textbook quantum mechanics carried out by people who have thought carefully about the subject, p_k is always considered to be the probability for some macroscopic state of affairs (e.g., "the pointer indicates $k = 3$") at the end of an experiment which began with a "preparation" resulting in $|\psi\rangle$ or ρ. While a POVM is often referred to as a "measurement," it is, at least within this framework, difficult to justify that term, since it is not clear that any property of the microscopic system is actually being measured. We shall come back to the problem of physical interpretation later.

★ Returning to Fig. 1. The probability that f is in the state $|f^k\rangle$ (as revealed by a simple measurement) is, by combining (4) and (6),

$$p_k = \langle\psi|K_k^\dagger K_k|\psi\rangle. \tag{44}$$

□ Exercise. Show that any operator of the form $K^\dagger K$ is positive; i.e., for any $|\psi\rangle$ in the Hilbert space on which K acts it is the case that $\langle\psi|K^\dagger K|\psi\rangle \geq 0$.

• Consequently, in light of (9) the collection $\{G_k = K_k^\dagger K_k\}$ constitutes a POVM, and this POVM can be used to address the question of what a measurement of f *after* the interaction represented by T in Fig. 1 tells us about the state of the system a *before* the interaction. But what does it tell us? In particular, what do we learn if the outcome of our simple measurement is k, which is to say f was in the state $|f^k\rangle$ at the end of the interaction?

★ In the particular case in which G_k is a rank-one operator, which means (because it is positive) that it is proportional to some $|\omega\rangle\langle\omega|$ for some $|\omega\rangle$, the following interpretation can be justified by a more detailed analysis (using consistent histories). In the particular experiment in which the final outcome was k, the system a was in the physical state $|\omega\rangle$ just before it interacted with e.

• But does this make sense? Suppose that $G_1 = |\omega^1\rangle\langle\omega^1|$ and $G_2 = |\omega^2\rangle\langle\omega^2|$, and that $|\omega^1\rangle$ and $|\omega^2\rangle$ are incompatible. Then surely it does not make sense to ask if the system was in one state or the other?

• The solution to this problem is to stress the qualification, "In the particular experiment in which the final outcome was k." Think of the final experimental outcome as a *label* on the earlier state, somewhat analogous to the labels introduced earlier for the states of an ensemble.

○ But doesn't this mean that somehow the future (outcome of the experiment) is mysteriously influencing the past (state of the quantum system)? No. See the comments in CQT towards the bottom of p. 200.

★ When G_k has rank greater than one but is not a projector it is not so clear what one learns from the POVM when the outcome is k. One can assert that in this instance the system a at the time of interest had the property given by the *support* of G_k: the smallest projector P such that $PG_k = G_k$, see CQT, p. 44. This, however, need not be very informative. In particular there are cases in which the support of G_k is I_a, in which case the fact that a had this (always true) property tells us nothing.

• Given the generality allowed in the definition of a POVM, this lack of clarity is not too surprising. Think of a situation in which two vehicles collide and a wheel spins off of one of them. What does this tell one about the state of affairs before the collision? Probably it says something, but it might be rather difficult to say just what it is.

○ For this reason it is a bit odd to refer to a general POVM as a "measurement." However, this terminology is by now embedded in the literature, and the concept of a POVM turns out to be extremely useful in many circumstances. Let us refer to it as a "complex measurement", or "POVM measurement," in contrast to the simple measurement introduced earlier, which detects a specific quantum property the system had just before the measurement took place.

★ In the special case in which the G_k are projectors, one refers to the POVM as a *projective measurement.* In this situation the term "measurement" is justified by the fact that when the outcome is k, the system of interest had the property represented by the projector G_k just before the measurement took place.

5.2 Example of a POVM

★ Consider the following quantum game. Alice prepares a qubit randomly in one of the three states $|\psi_j\rangle$, where

$$|\psi_1\rangle = \big(|0\rangle + |1\rangle\big)/\sqrt{2}, \quad |\psi_2\rangle = \big(|0\rangle + e^{i2\pi/3}|1\rangle\big)\sqrt{2}, \quad |\psi_3\rangle = \big(|0\rangle + e^{-i2\pi/3}|1\rangle\big)/\sqrt{2}, \tag{45}$$

and sends it to Carol, who on the basis of whatever measurements she wants to carry out is to specify a number $k = 1$ or 2 or 3 in such a way that k is *not* equal to j. Let us say that Carol wins \$1 every time k is unequal to j, but loses \$10 every time $k = j$.

• It is hard to think of a simple measurement strategy which will work. For example, suppose Carol measures the qubit in the S_x basis. If the outcome is $-1/2$, meaning the state orthogonal to $|\psi_1\rangle$, she is sure that Alice did *not* send $|\psi_1\rangle$, so it is safe to set $k = 1$. But suppose the measurement outcome is $S_x = +1/2$. This could mean that Alice sent $|\psi_1\rangle$, in which case $k = 2$ or $k = 3$ would be appropriate choices, but if Alice had sent $|\psi_2\rangle$ or $|\psi_3\rangle$, there is a finite (positive) probability that Carol will obtain $S_x = +1/2$, and so she could possibly lose by specifying $k = 2$ or $k = 3$.

□ Exercise. Work out the probabilities assuming a measurement of the sort just mentioned. Will Carol win or lose money in the long run if she makes optimum use of the information she obtains?

★ A POVM provides a better strategy. Choose $|\omega_k\rangle$ with $k = 1, 2, 3$ to be *orthogonal* to $|\psi_k\rangle$, and $G_k = |\omega_k\rangle\langle\omega_k|$ up to a constant of proportionality. In this case if Carol's measurement outcome is k, she knows that Alice did *not* send $|\psi_k\rangle$, so specifying the POVM output will win the game every time.

○ To check that this strategy really works, a certain number of details need to be worked out, see the following exercise.

□ Exercise. Assuming that the kets $|\omega_k\rangle$ are normalized, find the appropriate constant of proportionality relating G_k to $|\omega_k\rangle\langle\omega_k|$, and show that $\sum_k G_k = I$.

□ Exercise. The states in (45) lie on the equator of the Bloch sphere. If, instead, they all lie somewhere in the northern hemisphere, say at the same latitude and separate by 120° in longitude, the POVM strategy will no longer work, or at least it won't work perfectly. Can you explain why? [Hint. If you think of the Bloch ball as filled with density operators, projectors of rank 1 reside on the surface, and $I/2$ is at the origin. Give a geometrical interpretation to $\sum_k G_k = I$.]

★ How does Carol actually carry out the POVM? She has a well-equipped quantum laboratory, and when she receives the qubit a from Alice, she prepares a system e, known in this context as an *ancillary system* or *ancilla*, in a known state $|\hat{e}\rangle$ and lets it interact in a suitable way, represented by T in Fig. 1 with a. Then she carries out a simple measurement on f corresponding either to an orthonormal basis or some other decomposition of the identity.

• In fact f may be nothing but the combined system a and e after the interaction. (This corresponds to a one-dimensional b in Fig. 1, meaning the complex numbers; one-dimensional subsystems entering a tensor product are always trivial.)

5.3 Naimark extension

★ The Naimark extension (or dilation) theorem says that any POVM measurement can be realized as a projective measurement on a space of higher dimension, where "realized" means that the probabilities of the measurement outcomes are the same.

• Probably the simplest way to think about it is that used for the example in Sec. 5.2. In addition to the system to be measured with Hilbert space $\mathcal{H}_s$ there is an ancillary system $\mathcal{H}_r$ which starts of in a known pure state $|r_0\rangle$. Let $\mathcal{H}_t = \mathcal{H}_s \otimes \mathcal{H}_r$ be Hilbert space for the two systems together. Then there is a projective decomposition $\{P_k\}$ of the identity I_t on $\mathcal{H}_t$ such that for any density operator ρ_s on $\mathcal{H}_s$ it is the case that

$$\mathrm{Tr}_s\Big(G_k\rho_s\Big) = \mathrm{Tr}_t\Big(P_k(\rho_s \otimes [r_0])\Big) \tag{46}$$

○ That is to say, the P_k are chosen so that the partial trace

$$\mathrm{Tr}_r\Big(P_k\Big) = G_k \tag{47}$$

gives the desired answer.

• An alternative way to think about the Naimark extension is to suppose that the original Hilbert space $\mathcal{H}_s$ is a subspace of a suitably-chosen larger Hilbert space $\mathcal{H}_t$, and that E is a projector on $\mathcal{H}_t$ that projects onto $\mathcal{H}_s$. Then if $\mathcal{H}_t$ is large enough, there is a projective decomposition $\{P_k\}$ of its identity such that

$$G_k = EP_kE \tag{48}$$

for every k. As defined in (48) this G_k acts on the entire space $\mathcal{H}_t$, but since it gives zero when applied to any ket in the orthogonal complement of the subspace $\mathcal{H}_s$, it can just as well be thought of as acting on $\mathcal{H}_s$.

• Proofs of these results are not altogether straightforward. See Sec. 9-6 of Peres for a derivation of the Naimark result.