QUANTUM MEASUREMENTS

Lecture notes for WAQCT Summer School 21-26 August 2019

Erika Andersson, Heriot-Watt University, Edinburgh, UK

Contents

1. Introduction

Measurements are what links the world of quantum mechanics to the "classical" world. Through measurements, we extract information about quantum systems. A standard way to introduce quantum measurements in undergraduate textbooks is as a projection in the eigenbasis of some observable. We shall first describe such projective quantum measurements, and then discuss why they cannot describe all measurements we can make on a quantum system. This leads us to consider generalised quantum measurements, also called Probability Operator Measures (POMs) or Positive Operator-Valued Measures (POVMs). These can be viewed and realised as projective measurements on an enlarged quantum system. We will also describe how a quantum state is changed by a measurement. As an important application of generalised quantum measurement, we will discuss quantum state discrimination.

Neither projective nor generalised measurements solve the "measurement problem", that is, what really happens when a quantum system interacts with the measurement apparatus and one somehow obtains classical information about the system. In particular, quantum measurements are probabilistic. We can in general only predict probabilities for experimental outcomes, not exactly what will occur in individual cases. Nevertheless, if we accept this, generalised quantum measurements can describe any measurement we can make on a quantum system. Conversely, any generalised measurement (that can be written down "theoretically") can also be realised. This means that we can first find the optimal measurement theoretically, without worrying too much about how to realise it. Once we know the theoretical description, there is a "recipe" for figuring out how to realise it on the physical system at hand. This makes generalised measurements a very useful theoretical tool. Realisations using optical networks are well suited for distinguishing between quantum states encoded in photons, and this will be described as an example of how to construct a physical realisation of a generalised measurement. The same principles can however be applied also to other systems, including trapped ions or superconducting qubits.

For an excellent introduction to generalised quantum measurements (and quantum information science generally), see [1]. There are many useful and accessible review papers on state discrimination specifically, for example [2, 3, 4, 5]. The description of how to realise generalised measurements e.g. with optical networks hasn't made its way to any review papers yet, as far as the author knows.

2. Projective or von Neumann measurements

A standard way to describe quantum measurements is as a projection in the eigenbasis of some observable. If an observable is described by the Hermitian operator $\hat{A}$, then we can express $\hat{A}$ using its eigenvalues λ_i and eigenvectors $|v_i\rangle$ as

$$\hat{A} = \sum_i \lambda_i |v_i\rangle\langle v_i|. \tag{1}$$

The eigenvalues λ_i are the values that a measurement of $\hat{A}$ on an individual quantum system may give, that is, the measurement outcomes. The measurement operator associated with a measurement outcome λ_i is the projector $\hat{P}_i = |v_i\rangle\langle v_i|$ onto the corresponding eigenstate, and the probability for outcome λ_i is

$$p(\lambda_i|\hat{\rho}) = \mathrm{Tr}\left(\hat{P}_i\hat{\rho}\right), \tag{2}$$

where $\hat{\rho}$ is the density matrix for the measured state. If an ensemble of quantum systems is measured, then the average or expectation value for the observable $\hat{A}$ is given by

$$\langle \hat{A} \rangle = \mathrm{Tr}\left(\hat{A}\hat{\rho}\right) = \sum_i \lambda_i \mathrm{Tr}\left(|v_i\rangle\langle v_i|\hat{\rho}\right) = \sum_i \lambda_i p(\lambda_i|\hat{\rho}), \tag{3}$$

where $\hat{\rho}$ describes the ensemble.

It is usually stated that after a projective measurement, the quantum system is left in the (non-degenerate) eigenstate corresponding to λ_i, so that after the measurement, $\rho \to \hat{\rho}'_i = |v_i\rangle\langle v_i|$. If λ_i corresponds to more than one eigenstate, then the density matrix is transformed as

$$\rho \to \hat{\rho}'_i = \frac{\hat{P}_i\hat{\rho}\hat{P}_i}{\mathrm{Tr}(\hat{P}_i\hat{\rho}\hat{P}_i)} \tag{4}$$

(which obviously also holds if there is only one eigenstate). We recognise the denominator as the probability to obtain result i; this makes sure that the post-measurement state is normalised.

One easily realises that Equation (4) might not always be true for real measurements. Typically, an experimental realisation of a measurement changes or destroys the measured quantum system in a different way than leaving it in the eigenstate corresponding to the obtained measurement result. For example, detecting an atom or ion in an excited state often involves detecting the emission from when it becomes de-excited, and unless the atom or ion is re-excited, it is left in a lower-lying state. Similarly, when detecting light, it is usually absorbed by the detector, and the post-measurement state is really the vacuum state rather than, say, a number state or a quadrature eigenstate. It would be possible to modify the description of projective measurements, keeping their projective character, but allowing for other post-measurement states. We will not do this, however, since we will describe a more general kind of measurement in the next section.

3. Generalised quantum measurements

3.1. **Probability operator measures.** It turns out that projective quantum measurements are too restrictive. They cannot conveniently describe all measurements that can be realised in actual experiments, as will become clear. We can conceive of more general measurements, so-called probability operator measures (POMs) or positive operator-values measures (POVMs). These generalised quantum measurements can describe any measurement that is possible within the framework of quantum mechanics.

What properties would we like a description of quantum measurements to have? Suppose that a measurement should be described by a set of measurement operators $\hat{\Pi}_i$, one corresponding to each outcome of the measurement. We also want to keep the trace rule, that is, the probabilities for the outcomes should be given by

$$p(i|\hat{\rho}) = \mathrm{Tr}\left(\hat{\Pi}_i\hat{\rho}\right), \tag{5}$$

for any measured state $\hat{\rho}$. We will here describe the finite-dimensional case, but the treatment can also be extended to hold for infinite dimensions and continuous variables.

It is natural to demand that probabilities should be real and nonnegative, and that if we add up the probabilities for all outcomes that can occur, the sum should be 1. These two conditions should hold for any measured state $\hat{\rho}$. It follows that the operators $\hat{\Pi}_i$ must be Hermitian, and must satisfy the conditions

$$\hat{\Pi}_i \geq 0 \tag{6}$$

$$\sum_i \hat{\Pi}_i = \hat{\mathbb{I}}, \tag{7}$$

where $\hat{\mathbb{I}}$ is the identity operator. The first condition, $\hat{\Pi}_i \geq 0$, means that all eigenvalues of $\hat{\Pi}_i$ are greater than or equal to zero. If this were not the case, then when the measured state is an eigenstate of $\hat{\Pi}_i$ with a negative eigenvalue, the corresponding probability for that outcome would be negative. To obtain the second condition, we sum the probabilities for all outcomes,

$$1 = \sum_i p_i = \sum_i \mathrm{Tr}\left(\hat{\Pi}_i \hat{\rho}\right) = \mathrm{Tr}\left(\sum_i \hat{\Pi}_i \hat{\rho}\right). \tag{8}$$

Since this has to be equal to 1 no matter what $\hat{\rho}$ is, (7) must hold. Note that this condition, that the total probability for all outcomes is 1, means that we should include cases where "nothing happens" as an outcome. For example, when no detector clicks because of losses. Such "non-events" will also have a corresponding probability and measurement operator.

Conditions (6) and (7) also hold for projective measurements, but in addition, the measurement operators have to be projectors. While this can serve as a starting point for describing quantum measurements, its physical motivation is not very clear. For generalised quantum measurements, the measurement operators are not necessarily projectors. As will be shown below in section 5, however, if we couple the quantum system to be measured to an ancillary system, and then perform a projective measurement on the original system plus ancillary system, then the result is a generalised quantum measurement of the original system. Conversely, any generalised quantum measurement can be realised by suitably enlarging the Hilbert space of the measured system, for example by coupling it to an ancillary quantum system which may be part of a measuring device, and performing a projective measurement in the enlarged space. One might argue, therefore, that we could keep the description with only projective measurements. The coupling to the extra degrees of freedom may however be done in many different ways for the same generalised quantum measurement, and exactly how a particular realisation is made often depends on what is possible for the actual physical system at hand. In other words, the same generalised measurement may be realised in many different ways. It is convenient to have a description of the measurement that is independent of a particular realisation. It may also be useful to first theoretically optimise the generalised measurement, without referring to a realisation. Once the optimal measurement is clear, one can think of how to best implement it in practice. This makes generalised quantum measurements a useful description.

How then does the measured state change in a generalised quantum measurement? If outcome i is obtained, then $\hat{\rho}$ changes as

$$\hat{\rho} \to \hat{\rho}_i = \frac{\hat{A}_i \hat{\rho} \hat{A}_i^\dagger}{\mathrm{Tr}(\hat{A}_i \hat{\rho} \hat{A}_i^\dagger)}, \tag{9}$$

where the denominator is the probability to obtain result i, making sure that the post-measurement state is normalised, and the so-called Kraus operators $\hat{A}_i$ obey $\hat{\Pi}_i = \hat{A}_i^\dagger \hat{A}_i$. This condition is satisfied for any $\hat{A}_i = \hat{U}_i \Pi_i^{1/2}$, where the $\hat{U}_i$ are arbitrary unitary operators. (The square root of Π_i is defined as $\Pi_i^{1/2} = \sum_j \sqrt{\lambda_j} |v_j\rangle\langle v_j|$, where λ_j and $|v_j\rangle$ are the eigenvalues and eigenstates of Π_i.) Depending on how we choose the $\hat{U}_i$, equation (9) can cover any physical situation, such as a de-excited atom that is left in the ground state regardless of what the measurement outcome was, or light that is absorbed by a detector, resulting in a post-measurement state equal to the vacuum.

If we do not learn the outcome of the measurement, then the post-measurement state is a statistical mixture of the different post-measurement $\hat{\rho}_i$, with respective probabilities,

$$\hat{\rho} \to \hat{\rho}' = \sum_i \hat{A}_i \hat{\rho} \hat{A}_i^\dagger. \tag{10}$$

The reader may recognise this as a the Kraus form of a completely positive trace-preserving (CPTP) map. Such maps describe the most general transformations that quantum states can undergo. We can think of them as what happens to a quantum system if it is coupled to another quantum system by a unitary transform, and we then trace out the other system. This is closely related to how we can realise generalised measurements.

3.2. **An example: imperfect photon counting.** We will now give an example, due to Steve Barnett, that shows how the formalism of generalised quantum measurements is useful for describing experimental imperfections. Suppose that we have an imperfect photodetector, where each incoming photon is detected with probability η and undetected with probability $1-\eta$. The detector in this example does have photon number resolution. On the other hand, we are not considering the possibility of dark counts, that is, that the detector might fire when there are no photons present. (Most photodetectors are "bucket detectors" which only tell us either that there are "no photons detected" or "some photons detected". We could easily have picked that as an example too, and included dark counts.)

If there are m photons present in the incoming light, the probability $p_{\text{det}}(n|m)$ to detect n of these depends on the detector. We will also be able to deal with incoming states of light other than photon number states, but as usual in quantum mechanics, it is sufficient to know what happens for some complete set of basis states. For the detector we are considering, we have that

$$p_{\text{det}}(n|m) = \eta^n (1-\eta)^{m-n} \binom{m}{n}, \tag{11}$$

since the probability to detect n photons is η^n, the probability not to detect the remaining $m-n$ photon is $(1-\eta)^{m-n}$, and the last factor arises since there are $\binom{m}{n}$ ways to choose n out of the m photons present. If p_m is the probability that the incoming light has m photons present, then the overall probability to detect n photons is

$$p_{\text{det}}(n) = \sum_{m=n}^{\infty} p(n|m) p_m = \sum_{m=n}^{\infty} \eta^n (1-\eta)^{m-n} \binom{m}{n} p_m. \tag{12}$$

If we would like to write this as $p_{\text{det}}(n) = \text{Tr}\,[\Pi_n \hat{\rho}]$, where $\hat{\rho}$ is the density matrix describing the incoming light, and Π_n is a measurement operator characterising the detector, then it is easy to show that

$$\Pi_n = \sum_{m=n}^{\infty} \eta^n (1-\eta)^{m-n} \binom{m}{n} |m\rangle\langle m| \tag{13}$$

does the job. The Π_n are not orthogonal projectors, but mixtures of un-normalised projectors. They do sum to the identity operator, and they are of course positive (more precisely, positive semi-definite, that is, their eigenvalues are positive or zero). In the limit when $\eta = 1$, the measurement is a perfect projection in the number-state basis. But otherwise, there is a possibility to detect fewer photons than were actually present, and therefore the measurement operators must reflect this possibility. That is, if we want to describe a lossy detector, we are naturally led to non-orthogonal measurement operators. This is true in general. In order to describe the imperfections present in any experimental measurement, generalised quantum measurements arise instead of projective measurements.

4. Identifying quantum states

We have seen that imperfect projective measurements are conveniently described in terms of POMs. There are also situations, however, where we deliberately might want to use a generalised measurement rather than a projective measurement, not just because of experimental limitations. In many cases, a generalised measurement is better than a projective measurement at obtaining exactly the information we want about the measured system. One example is when we want to unambiguously distinguish between two non-orthogonal quantum states. This is a special case of quantum state discrimination [6], which we will give a brief introduction to in this section. Quantum state discrimination can also be called "quantum hypothesis testing". We want to test different hypotheses about the quantum system that is measured to try to say what the best guess is. For more detailed reviews of this subject, see [2, 3, 4, 5].

Identifying quantum states is an important application of generalised quantum measurements. In a communication situation, for example, the receiver has to distinguish between different signal states. These states may have been affected by noise during preparation, transmission and detection, and this will affect what measurement the receiver should implement. Even for signal states that initially were orthogonal, so that they in principle could be distinguished perfectly, they may not be orthogonal when they reach the receiver. Sometimes one uses non-orthogonal quantum states on purpose, for example in quantum key distribution [7]. Here it is precisely the fact that non-orthogonal quantum states cannot be distinguished perfectly that guarantees the security of the protocol. If we want to determine how well a quantum communication protocol can work, both how well a legitimate participant can do, and how well it is possible to cheat in a cryptographic protocol, then we need to know what the relevant optimal quantum measurements are.

It is also possible to draw parallels between quantum state discrimination and other tasks in quantum information science, such as cloning and entanglement purification. For example, bounds on the success probability for cloning a set of quantum states can be derived from the maximum success probability for state discrimination between the states is known [3]. A read-out measurement is also typically the endpoint of any quantum computation, and measurements can also be used during the computation e.g. for error correction.

Suppose then that the states that we want to distinguish between are described by $\hat{\rho}_i$, occurring with prior probabilities $p(\hat{\rho}_i) = p_i$, where $i = 1, 2, \ldots N$. When asking how well we can distinguish between these quantum states, we first have to specify what quantity we would like to optimise. One possible choice is to maximise the probability to obtain the correct result, resulting in a so-called *minimum-error* measurement. This can be generalised to measurements that give minimum average cost, where some errors may be worse than others so that there is a cost C_{ij} associated with obtaining outcome j when the state actually was $\hat{\rho}_i$ [6]. Another possibility is measurements that give *unambiguous* or error-free results, at the expense of sometimes not giving a result at all. Such measurements are not always possible, but we can always make a measurement that gives *maximum confidence* in the result. That is, whenever we do obtain a result, it has the maximum possible probability to be correct. Another figure of merit is *mutual information*, leading to measurements that maximise the information the result j contains about which state $\hat{\rho}_i$ was sent, measured e.g. in bits. A *maximum fidelity* measurement is again optimal when, based on the measurement result, we need to prepare a new copy of the state which matches the original state as well as possible in terms of fidelity. There is no limit on the number of copies we can prepare, since the measurement result is classical information, and therefore this type of measurement is one way to achieve cloning to infinitely many copies. It is also possible to come up with other figures of merit, or to interpolate between different figures of merit. For example, one may ask what the minimum possible error is, given a fixed maximum allowed probability for a measurement to fail to give result at all. Of these possibilities, we will look at minimum-error, unambiguous, maximum confidence and maximum mutual information measurements.

4.1. **Minimum-error measurements.** For a minimum error measurement, we want to maximise the probability to obtain the correct result. Since it always pays to make a guess, the measurement will always give one of the states as a result. Consequently there is one measurement operator $\hat{\Pi}_i$ corresponding to each state $\hat{\rho}_i$ – some of them may be zero operators – and no more measurement operators. If the prepared state is $\hat{\rho}_i$, then the probability for result j is, as usual, given by $p(j|\hat{\rho}_i) = \mathrm{Tr}\left(\hat{\Pi}_j\hat{\rho}_i\right)$. The average probability p_c to identify the state correctly is

$$p_c = \sum_i p_i p(i|\hat{\rho}_i) = \sum_i p_i \mathrm{Tr}\left(\hat{\rho}_i\hat{\Pi}_i\right), \tag{14}$$

where p_i, as before, is the prior probability for state ρ_i. Maximising p_c is equivalent to minimising the average probability to obtain the wrong result, that is, the error probability

$$p_e = 1 - p_c = \sum_{i,j:i\neq j} p_i p(j|\hat{\rho}_i) = \sum_{i,j:i\neq j} p_i \mathrm{Tr}\left(\hat{\rho}_i\hat{\Pi}_j\right), \tag{15}$$

and this is why the measurement is called a “minimum-error” measurement. Unambiguous and maximum confidence measurements, which we will consider next, actually have a smaller probability to give the wrong result, but these types of measurements do not always give a result and will have a lower average probability to correctly identify the state.

For a general set of states and prior probabilities, finding the optimal minimum error measurement analytically is very difficult. The best measurement may of course be found numerically if we specify states and prior probabilities (semi-definite programming is a useful tool). In any event, it is known that the optimal measurement must satisfy the conditions [6]

$$\begin{aligned} \hat{\Pi}_i\left(p_i\hat{\rho}_i - p_j\hat{\rho}_j\right)\hat{\Pi}_j &= 0 \ \forall\ i,j, & (16)\\ \hat{\Gamma} - p_i\hat{\rho}_i &\geq 0 \ \forall\ i, & (17)\end{aligned}$$

where the operator $\hat{\Gamma} = \sum_i p_i\hat{\Pi}_i\hat{\rho}_i$, so that $p_c = \mathrm{Tr}(\hat{\Gamma})$. Roughly speaking, condition (16) comes from the fact that the error probability should have an extremum for the best measurement strategy, and condition (17) comes from the fact that the extremum should be a minimum. A full derivation of these conditions is beyond the short treatment here, but Helstrom’s book [6] has a motivation in Chapters II and IV; the “quantum” case is analogous to the “classical” one. Unfortunately, these conditions are usually not too helpful in finding the optimal strategy. They can nevertheless be used to check whether a measurement is optimal if we can make an educated guess as to what the optimal measurement might be.

To prove that conditions (16) and (17) are sufficient to guarantee optimality, consider another measurement strategy with measurement operators $\Pi'_i, i = 1,\ldots,N$. The difference between the average probabilities to be correct for the two measurement strategies is then given by

$$p'_c - p_c = \sum_i p_i \mathrm{Tr}(\hat{\rho}_i\Pi'_i) - \mathrm{Tr}(\hat{\Gamma}) = -\mathrm{Tr}\left[\sum_i\left(\hat{\Gamma} - p_i\hat{\rho}_i\right)\Pi'_i\right], \tag{18}$$

where we used the fact that the Π_i' are complete, that is, they sum to the identity. The operators Π_i' are positive, and by condition (17), so are $\hat{\Gamma} - p_i\hat{\rho}_i$. Therefore $p_c' - p_c$ is negative, and the primed measurement strategy must have a smaller probability to be correct. For a proof that conditions (16) and (17) are also necessary for a measurement strategy to be optimal, see appendix I in [1]. The optimal measurement strategy may not be unique, but all optimal measurement strategies will have the same $\hat{\Gamma}$ operator.

One case where the optimal solution is known analytically is for distinguishing between two states $\hat{\rho}_1$ and $\hat{\rho}_2$, occurring with probabilities p_1 and p_2 [6]. The optimal measurement is then a projective measurement in the basis in which the operator

$$\hat{O} = p_1\hat{\rho}_1 - p_2\hat{\rho}_2 \tag{19}$$

is diagonal. If the result corresponds to an eigenvector with a positive eigenvalue, then the most likely state was $\hat{\rho}_1$, and if the result corresponds to a negative eigenvalue, then the most likely state was $\hat{\rho}_2$. If there are eigenvalues equal to zero, then if such an outcome is obtained, $\hat{\rho}_1$ and $\hat{\rho}_2$ are equally likely, and we may guess either $\hat{\rho}_1$ or $\hat{\rho}_2$ with whatever probabilities we like, for example, 1/2 each. The error probability for the optimal measurement is given by

$$p_e = \frac{1}{2}\left[1 - \sqrt{4p_1p_2\mathrm{Tr}\left(\hat{\rho}_1\hat{\rho}_2\right)}\right]. \tag{20}$$

As another example, consider distinguishing between three equiprobable symmetric states given by

$$|\psi_1\rangle = -|0\rangle,\ |\psi_2\rangle = \frac{1}{2}(|0\rangle + \sqrt{3})|1\rangle,\ |\psi_3\rangle = \frac{1}{2}(|0\rangle - \sqrt{3})|1\rangle, \tag{21}$$

where $|0\rangle$ and $|1\rangle$ are orthonormal basis states [6]. (A set of N states $\{|\Psi_1\rangle, |\Psi_2\rangle, \ldots, |\Psi_N\rangle\}$ is called symmetric if there is a unitary operation $\hat{U}$ so that $|\Psi_i\rangle = \hat{U}^{i-1}|\Psi_1\rangle$, with $\hat{U}^N = \mathbb{I}$.) It is easy to verify, using conditions (16) and (17), that the optimal measurement that distinguishes between these three states with minimum error has the measurement operators $\Pi_i = 2/3|\psi_i\rangle\langle\psi_i| = 2/3\rho_i$ for $i = 1, 2, 3$ [6]. This is an example where the optimal measurement has more measurement operators than there are dimensions in the state space; the measurement operators are unnormalised projectors.

4.2. **Unambiguous measurements.** One can also choose to make a measurement that never identifies a state incorrectly, at the cost of sometimes failing to give an answer. That is, we demand that the error probability $p_e = 0$, meaning that if $i \neq j$, then

$$p(i|\hat{\rho}_j) = \mathrm{Tr}\left(\hat{\Pi}_i\hat{\rho}_j\right) = 0. \tag{22}$$

Failing to give an answer corresponds to a measurement operator $\Pi_?$. For optimal unambiguous state identification, the success probability

$$p_c = \sum_i p_i p(i|\hat{\rho}_i) = \sum_i p_i \mathrm{Tr}\left(\hat{\Pi}_i\hat{\rho}_i\right) \tag{23}$$

is maximised, subject to the condition $p_e = 0$. Equivalently, the probability to give an inconclusive answer,

$$p_? = \sum_i p_i p(?|\hat{\rho}_i) = \sum_i p_i \mathrm{Tr}\left(\Pi_?\hat{\rho}_i\right) = 1 - p_c, \tag{24}$$

is minimised. Unambiguous measurements have a lower probability to give a correct result than minimum-error measurements, but in return they guarantee that an obtained result is correct.

Finding optimal unambiguous measurement strategies is, at least for sets of pure states, somewhat easier that finding minimum-error measurement strategies, due to the simplicity of condition (22). For a set of pure states, unambiguous state identification is possible with a non-zero success probability if and only if the states $\hat{\rho}_i$ are linearly independent [8]. For the three symmetric states in equation (21), for example, unambiguous discrimination is impossible, no matter what their (non-zero) prior probabilities are. Unambiguous discrimination between mixed states is also possible. It is easy to realise that in general, for us to be able to unambiguously identify a state $\hat{\rho}_i$ with non-zero probability, this state must have a component which is orthogonal to all the other possible states. The analytical solution for the optimal measurement for a general set of mixed states is difficult, but one can obtain results for special cases [5, 2].

As an example of an unambiguous measurement, suppose that we want to distinguish between two non-orthogonal pure states $|a\rangle$ and $|b\rangle$ which occur with prior probabilities p_a and p_b. We could choose to make a projective measurement in the basis $\{|a\rangle, |a^\perp\rangle\}$, where $|a^\perp\rangle$ is orthogonal to $|a\rangle$. If the outcome corresponding to $|a^\perp\rangle$ is obtained, then the state must have been $|b\rangle$. The success probability for this strategy would be $p_c = |\langle a^\perp|b\rangle|^2 p_b$, and the measurement operators are $\Pi_a = 0$, $\Pi_b = |a^\perp\rangle\langle a^\perp|$, and $\Pi_? = |a\rangle\langle a|$. Similarly, we might choose to measure in the basis $\{|b\rangle, |b^\perp\rangle\}$, with a success probability $p_c = |\langle b^\perp|a\rangle|^2 p_a$.

Neither of these measurements are optimal in general. They sometimes are, depending on the prior probabilities of $|a\rangle$ and $|b\rangle$. But when the prior probabilities p_a and p_b are similar, then the optimal unambiguous measurement will have three measurement operators. The measurement operators are in general given by

$$\begin{aligned} \Pi_a &= k_a|b^\perp\rangle\langle b^\perp| \\ \Pi_b &= k_b|a^\perp\rangle\langle a^\perp| \\ \Pi_? &= \hat{\mathbb{I}} - \Pi_a - \Pi_b. \end{aligned} \tag{25}$$

The nonnegative constants of proportionality k_a and k_b depend on the prior probabilities p_a and p_b and on the overlap $\langle a|b\rangle$, and must be chosen so that $\Pi_? \geq 0$. For equal prior probabilities, the minimum possible probability for the inconclusive outcome can be shown to be $p_? = |\langle a|b\rangle|$, so that the maximum possible success probability is given by $p_c = 1 - |\langle a|b\rangle|$, and $k_a = k_b = (1 + |\langle a|b\rangle|)^{-1}$. Below in section 5.5 we'll see how to realise this measurement for polarisation states of a photon, and also for two coherent states $|\alpha\rangle$ and $|\beta\rangle$.

4.3. **Maximum confidence measurements.** As mentioned above, if a set of states is not linearly independent, then it is not possible to distinguish unambiguously between them. We can however still maximise the conditional probability $p(\hat{\rho}_i|i)$ that an obtained outcome is correct, that is, the probability that if we obtain result "i", the prepared state really was $\hat{\rho}_i$ [9]. For an unambiguous measurement, $p(\hat{\rho}_i|i) = 1$ holds.

It is important to realise that in general $p(i|\hat{\rho}_i) \neq p(\hat{\rho}_i|i)$. Using Bayes' rule,

$$p(a, b) = p(a)p(b|a) = p(b)p(a|b), \tag{26}$$

we can write

$$p(\hat{\rho}_i|i) = \frac{p(i|\hat{\rho}_i)p_i}{p(i)} = \frac{\mathrm{Tr}\left(\hat{\Pi}_i\hat{\rho}_i\right)p_i}{\mathrm{Tr}\left(\hat{\rho}\hat{\Pi}_i\right)}, \tag{27}$$

where, as before, we want to distinguish between the states $\hat{\rho}_i$ which have prior probabilities p_i, so that $\hat{\rho} = \sum_i p_i\hat{\rho}_i$ is the prior density matrix.

For a set of pure states $|\Psi_i\rangle$ occurring with prior probabilities p_i, it can be shown that the maximum confidence measurement operators will be given by [9]

$$\hat{\Pi}_i = k_i\hat{\rho}^{-1}|\Psi_i\rangle\langle\Psi_i|\hat{\rho}^{-1}. \tag{28}$$

The nonnegative constants of proportionality k_i should be chosen so that

$$\Pi_? = \hat{\mathbb{I}} - \sum_i \hat{\Pi}_i \geq 0, \tag{29}$$

but can otherwise be chosen arbitrarily, since we do not care how seldom a result other than "?" is obtained, only whether it is correct or not when it is actually obtained. We may of course choose the constants of proportionality so that the probability associated with $\Pi_?$ is as small as possible. In analogy with optimal unambiguous measurements, this is called an optimal maximum confidence measurement.

For a set of mixed states, the optimal measurement operators will have the property

$$\hat{\Pi}_i \propto \hat{\rho}^{-1/2}|\lambda_i^{max}\rangle\langle\lambda_i^{max}|\hat{\rho}^{-1/2}, \tag{30}$$

where $|\lambda_i^{max}\rangle$ is the eigenket of the operator

$$\hat{\rho}_i' = \frac{\hat{\rho}^{-1/2}\hat{\rho}_i\hat{\rho}^{-1/2}}{\mathrm{Tr}\,(\hat{\rho}_i\hat{\rho}^{-1})} \tag{31}$$

corresponding to the largest eigenvalue of this operator. Again, the constants of proportionality can then also be chosen so that the probability for an inconclusive outcome is as small as possible.

4.4. **Maximum mutual information.** Especially in a communication situation, it makes sense to choose the measurement so that one obtains maximum information about which state was sent. Information is measured by entropy-like quantities. The mutual information quantifies how much information the measurement result j gives about which state $\hat{\rho}_i$ was actually sent, and is given by

$$I(i:j) = \sum_{ij} p(i,j)\log\left(\frac{p(i,j)}{p_i p_j}\right) = \sum_{ij} p_i \mathrm{Tr}(\hat{\Pi}_j\hat{\rho}_i)\log\left(\frac{\mathrm{Tr}(\Pi_j\rho_i)}{\mathrm{Tr}(\hat{\Pi}_j\rho)}\right), \tag{32}$$

where the states $\hat{\rho}_i$ occur with prior probabilities p_i, result j corresponds to the measurement operator $\hat{\Pi}_j$, and where $\hat{\rho} = \sum_i p_i\hat{\rho}_i$. (Note that this is information about *which* of the states $\hat{\rho}_i$ was sent, not about *what* the possible states $\hat{\rho}_i$ are, which is assumed to be prior knowledge when designing the measurement strategy.)

Due to the logarithm in the definition of mutual information, finding measurements that optimise this quantity is usually more difficult than finding other kinds of optimal measurements. Nevertheless, the optimal measurement is known analytically in some special cases, such as for two non-orthogonal states and for sets of equiprobable symmetric states. As an example, consider again the three equiprobable symmetric states in equation (21). The optimal maximum mutual information measurement strategy for these states has the measurement operators

$$\begin{aligned} \Pi_1 &= \frac{2}{3}|\psi_1^\perp\rangle\langle\Psi_1^\perp| \\ \Pi_2 &= \frac{2}{3}|\psi_2^\perp\rangle\langle\Psi_2^\perp| \\ \Pi_3 &= \frac{2}{3}|\psi_3^\perp\rangle\langle\Psi_3^\perp|, \end{aligned} \tag{33}$$

where $|\psi_i^\perp\rangle$ is a state orthonormal to $|\psi_i\rangle$. This measurement rules out one of the states perfectly, leaving the remaining two equally likely. In this case, ruling out one state completely gives more information than trying to identify which single state was the most likely one. (But it is not the case that maximising the mutual information in general means ruling out states. It just happens in this particular example.)

The maximum mutual information one then obtains, when given one copy of one of the states in (21), is ≈ 0.67. Note that this is *less* than one bit. If we could perfectly distinguish between the three equiprobable states, then we'd obtain $\log_2 3 \approx 1.58$ bits. But if we are measuring a qubit, then we can obtain at most one bit of information, and in this case even less. The mutual information is bounded by the accessible information or the Holevo bound [1]

$$\chi = S(\hat{\rho}) - \sum_i p_i S(\hat{\rho}_i), \tag{34}$$

where $\hat{\rho} = \sum_i p_i \hat{\rho}_i$ is the average prior density matrix, and $S(\hat{\rho}) = -\mathrm{Tr}\,(\hat{\rho}\log\hat{\rho})$ is the von Neumann entropy. If we measure information in bits, then the basis of the logarithm is 2.

The accessible information bounds the information that can be obtained by any quantum measurement. It is an important quantity e.g. when assessing the security of quantum key distribution, where one needs to be able to bound the information that an eavesdropper could obtain. For pure signal states $\hat{\rho}_i = |\Psi_i\rangle\langle\Psi_i|$, the von Neumann entropies $S(\hat{\rho}_i) = 0$, so that $\chi = S(\hat{\rho})$. This means for example that at most one bit of classical information can be transmitted by a single qubit, since $S(\hat{\rho})$ is at most 1 bit for a qubit. $S(\hat{\rho}) = 1$ for two equiprobable orthogonal pure states, in which case a measurement in that same basis will give the maximum accessible information of 1 bit.

The accessible information is 1 bit also for the three equiprobable symmetric states we have been considering, or for any set of states in two dimensions for which the average prior density matrix is $\hat{\rho} = 1/2\hat{\mathbb{1}}$. But as mentioned above, for the three symmetric states, the maximum mutual information one can obtain is only about 0.67 if only one copy is measured. This illustrates the fact that it may be possible to reach the bound given by χ only in the limit of collective measurements (in an entangled basis) on many transmitted states.

To summarise, we have looked at minimum-error measurements, unambiguous measurements, maximum confidence measurements and measurements that maximise the mutual information. These are not the only possibilities, however. We could also look at estimating some parameter of the measured state, and optimise the measurement e.g. to maximise the information about whatever we are interested in. Generalised measurements can describe any measurement which is possible within quantum mechanics. Next, we will look at how to realise a measurement once we know its mathematical description.

5. Realisation of generalised measurements

5.1. Projective measurement in an extended space.

Any generalised quantum measurement, at least in the finite-dimensional case, can be realised as a projective von Neumann-measurement in an extended higher-dimensional space. This is often referred to as the Neumark or Naimark extension. It is easy to see that the converse should be true. That is, if we have a projective quantum measurement in M dimensions, described by the projection operators P_m, and look at the resulting measurement in any N-dimensional subspace, this will be a generalised measurement in that subspace, with at least N and at most M outcomes.

For example, assume that we couple the system to be measured, initially in the state $|\psi\rangle_S$, to some known ancillary state $|\phi\rangle_{aux}$ using a unitary transform, $\hat{U}|\psi\rangle_S \otimes |\phi\rangle_{aux}$. This is followed by a projective von Neumann measurement in the system-ancilla basis $|i\rangle_S \otimes |j\rangle_{aux}$. Often we are experimentally constrained to make the final measurement in a specific basis, and we can view the unitary transform $\hat{U}$ as allowing us to select any orthonormal complete measurement basis in the system-ancilla space. (We could also easily instead consider the situation where we couple the system to be measured to a probe system, and then measure only the probe. The result would be the same, that is, this procedure results in a generalised measurement on the system to be measured.)

The probability for result (i, j), where i labels the result for the system and j the result for the ancillary system, is then

$$p(i,j) = |_{aux}\langle j| \otimes_S \langle i|\hat{U}|\psi\rangle_S \otimes |\phi\rangle_{aux}|^2 =_S \langle\psi|\hat{\Pi}_{ij}|\psi\rangle_S, \tag{35}$$

where

$$\hat{\Pi}_{ij} =_{aux} \langle\phi|\hat{U}^\dagger||i\rangle_S \otimes |j\rangle_{aux\ aux}\langle j| \otimes_S \langle i|\hat{U}|\phi\rangle_{aux}. \tag{36}$$

These measurement operators describe a generalised quantum measurement, since it can be verified that they are positive and sum to the identity operator in the space of the system to be measured. If the system has d_S dimensions and the ancillary system d_A dimensions, then there are $d_S d_A$ outcomes in total. Therefore, *some* generalised quantum measurements can clearly be realised as projections in a higher-dimensional space.

We will now show that *any* generalised measurement may be realised as a projective measurement in a higher-dimensional Hilbert space. Suppose that a generalised measurement in N dimensions has M measurement operators $\hat{\Pi}_i$, where $i = 1, 2, \ldots M$. Without loss of generality, we can assume that all measurement operators $\hat{\Pi}_i$ are rank one operators. That is, they can be written as

$$\hat{\Pi}_i = |\Psi_i\rangle\langle\Psi_i| \tag{37}$$

for some $|\Psi_i\rangle$, where $|\langle\Psi_i|\Psi_i\rangle| \leq 1$. That is, the states $|\Psi_i\rangle$ might not be normalised. It will hold that $M \geq N$.

We can assume rank-1 measurement operators without loss of generality, because if some of the original measurement operators are not rank one, then we could write them as

$$\Pi_k = \sum_{l=1}^{m} \Pi_{kl}, \tag{38}$$

where $\Pi_{kl} = |\Psi_{kl}\rangle\langle\Psi_{kl}|$ are rank-1 operators, with m chosen sufficiently large (and again, $|\Psi_{k,j}\rangle$ might not be normalised). This way we can construct a new generalised measurement, with the higher-rank (mixed) measurement operators Π_k replaced by sufficiently many rank-1 measurement operators Π_{kl}. The new generalised measurement has only rank-1 measurement operators. The original generalised measurement can clearly be realised as a coarse-graining of the new measurement. Coarse-graining here means that obtaining any of the measurement outcomes kl, corresponding to Π_{kl}, means that the measurement outcome is taken to be k, corresponding to Π_k – we just forget about l.

Suppose therefore that the generalised measurement we are considering has M rank-1 measurement operators $\hat{\Pi}_i$, with $M \geq N$. The M states $|\Psi_i\rangle$ can be written as

$$|\Psi_i\rangle = \sum_{j=1}^{N} a_{ij}|j\rangle, \tag{39}$$

where $\{|j\rangle\}, j = 1, 2, \ldots N$, is any choice of orthonormal basis states in the N-dimensional space of the quantum system to be measured. The condition (7) that $\sum \hat{\Pi}_i = \hat{\mathbb{I}}$, which has to hold since probabilities for all outcomes must sum to one, becomes

$$\sum_{i=1}^{M} |\Psi_i\rangle\langle\Psi_i| = \sum_{i=1}^{M} \sum_{j,k=1}^{N} a_{ij}|j\rangle\langle k|a_{ik}^* = \hat{\mathbb{I}}. \tag{40}$$

This means that

$$\sum_{i=1}^{M} a_{ij}a_{ik}^{*} = \delta_{jk} \tag{41}$$

has to hold, meaning that the N M-dimensional states $|\phi_j\rangle = \sum_i a_{ij}|i\rangle$ are orthonormal. This is the same as saying that we can arrange the coefficients a_{ij} as an $M \times N$ matrix,

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{pmatrix}, \tag{42}$$

where the M rows correspond to the different $|\Psi_i\rangle$ and the N columns correspond to the different $|\phi_j\rangle$. This is not a square matrix unless $M = N$, but since its columns are orthonormal, it can be completed to become a square unitary matrix U_M by adding $M - N$ suitably chosen columns. If $M = N + 1$, then the last column of U_M is unique up to a phase factor, but otherwise there are infinitely many possible ways of completing the matrix.

How does this relate to a realisation of the original generalised measurement? Suppose that we find $M - N$ auxiliary basis states, and then apply the unitary transform U_M in the extended M-dimensional space (the original N basis states, plus the $M - N$ extra basis states). As will be further discussed below, the auxiliary basis states could for example be other energy levels (if we want to measure a trapped ion, say) or more spatial paths (if we are measuring a photon, say), or we could make use of one or more auxiliary qubits of whatever kind is suitable. Then for any state $\hat{\rho}$ that is restricted to the original N-dimensional subspace, the probability for outcome i is given by

$$p(i) = \mathrm{Tr}\left(\hat{\rho}\hat{\Pi}_i\right) = \langle\Psi_i|\hat{\rho}|\Psi_i\rangle = \langle i|U_M\hat{\rho}U_M^{\dagger}|i\rangle. \tag{43}$$

Here, the N first columns of U_M are given by the $M \times N$ matrix in (42), where the ith row contains the coefficients in $|\Psi_i\rangle$. This means that it is possible to realise the generalised measurement by first applying U_M, thereby coupling the quantum system that should be measured to the auxiliary degrees of freedom in a specified way. After that, one makes a projective measurement in the total M-dimensional space. This realises the measurement because it gives the correct probabilities for all the outcomes, for any state that we might measure.

5.2. **The auxiliary basis states.** What the auxiliary basis states are does not matter from a theoretical point of view. They can be chosen as best suits the experimental realisation at hand. From our treatment above, it seems that the auxiliary dimensions are added as a direct sum,

$$\mathcal{H}_M^{tot} = \mathcal{H}_N^S \oplus \mathcal{H}_{M-N}^{aux}, \tag{44}$$

where $\mathcal{H}_M^{tot}$ denotes the total Hilbert space, $\mathcal{H}_N^S$ the space of the original system, and $\mathcal{H}_{M-N}^{aux}$ the auxiliary space. This would naturally be the case for example if the original basis states are energy levels of an atom or an ion, and the extra basis states are other energy levels of that same atom or ion. The unitary transform U_M then acts only on that single quantum system.

We can however also add the auxiliary basis states through a tensor product, by adding a completely new D-dimensional quantum system. The dimension of the resulting total Hilbert space,

$$\mathcal{H}_{DN}^{tot} = \mathcal{H}_N^S \otimes \mathcal{H}_D^{aux}, \tag{45}$$

is then DN, which should be at least M, but may be larger. The unitary transform U_M will now couple the original quantum system to the auxiliary D-dimensional system. The basis states corresponding to the N original columns of U_M may then be chosen for example as $|j\rangle^S \otimes |0\rangle^{aux}$, where $|j\rangle^S$ are the basis states of the system to be measured, and $|0\rangle^{aux}$ is a suitable basis state for the auxiliary quantum system. The remaining $M - N$ states are chosen among $|j\rangle^S \otimes |k\rangle^{aux}$, where $k \neq 0$. If $DN > M$, there are more basis states than needed, and the action of the coupling between original and auxiliary quantum system on the additional basis states does not matter. It may e.g. be chosen as the identity operation, but could be whatever is experimentally convenient.

For optical realisations, the auxiliary basis states are often added as new paths which the photons can take. For example, if the original basis states are the polarisation states $|H\rangle, |V\rangle$ of a single photon, then we can use a beam splitter to couple in new spatial modes, for example increasing the number of spatial modes from one to two. If the two input (and output) modes are labelled by a and b, then the resulting state space now has four basis states, $|H\rangle_a, |V\rangle_a, |H\rangle_b$ and $|V\rangle_b$. By adding more paths, the total number of basis states can be made as high as necessary. The unitary transform U_M will then take the form of an optical network consisting of beam splitters, wave plates and phase shifters (we'll see below how to construct this once we know U_M).

We could also introduce the extra basis states by a tensor product with another quantum system, e.g. another photon or an atom. Photon-photon interactions are however very weak, meaning that this is not so good for measurements on photons from an experimental point of view, since it will be difficult to implement U_M. For the same reason, it is also not so easy to implement measurements where the "original" system contains multi-photon states. For example, a Bell measurement on two photons is a measurement in an entangled basis and therefore difficult to implement – impossible using linear optics – even if it is "just" a projective measurement in a 4-dimensional space. Photon-atom interactions can be stronger, especially in a cavity QED setting. For any sort of interacting qubits in a quantum computer, adding more qubits to introduce the extra dimensions is also clearly an option, and probably the preferred one. For generalised measurement on single photons, a realisation using an optical network will usually be simpler, adding the auxiliary basis states by introducing new paths rather than coupling them to some other kind of qubits.

Next, we will explain how to decompose the unitary transform U_M using simpler basic operations that correspond to 2×2 "beam splitters" and phase shifters. As examples, we will look at unambiguous discrimination of two non-orthogonal states and two polarisation states.

5.3. **Decomposition of unitary transforms.** We've just established that a generalised measurement can be realised as a projective measurement in a higher-dimensional space. The quantum system to be measured should be coupled to the auxiliary basis states using a unitary transform U_M. To give one explicit recipe for how to implement U_M, suitable for experimental realisation, we will describe a way of decomposing any $M \times M$ unitary transform using 2×2 unitaries [10]. There are also other ways of decomposing unitary transforms, e.g. using Householder transforms. The method we will use is particularly suited for realisations using optical networks, where the basic elements, wave plates and beam splitters, implement 2×2 unitary transforms, or for any implementation where we can couple two basis states at a time using e.g. laser pulses or some other control mechanism.

A general 2×2 unitary transform may be written as

$$U_2 = e^{i\phi} \begin{pmatrix} a & -b \\ b^* & a^* \end{pmatrix}, \tag{46}$$

where $|a|^2 + |b|^2 = 1$ and the global phase $e^{i\phi}$ will not matter for our present purpose. We can think of this as a lossless "beam splitter" coupling the two modes, with phase shifts added to input and/or output ports. With T_{pq}, let us denote an $M \times M$ identity matrix with elements I_{pp}, I_{pq}, I_{qp} and I_{qq} replaced by those of U_2 above. This matrix acts as a beam splitter for modes p and q and leaves all other modes unchanged.

Suppose that the matrix elements of the transform U_M are a_{ij}, where $1 \leq i, j \leq M$. By multiplying U_M from the right with a sequence of matrices T_{pq}, we can transform U_M into a diagonal matrix. Starting with the last row of U_M, taking $p = M$ and $q = M-1, M-2, \ldots 1$, we can successively make elements $a_{M,M-1}, a_{M,M-2}, \ldots, a_{M,1}$ zero. Since all transforms are unitary, the last column must then also contain only zeros, except for the diagonal element,

$$U_M \cdot T_{M,M-1} \cdot T_{M,M-2} \cdots T_{M,1} = \left(\begin{array}{cccc|c} & & & & 0 \\ & U_{M-1} & & & 0 \\ & & & & \vdots \\ & & & & 0 \\ \hline 0 & 0 & \cdots & 0 & e^{i\phi_M} \end{array} \right). \tag{47}$$

By applying the same procedure to U_{M-1}, we can make the second last row contain only zeros except for the diagonal element. This process is repeated until U_M has been transformed into a diagonal matrix,

$$U_M \cdot T_{M,M-1} \cdot T_{M,M-2} \cdots T_{2,1} = D = \begin{pmatrix} e^{i\phi_1} & 0 & \cdots & 0 \\ 0 & e^{i\phi_2} & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & e^{i\phi_M} \end{pmatrix}. \tag{48}$$

We can therefore write U_M as a matrix product,

$$U_M = D T_{2,1}^\dagger \cdots T_{M,M-2}^\dagger T_{M,M-1}^\dagger. \tag{49}$$

It follows that the matrix U_M can be realised as a sequence of "beam splitter" operations, followed by final phase shifts on the basis states. When realising generalised quantum measurements, these last phase shifts may be omitted. They will not matter for the final projective measurement.

5.4. **Realisations using sequential measurements and feedback.** Using the construction above, the number of 2×2 beam splitter transforms needed is at most $M(M-1)$. Using sequential projective measurements this number can in fact be further reduced. This can be convenient for experimental realisations. Essentially, one makes use of the considerable freedom in the choice of the transform U_M. The number of auxiliary dimensions that are needed at any one point in an experimental realisation is then also reduced.

In particular, we can realise the measurement in at most M steps using only one extra dimension at a time, that is, only $N+1$ dimensions in the extended space [11]. Essentially, we are checking the measurement outcomes one by one, making a projective measurement to at each step ask "is it this particular measurement outcome or any of the others?". If we identify the outcome, the we can stop, otherwise, we have ruled out that outcome and proceed to pick the next one to test.

A sequential realisation using only an ancillary qubit (two-level system), where the dimension of the space of the system to be measured thus is at most doubled a any one time, requires only a number of steps proportional to $\log M$ [12]. The system to be measured is coupled to a qubit, the qubit measured with a projective measurement with two outcomes, and the process repeated. At each step, we are effectively dividing the

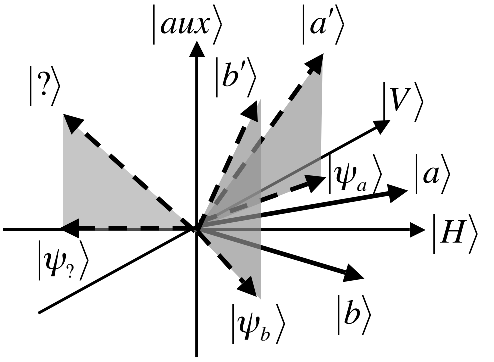

FIGURE 1. Optimal unambiguous identification of two non-orthogonal polarisation states $|a\rangle$ and $|b\rangle$. If we add a third basis state $|aux\rangle$, then we can find a three-dimensional projective measurement in the basis $\{|a'\rangle, |b'\rangle, |?\rangle\}$, such that $|a\rangle$ is orthogonal to $|b'\rangle$ and $|b\rangle$ is orthogonal to $|a'\rangle$. When projecting down these three-dimensional basis states to the original space, spanned by the states $|H\rangle$ and $|V\rangle$, the resulting states are $|\psi_a\rangle, |\psi_b\rangle$ and $|\psi_?\rangle$. The measurement operators are given by $\Pi_a = |\psi_a\rangle\langle\psi_a|, \Pi_b = |\psi_b\rangle\langle\psi_b|, \Pi_? = |\psi_?\rangle\langle\psi_?|$.

outcomes in two groups, and asking in which of the two groups the outcome is. This realisation requires feedback in the sense that subsequent steps in the measurement depends on the result in previous steps. But from an experimental point of view it can be highly advantageous to be able to limit the number of auxiliary degrees of freedom and the number of steps needed to realise a measurement.

5.5. **Examples: polarisation states and coherent states.** As examples we will first look at how to realise optimal unambiguous discrimination between two equiprobable non-orthogonal polarisation states, and then between two coherent states.

5.5.1. *Distinguishing between two polarisation states.* The two non-orthogonal polarisation states $|a\rangle$ and $|b\rangle$ of a photon in section 4.2 were given by

$$|a\rangle = \cos\theta|H\rangle + \sin\theta|V\rangle, |b\rangle = \cos\theta|H\rangle - \sin\theta|V\rangle. \tag{50}$$

Here $0 \leq \theta \leq \pi/4$, and $|H\rangle$ and $|V\rangle$ denote horizontal and vertical polarization. In this case, there will be three measurement outcomes, "the state was definitely $|a\rangle$", "the state was definitely $|b\rangle$" and "don't know". This means that our extended space will only need three dimensions, that is, one auxiliary dimension in addition to the original two. Since there are only three dimensions and the coefficients of the states $|a\rangle$ and $|b\rangle$ are real, we can also visualize what U_M really means instead of just "following the recipe". If we extend the two-dimensional space $\{|H\rangle, |V\rangle\}$ with a third basis state $|aux\rangle$, then, as shown in Fig. 1, we can find a 3D orthonormal basis $\{|a'\rangle, |b'\rangle, |?\rangle\}$, so that $|a\rangle$ is orthogonal to $|b'\rangle$ and $|b\rangle$ is orthogonal to $|a'\rangle$. The choice of these basis states shown in the figure is given by

$$\begin{aligned}
|a'\rangle &= \frac{1}{\sqrt{2}}\left(\tan\theta|H\rangle + |V\rangle + \sqrt{1-\tan^2\theta}|aux\rangle\right), \\
|b'\rangle &= \frac{1}{\sqrt{2}}\left(\tan\theta|H\rangle - |V\rangle + \sqrt{1-\tan^2\theta}|aux\rangle\right), \\
|?\rangle &= -\sqrt{1-\tan^2\theta}|H\rangle + \tan\theta|aux\rangle.
\end{aligned} \tag{51}$$

When projecting down these three-dimensional basis states to the original space spanned by the states $|H\rangle$ and $|V\rangle$, the resulting unnormalised states are $|\psi_a\rangle, |\psi_b\rangle$ and $|\psi_?\rangle$, corresponding to the states $|\Psi_i\rangle$ in equation (39). The measurement operators are given by $\Pi_a = |\psi_a\rangle\langle\psi_a|, \Pi_b = |\psi_b\rangle\langle\psi_b|, \Pi_? = |\psi_?\rangle\langle\psi_?|$, and they are unnormalised projectors; these are the same measurement operators as were given in 4.2. By construction, this measurement has $p(b|a) = \langle a|\Pi_b|a\rangle = 0$ and $p(a|b) = \langle b|\Pi_a|b\rangle = 0$, so the results are error-free (unambiguous).

We can realise the measurement by first implementing

$$\hat{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} \tan\theta & 1 & \sqrt{1-\tan^2\theta} \\ \tan\theta & -1 & \sqrt{1-\tan^2\theta} \\ -\sqrt{2(1-\tan^2\theta)} & 0 & \sqrt{2}\tan\theta \end{bmatrix}, \tag{52}$$

where the initial state only has nonzero components for $|H\rangle$ and $V\rangle$, and then measuring in the $\{|H\rangle, |V\rangle, |aux\rangle\}$ basis. The two first columns of this matrix come from the coefficients for $|H\rangle$ and $|V\rangle$ in (51), which are also the coefficients in $|\psi_a\rangle, |\psi_b\rangle$ and $|\psi_?\rangle$, and the third column is the one we add to complete the matrix to a unitary 3 × 3 transform. The final result H then corresponds to $|a\rangle$, V to b and $|aux\rangle$ corresponds to the inconclusive outcome. Any other assignment of final basis states is of course equally possible and would be reflected in the corresponding $\hat{U}$.

A possible decomposition of $\hat{U}$, using the method described above, is $\hat{U} = \hat{T}_{H,V}\hat{T}_{H,aux}$, where

$$\hat{T}_{H,V} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}, \tag{53}$$

$$\hat{T}_{H,aux} = \begin{bmatrix} \tan\theta & 0 & \sqrt{1-\tan^2\theta} \\ 0 & 1 & 0 \\ -\sqrt{1-\tan^2\theta} & 0 & \tan\theta \end{bmatrix}. \tag{54}$$

A possible realisation of this measurement using an optical network is shown in Fig. 1. This is essentially how the measurement was realised by Clarke et al. on polarised weak coherent states [13]. The polarising beam splitter PBS1 transmits horizontal and reflects vertical polarization, and the auxiliary degree of freedom is introduced at the beam splitter BS, which corresponds to $\hat{T}_{H,aux}$, coupling horizontally polarised light in the lower path marked by "L" to (horizontally) polarized light that would enter in its second input port, indicated by a dotted arrow. Only vacuum is incident on this port, since the states we are distinguishing do not have a component along $|aux\rangle$. The transmission coefficient for the beam splitter BS should be chosen as $T = 1 - \tan^2\theta$ and the reflection coefficient as $R = \tan^2\theta$, since the mode that is labelled aux is reflected at the beam splitter BS.

The second beam splitter transform, $\hat{T}_{H,V}$, is implemented by recombining the states $|H\rangle$ and $|V\rangle$ using PBS2, and then rotating polarisation 45° using the half-wave plate. Result a corresponds to horizontal polarisation and b to vertical polarisation, and after the final polarising beam splitter PBS3, these two possibilities can be distinguished by detectors at the outputs. Essentially the same experimental setup, with the transmission and reflection coefficients of the beam splitter BS properly adjusted, can also be used for minimum-error discrimination between three symmetric states.

Finally, while we have described a realisation using an optical setup, let's note that the same principles can be used also for other realisations. We just have to implement the unitary transform U_M on the relevant quantum systems. Sometimes the decomposition in 2 × 2 transforms we've described is useful, but especially when there are more than

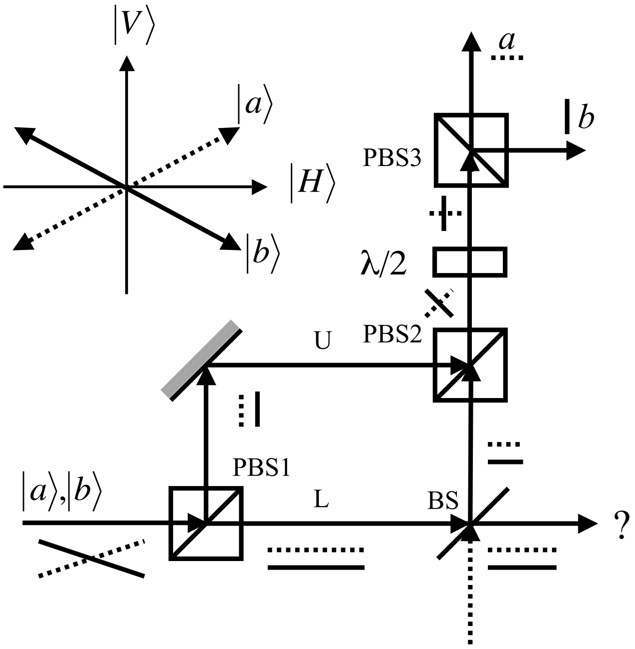

FIGURE 2. Experimental setup for optimal unambiguous identification of two non-orthogonal polarisation states $|a\rangle$ and $|b\rangle$. The polarising beam splitters $PBS1$, $PBS2$, and $PBS3$ transmit horizontal and reflect vertical polarisation. The state space is extended at the beam splitter BS, with vacuum incident on the port where the auxiliary degree of freedom is coupled in, indicated by a dotted arrow. The amplitudes of the states $|a\rangle$ and $|b\rangle$ are schematically indicated by dotted and solid lines at each stage in the network.

one quantum system involved, some other way of decomposing U_M might be more convenient. In quantum computing, we also need to decompose "large" unitary transforms into manageable steps ("basic" quantum gates), and any such decomposition method can of course be used also here.

5.5.2. *Distinguishing between two coherent states.* The optimal unambiguous measurement to unambiguously distinguish between two coherent states $|\alpha\rangle$ and $|-\alpha\rangle$ can also be realised using only linear optics [14], as shown in Fig. 3. That is, we know the amplitude of the coherent state, and that the phase is either $+1$ or -1 relative to some phase reference. Here, we won't directly make use of the constructions above, but the example is nevertheless a nice illustration, and is used in many protocols for quantum communication, including realisations of quantum digital signatures and quantum fingerprinting [15].

The state to be identified, $|\alpha\rangle$ or $|-\alpha\rangle$, is directed onto a balanced beam splitter, with a fixed state $|\alpha\rangle$ incident on the other input port. If the phase relationships between output and input ports are arranged so that the beam splitter transforms $|\alpha\rangle_1 \otimes |\beta\rangle_2$ to $|(\alpha+\beta)/\sqrt{2}\rangle_1 \otimes |(\alpha-\beta)/\sqrt{2}\rangle_2$, we see that if the state to be identified was $|\alpha\rangle$, then output port 1 will contain $|\sqrt{2}\alpha\rangle$ and port 2 will be empty, and if it was $|-\alpha\rangle$, then output port 1 will be empty and output mode 2 contain $|-\sqrt{2}\alpha\rangle$. By detecting photons in the output ports, we can therefore unambiguously tell whether the state in input port 1 was $|\alpha\rangle$ or $|-\alpha\rangle$. Since any coherent state contains a vacuum component, we

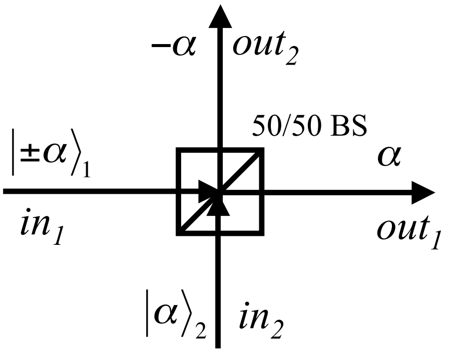

FIGURE 3. Optimal unambiguous identification of a coherent state, known to be either $|\alpha\rangle$ or $|-\alpha\rangle$. The state $|\pm\alpha\rangle$ is interfered with the state $|\alpha\rangle$ on a balanced beam splitter (BS), and photons detected at the output ports. If light is detected in output port 1, the state in input port 1 must have been $|\alpha\rangle$, and if light is detected in output port 2, it must have been $|-\alpha\rangle$. If no photons are detected, this corresponds to the inconclusive outcome.

might not see any photons at all, which corresponds to the inconclusive outcome. The probability for this is $\langle 0|\sqrt{2}\alpha\rangle = \langle -\alpha|\alpha\rangle = \exp(-|\alpha|^2)$, which is the optimal (minimal) failure probability. Clearly, no photon counting is required, only being able to tell the difference between the vacuum and any nonzero number of photons.

For a balanced beam splitter with other phase relationships, we can adjust the phase of the fixed state in input port 2 so that the procedure still works. Also, if the two states to be distinguished are not $|\pm\alpha\rangle$ but $|\alpha\rangle$ and $|\beta\rangle$, then we can precede the described measurement with displacement by $-(\alpha+\beta)/2$ using a beam splitter, and then distinguish $|\pm(\alpha-\beta)/2\rangle$ using the technique above.

In this example, we did not explicitly make use of the construction of a unitary transform to couple the system to be measured to auxiliary degrees of freedom. This is, however, is exactly what is happening. The extra degrees of freedom are introduced when the state $|\pm\alpha\rangle$ is coupled to $|\alpha\rangle$ using the first beam splitter. That the measurement can be realised using only passive linear optics, and with such a simple setup, is rather lucky. Minimum-error discrimination between two coherent states, or optimal unambiguous discrimination between more than two coherent states, are considerably more difficult to realise. The former can only be realised asymptotically using linear optics, and the latter can likely only be realised either suboptimally or asymptotically.

References

[1] S.M. Barnett. *Quantum Information*. Oxford master series in atomic, optical and laser physics. Oxford University Press, 2009.

[2] S.M. Barnett and S. Croke. Quantum state discrimination. *Adv. Opt. Photon*, 1(1):238, 2009.

[3] A. Chefles. Quantum state discrimination. *Contemporary Physics*, 41(6):401–424, 2000.

[4] M. Paris and J. Řeháček. *Quantum State Estimation*. Lecture Notes in Physics. Springer, 2004.

[5] J.A. Bergou. Quantum state discrimination and selected applications. *Journal of Physics: Conference Series*, 84(1):012001, 2007.

[6] C.W. Helstrom. *Quantum Detection and Estimation Theory*. Mathematics in Science and Engineering. Academic Press, 1976.

[7] N. Gisin, G.G. Ribordy, W. Tittel, and H. Zbinden. Quantum cryptography. *Rev. Mod. Phys.*, 74(1):145–195, 2002.

[8] A. Chefles. Unambiguous discrimination between linearly independent quantum states. *Phys. Lett. A*, 239(6):339–347, 1998.

[9] S. Croke, E. Andersson, S. M. Barnett, C. R. Gilson, and J. Jeffers. Maximum confidence quantum measurements. *Phys. Rev. Lett.*, 96:070401, 2006.

[10] M. Reck and A. Zeilinger. Experimental realization of any discrete unitary operator. *Physical Review Letters*, 73:58–61, 1994.

[11] G. Wang and M. Ying. Realization of positive-operator-valued measures by projective measurements without introducing ancillary dimensions. *Arxiv preprint quant-ph/0608235*, 2006.

[12] E. Andersson and D.K.L. Oi. Binary search trees for generalized measurements. *Physical Review A*, 77(5):052104, 2008.

[13] R.B.M. Clarke, A. Chefles, S.M. Barnett, and E. Riis. Experimental demonstration of optimal unambiguous state discrimination. *Physical Review A*, 63(4):040305, 2001.

[14] B. Huttner, N. Imoto, N. Gisin, and T. Mor. Quantum cryptography with coherent states. *Physical Review A*, 51(3):1863, 1995.

[15] J.M. Arrazola and N. Lütkenhaus. Quantum fingerprinting with coherent states and a constant mean number of photons. *Phys. Rev. A*, 89:062305, 2014.