Quantum State Tomography

J. B. Altepeter[1], D. F. V. James[2], and P. G. Kwiat[1]

[1] Dept. of Physics, University of Illinois at Urbana-Champaign, Urbana IL 61801
altepete@uiuc.edu, kwiat@uiuc.edu
[2] Theoretical Division T-4, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 dfvj@lanl.gov

Much like its classical counterpart, which aims at reconstructing three-dimensional images via a series of two-dimensional projections along various 'cuts', quantum tomography characterizes the complete quantum state of a particle or particles through a series of measurements in different bases. While the characterization of a classical object can involve a series of measurements on the same subject, measuring a single quantum particle perturbs its state, often making its further investigation uninformative. For this reason, quantum tomography must be carried out in stages on a number of identical copies of the same state, and can never be successfully applied to a single unknown particle. The etymology of tomography is therefore descriptive (the Greek *tomos* means section) – a series of measurements on identical particle ensembles each allow a glimpse into a distinct aspect of a quantum state's reality. Each new type of measurement illuminates a new dimension of an unknown state; subjecting more identical copies of that state to a single type of measurement brings that particular observable into sharper relief.

This chapter will present a review of the representation of quantum states, a procedure for and an explanation of the reconstruction of an unknown state from a series of ideal measurements on an ensemble of identical particles, and the adaptation of that method to real systems, i.e., non-ideal measurements. Each section will be presented first in general, without respect to a particular physical qubit implementation, followed by the application of that theory to the specific case of qubits encoded into the polarization states of photons. We chose this system for convenience and availability of clean results. However, the techniques presented here can be applied to determine the quantum state of *any* system of one or more qubits (2-level systems). This includes photons [1-10], spin-$\frac{1}{2}$ particles (as, e.g., are used in NMR quantum computing [11, 12, 13, 14]), and (effectively) 2-level atoms [15, 16]. In order to facilitate the use of these techniques by groups and individuals working in any field, a website is available which provides both further details about these techniques and working, documented code for implementing them.[3]

[3] http://www.physics.uiuc.edu/research/QuantumPhotonics/Tomography/

1 State Representation

Before states can be analyzed, it is necessary to understand their representation. In particular, the reconstruction of an unknown state is often simplified by a specific state parametrization.

1.1 Representation of Single Qubit States

Rather than begin with a general treatment of tomography for an arbitrary number of qubits, throughout this chapter the single-qubit case will be investigated initially. This provides the opportunity to strengthen an intuitive grasp of the fundamentals of state representation and tomography before moving on to the more complex (and more useful) general case. In pursuance of this goal, we will use several graphical representations only available at the single-qubit level.

Pure States, Mixed States, and Diagonal Representations

In general, any single-qubit in a pure state can be represented by

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{1}$$

where α and β are complex and $|\alpha|^2 + |\beta|^2 = 1$ [19]. If the normalization is written implicitly and the global phase is ignored, this can be rewritten as

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)e^{i\phi}|1\rangle. \tag{2}$$

These representations are sufficient to enable the description of the action of any operator (e.g., projectors or unitary rotations) on a pure state, and therefore to carry out tomography on that state. However, as previously discussed, any such tomography would require an ensemble of such states. What if the measured ensemble – being as yet unknown – contained an ensemble of *different* pure states? Or what if the members of the ensemble were themselves not in pure states (perhaps because they are entangled with unobserved degrees of freedom)? In this case the overall state is *mixed*.

In general, these mixed states may be described by a probabilistically weighted incoherent sum of pure states. In other words, it is as if any particle in the ensemble has a specific probability of being in a given pure state, and this state is distinguishably labelled in some way. If it were not distinguishable, its constituent pure states would add coherently (with a definite relative phase), yielding a single pure state. A mixed state can be represented by a density matrix $\hat{\rho}$, where

$$\hat{\rho} = \sum_i P_i|\psi_i\rangle\langle\psi_i| = \begin{matrix} \\ |0\rangle \\ |1\rangle \end{matrix} \begin{matrix} \begin{matrix} \langle 0| & \langle 1| \end{matrix} \\ \begin{pmatrix} A & Ce^{i\phi} \\ Ce^{-i\phi} & B \end{pmatrix} \end{matrix}. \tag{3}$$

P_i is the probabalistic weighting ($\sum_i P_i = 1$), A, B and C are all real and non-negative, $A + B = 1$, and $C \leq \sqrt{AB}$ [19].

While any ensemble of pure states can be represented in this way, it is also true that *any* ensemble of single-qubit states can be represented by an ensemble of only *two* orthogonal pure states. (Two pure states $|\psi_i\rangle$ and $|\psi_j\rangle$ are orthogonal if $|\langle\psi_i|\psi_j\rangle| = 0$). For example, if the matrix from equation 3 were diagonal, then it would clearly be a probabalistic combination of two orthogonal states, as

$$\begin{array}{c} \\ |0\rangle \\ |1\rangle \end{array} \begin{array}{c} \begin{array}{cc} \langle 0| & \langle 1| \end{array} \\ \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \end{array} = A|0\rangle\langle 0| + B|1\rangle\langle 1|. \tag{4}$$

However, *any* physical density matrix can be diagonalized, such that

$$\hat{\rho} = \begin{array}{c} \\ |\psi\rangle \\ |\psi^{\perp}\rangle \end{array} \begin{array}{c} \begin{array}{cc} \langle \psi| & \langle \psi^{\perp}| \end{array} \\ \begin{pmatrix} E_1 & 0 \\ 0 & E_2 \end{pmatrix} \end{array} = E_1|\psi\rangle\langle\psi| + E_2|\psi^{\perp}\rangle\langle\psi^{\perp}|, \tag{5}$$

where $\{E_1, E_2\}$ are the eigenvalues of $\hat{\rho}$, and $\{|\psi\rangle, |\psi^{\perp}\rangle\}$ are the eigenvectors (recall that these eigenvectors can always be made mutually orthogonal, denoted here by the $\perp$ symbol). Thus the representation of any quantum state, no matter how it is constructed, is identical to that of an ensemble of two orthogonal pure states. [4]

Examples in the Photon Case

Throughout this chapter, examples will be provided using qubits encoded into the electric field polarization of photons. For a single photon, this system has two levels, e.g., horizontal ($|H\rangle \equiv |0\rangle$) and vertical ($|V\rangle \equiv |1\rangle$), with all possible pure polarization states constructed from coherent superpositions of these two states. For example, diagonal, antidiagonal, right-circular and left-circular light are respectively represented by

[4] It is an interesting question whether all physical states described by a mixed state (e.g., equation 5) are indeed completely equivalent. For example, Leonhardt discussed the notion that two types of unpolarized light could be considered, depending on whether the incoherence between polarization components arose purely due to an averaging over rapidly varying phases, or from an entanglement with another quantum system altogether [17]. This line of thought can even be pushed further, by asking whether all mixed states necessarily arise only from tracing over some unobserved degrees of freedom with which the quantum system has become entangled, or if indeed such entanglement may 'collapse' when the systems involved approach macroscopic size [18]. If the latter were true, then there would exist mixed states that could *not* be seen as pure in some larger Hilbert space. In any event, to our knowledge, at least insofar as state tomography is concerned, these subtleties of interpretation do not in any way affect experimental results.

$$
\begin{aligned}
|D\rangle &\equiv (|H\rangle + |V\rangle)/\sqrt{2}, \\
|A\rangle &\equiv (|H\rangle - |V\rangle)/\sqrt{2}, \\
|R\rangle &\equiv (|H\rangle + i|V\rangle)/\sqrt{2}, \\
\text{and } |L\rangle &\equiv (|H\rangle - i|V\rangle)/\sqrt{2}.
\end{aligned} \tag{6}
$$

Now consider measuring a source of photons which emits a one-photon wave packet each second, but alternates between horizontal, vertical, and diagonal polarizations. Their emission time labels these states (in principle) as distinguishable, and so if we ignore that timing information when they are measured, we must represent their state as a density matrix $\hat{\rho}$:

$$
\begin{aligned}
\hat{\rho} &= \frac{1}{3}(|H\rangle\langle H| + |V\rangle\langle V| + |D\rangle\langle D|) \\
&= \frac{1}{3}\left(\begin{array}{c} \\ |H\rangle \\ |V\rangle \end{array}\!\!\begin{array}{c} \begin{array}{cc} \langle H| & \langle V| \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \end{array} + \begin{array}{c} \\ |H\rangle \\ |V\rangle \end{array}\!\!\begin{array}{c} \begin{array}{cc} \langle H| & \langle V| \end{array} \\ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{array} + \begin{array}{c} \\ |H\rangle \\ |V\rangle \end{array}\!\!\begin{array}{c} \begin{array}{cc} \langle H| & \langle V| \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}\right) \\
&= \frac{1}{6}\left(\begin{array}{c} \\ |H\rangle \\ |V\rangle \end{array}\!\!\begin{array}{c} \begin{array}{cc} \langle H| & \langle V| \end{array} \\ \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \end{array}\right).
\end{aligned} \tag{7}
$$

When diagonalized,

$$
\hat{\rho} = \frac{1}{3}\left(\begin{array}{c} \\ |D\rangle \\ |A\rangle \end{array}\!\!\begin{array}{c} \begin{array}{cc} \langle D| & \langle A| \end{array} \\ \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \end{array}\right) = \frac{2}{3}|D\rangle\langle D| + \frac{1}{3}|A\rangle\langle A|, \tag{8}
$$

which, as predicted in equation 5, is a sum of only *two* orthogonal states.

Henceforth, the 'bra' and 'ket' labels will be suppressed from written density matrices where the basis is $\{|0\rangle, |1\rangle\}$ or $\{|H\rangle, |V\rangle\}$.

The Stokes Parameters and the Poincaré Sphere

Any single-qubit density matrix $\hat{\rho}$ can be uniquely represented by three parameters $\{S_1, S_2, S_3\}$:

$$
\hat{\rho} = \frac{1}{2}\sum_{i=0}^{3} S_i \hat{\sigma}_i. \tag{9}
$$

The $\hat{\sigma}_i$ matrices are

$$
\hat{\sigma}_0 \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \hat{\sigma}_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \hat{\sigma}_2 \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \hat{\sigma}_3 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{10}
$$

and the S_i values are given by

$$S_i \equiv \mathrm{Tr}\left\{\hat{\sigma}_i \hat{\rho}\right\} [19]. \tag{11}$$

For all pure states, $\sum_{i=1}^{3} S_i^2 = 1$; for mixed states, $\sum_{i=1}^{3} S_i^2 < 1$; for the completely mixed state, $\sum_{i=1}^{3} S_i^2 = 0$. Due to normalization, S_0 will always equal one.

Physically, each of these parameters directly corresponds to the outcome of a specific pair of projective measurements:

$$\begin{aligned} S_0 &= P_{|0\rangle} + P_{|1\rangle} \\ S_1 &= P_{\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)} - P_{\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)} \\ S_2 &= P_{\frac{1}{\sqrt{2}}(|0\rangle+i|1\rangle)} - P_{\frac{1}{\sqrt{2}}(|0\rangle-i|1\rangle)} \\ S_3 &= P_{|0\rangle} - P_{|1\rangle}, \end{aligned} \tag{12}$$

where $P_{|\psi\rangle}$ is the probability to measure the state $|\psi\rangle$. As we shall see below, these relationships between probabilities and S parameters are extremely useful in understanding more general operators. Because $P_{|\psi\rangle} + P_{|\psi^\perp\rangle} = 1$, these can be simplified in the single-qubit case, and

$$P_{|\psi\rangle} - P_{|\psi^\perp\rangle} = 2P_{|\psi\rangle} - 1, \tag{13}$$

where $|\psi^\perp\rangle$ denotes the state orthogonal to $|\psi\rangle$.

The probability of projecting a given state $\hat{\rho}$ into the state $|\psi\rangle$ (the probability of measuring $|\psi\rangle$) is given by [20]:

$$\begin{aligned} P_{|\psi\rangle} &= \langle\psi|\hat{\rho}|\psi\rangle \\ &= \mathrm{Tr}\left\{|\psi\rangle\langle\psi|\hat{\rho}\right\}. \end{aligned} \tag{14}$$

In 12 above, the S_i are defined with respect to three states, $|\phi\rangle_i$:

$$\begin{aligned} |\phi\rangle_1 &= \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \\ |\phi\rangle_2 &= \frac{1}{\sqrt{2}}\left(|0\rangle + i|1\rangle\right) \\ |\phi\rangle_3 &= |0\rangle, \end{aligned} \tag{15}$$

and their orthogonal compliments, $|\phi^\perp\rangle$; parameters similar to these and serving the same function can be defined with respect to any three arbitrary states, $|\psi_i\rangle$. (Arbitrary with one condition: the matrices $|\psi_i\rangle\langle\psi_i|$ along with the identity must be linearly independent.) Operators analogous to the $\hat{\sigma}$ operators can be defined relative to these states:

$$\hat{\tau}_i \equiv |\psi_i\rangle\langle\psi_i| - |\psi_i^\perp\rangle\langle\psi_i^\perp|. \tag{16}$$

We can further define an 'S-like' parameter T, given by:

$$T_i = \mathrm{Tr}\left\{\hat{\tau}_i \hat{\rho}\right\}. \tag{17}$$

Continuing the previous convention and to complete the set, we define $\hat{\tau}_0 \equiv \hat{\sigma}_0$, which then requires that $T_0 = 1$. Note that the S_i parameters are simply a special case of the T_i, for the case when $\hat{\tau}_i = \hat{\sigma}_i$.

Unlike the specific case of the S parameters which describe *orthogonal* measurement bases, for non-orthogonal measurements

$$\hat{\rho} \neq \frac{1}{2}\sum_{i=0}^{3} T_i \hat{\tau}_i. \tag{18}$$

In order to reconstruct the density matrix, the T parameters must first be transformed into the S parameters (see equation 21).

The Photon Case

For photon polarization, the S_i are the famous Stokes parameters (though normalized), and correspond to measurements of D/A, R/L, and H/V [21]. In terms of the $\hat{\tau}$ matrices just introduced, we would define a set of basis states $|\psi_1\rangle \equiv |D\rangle, |\psi_2\rangle \equiv |R\rangle$, and $|\psi_3\rangle \equiv |H\rangle$. For these analysis bases, $\hat{\tau}_1 = \hat{\sigma}_1, \hat{\tau}_2 = \hat{\sigma}_2$, and $\hat{\tau}_3 = \hat{\sigma}_3$ (and therefore $T_i = S_i$ for this specific choice of analysis bases).

As the simplest example, consider the input state $|H\rangle$. Applying equation (11), we find that

$$\begin{aligned} S_0 &= \mathrm{Tr}\left\{\sigma_0 \rho_H\right\} = 1 \\ S_1 &= \mathrm{Tr}\left\{\sigma_1 \rho_H\right\} = 0 \\ S_2 &= \mathrm{Tr}\left\{\sigma_2 \rho_H\right\} = 0 \\ S_3 &= \mathrm{Tr}\left\{\sigma_3 \rho_H\right\} = 1, \end{aligned} \tag{19}$$

which from equation 9 implies that

$$\rho_H = (\sigma_0 + \sigma_3) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{20}$$

When the Stokes parameters are used as coordinates in 3-space, the space of all legal states falls within a sphere of radius one (the Poincaré sphere for polarization, the Bloch sphere for electron spin or other two-level systems). The pure states are found on the surface, states of linear polarization on the equator, circular states at the poles, mixed states within, and the totally mixed state – corresponding to completely unpolarized photons – at the center of the sphere. This provides a very convenient way to visualize one-qubit states (see figure 1). The θ and ϕ values from equation 2 allow any pure state

to be easily mapped onto the sphere surface. These values are the polar coordinates of the pure state they represent on the Poincaré sphere. [5] In addition to mapping states, the sphere can be used to represent any unitary operation as a rotation about an arbitrary axis. For example, waveplates implement rotations about an axis that passes through the equator.

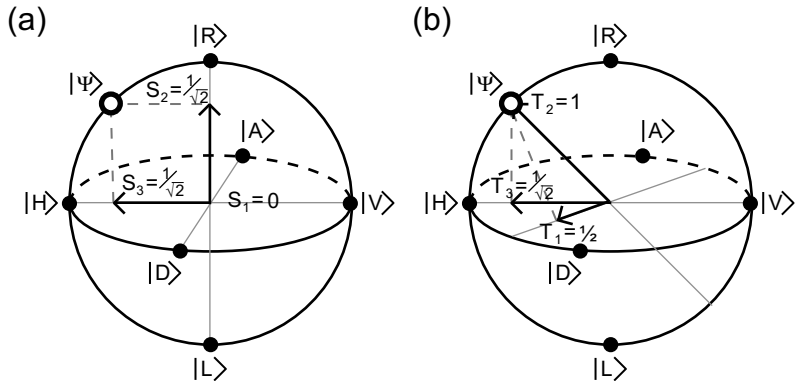


Fig. 1. The Bloch Poincaré sphere. Any single-qubit quantum state $\hat{\rho}$ can be represented by three parameters $T_i = \mathrm{Tr}\{\hat{\tau}_i\hat{\rho}\}$, as long as the operators $\hat{\tau}_i$ in addition to the identity are linearly independent. Physically, the T_i parameters directly correspond to the outcome of a specific projective measurement: $T_i = 2P_i - 1$, where P_i is the probability of success for the measurement. The T_i may be used as coordinates in 3-space. Then all 1-qubit quantum states fall on or within a sphere of radius one. The surface of the sphere corresponds to pure states, the interior to mixed states, and the origin to the totally mixed state. Shown is a particular pure state $|\psi\rangle$, which is completely specified by its projection onto a set of non-parallel basis vectors. (a) When $\hat{\tau}_i = \hat{\sigma}_i$ (the Pauli matrices), the basis vectors are orthogonal, and in this particular case the T_i are equal to the S_i, the well known Stokes parameters, corresponding to measurements of diagonal (S_1), right-circular (S_2), and horizontal (S_3) polarizations. (b) A non-orthogonal coordinate system in Poincaré space. It is possible to represent a state using its projection onto non-orthogonal axes in Poincaré space. This is of particular use when attempting to reconstruct a quantum state from non-orthogonal measurements. Shown here are the axes corresponding to measurements of 22.5° linear (T_1), elliptical light rotated 22.5° from H towards R (T_2), and horizontal (T_3).

Any state $|\psi\rangle$ and its orthogonal partner, $|\psi^{\perp}\rangle$, are found on opposite points of the Poincaré sphere. The line connecting these two points forms an

[5] These polar coordinates are by convention rotated by 90°, so that $\theta = 0$ is on the equator corresponding to the state $|H\rangle$ and $\theta = 90°, \phi = 90°$ is at the North Pole corresponding to the state $|R\rangle$. This 90° rotation is particular to the Poincaré representation of photon polarization [22]; representations of two-level systems on the Bloch sphere do not introduce it.

axis of the sphere, useful for visualizing the outcome of a measurement in the $|\psi\rangle/|\psi^{\perp}\rangle$ basis. The projection of any state $\hat{\rho}$ (through a line perpendicular to the $|\psi\rangle/|\psi^{\perp}\rangle$ axis), will lie a distance along this axis corresponding to the relevant Stokes-like parameter ($T = \langle\psi|\hat{\rho}|\psi\rangle - \langle\psi^{\perp}|\hat{\rho}|\psi^{\perp}\rangle$).

Thus, just as any point in three-dimensional space can be specified by its projection onto three linearly independent axes, any quantum state can be specified by the three parameters $T_i = \mathrm{Tr}\{\hat{\tau}_i\hat{\rho}\}$, where $\hat{\tau}_{i=1,2,3}$ are linearly independent matrices equal to $|\psi_i\rangle\langle\psi_i| - |\psi_i^{\perp}\rangle\langle\psi_i^{\perp}|$. The $\hat{\tau}_i$ correspond to general Stokes-like parameters for any three linearly independent axes on the Poincaré sphere. However, they can differ from the canonical Stokes axes and need not even be orthogonal. See figure 1b for an example of state representation using non-orthogonal axes.

In order to use these non-orthogonal Stokes-like parameters, it is necessary to be able to transform a state from a Stokes representation to a new nonorthogonal representation and vice-versa. In general, for any two representations $S_i = \mathrm{Tr}\{\hat{\sigma}_i\hat{\rho}\}$ and $T_i = \mathrm{Tr}\{\hat{\tau}_i\hat{\rho}\}$ it is possible to transform between them by using

$$\begin{pmatrix} T_0 \\ T_1 \\ T_2 \\ T_3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \mathrm{Tr}\{\hat{\tau}_0\hat{\sigma}_0\} & \mathrm{Tr}\{\hat{\tau}_0\hat{\sigma}_1\} & \mathrm{Tr}\{\hat{\tau}_0\hat{\sigma}_2\} & \mathrm{Tr}\{\hat{\tau}_0\hat{\sigma}_3\} \\ \mathrm{Tr}\{\hat{\tau}_1\hat{\sigma}_0\} & \mathrm{Tr}\{\hat{\tau}_1\hat{\sigma}_1\} & \mathrm{Tr}\{\hat{\tau}_1\hat{\sigma}_2\} & \mathrm{Tr}\{\hat{\tau}_1\hat{\sigma}_3\} \\ \mathrm{Tr}\{\hat{\tau}_2\hat{\sigma}_0\} & \mathrm{Tr}\{\hat{\tau}_2\hat{\sigma}_1\} & \mathrm{Tr}\{\hat{\tau}_2\hat{\sigma}_2\} & \mathrm{Tr}\{\hat{\tau}_2\hat{\sigma}_3\} \\ \mathrm{Tr}\{\hat{\tau}_3\hat{\sigma}_0\} & \mathrm{Tr}\{\hat{\tau}_3\hat{\sigma}_1\} & \mathrm{Tr}\{\hat{\tau}_3\hat{\sigma}_2\} & \mathrm{Tr}\{\hat{\tau}_3\hat{\sigma}_3\} \end{pmatrix} \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix}. \quad (21)$$

This relation allows S parameters to be transformed into any set of T parameters. In order to transform from T to S, we can invert the 4 by 4 matrix in equation 21 and multiply both sides by this new matrix. This inversion is possible because we have chosen the $\hat{\tau}_i$ operators to be linearly independent, as otherwise the T_i parameters would not specify a single point in Hilbert space.

1.2 Representation of Multiple Qubits

With the extension of these ideas to cover multiple qubits, it becomes possible to investigate non-classical features, including the quintessentially quantum mechanical phenomenon of entanglement.

Pure States, Mixed States, and Diagonal Representations

As the name implies, multiple-qubit states are constructed out of individual qubits. As such, the Hilbert space of a many qubit system is spanned by state vectors which are the tensor product of single-qubit state vectors. A general n-qubit system can be written as

$$|\psi\rangle = \sum_{i_1,i_2,\ldots i_n=0,1} \alpha_{i_1,i_2,\ldots i_n} |i_1\rangle \otimes |i_2\rangle \otimes \ldots \otimes |i_n\rangle. \quad (22)$$

Here the α_i are complex, $\sum_i |\alpha_i|^2 = 1$, and $\otimes$ denotes a tensor product, used to join component Hilbert spaces. For example, a general two-qubit pure state can be written

$$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle, \tag{23}$$

where $|00\rangle$ is shorthand for $|0\rangle_1 \otimes |0\rangle_2$.

As before, we represent a general mixed state through an incoherent sum of pure states:

$$\hat{\rho} = \sum_i P_i |\psi_i\rangle\langle\psi_i|. \tag{24}$$

And, as before, each n-qubit state can be represented by a 2^n by 2^n density matrix which, when diagonalized, allows any state to be written as

$$\hat{\rho} = \sum_{i=1}^{2^n} P_i |\phi_i\rangle\langle\phi_i|. \tag{25}$$

(24) differs from (25) in that the ϕ_i are necessarily orthogonal ($\langle\psi_i|\psi_j\rangle = \delta_{ij}$), and there are at most 2^n of them (in (24) there could be an arbitrary number of $|\psi_i\rangle$).

The Photon Case

Any two-qubit polarization state can be written as

$$\hat{\rho} = \begin{array}{c} \\ |HH\rangle \\ |HV\rangle \\ |VH\rangle \\ |VV\rangle \end{array} \begin{array}{c} \begin{array}{cccc} \langle HH| & \langle HV| & \langle VH| & \langle VV| \end{array} \\ \begin{pmatrix} A_1 & B_1 e^{i\phi_1} & B_2 e^{i\phi_2} & B_3 e^{i\phi_3} \\ B_1 e^{-i\phi_1} & A_2 & B_4 e^{i\phi_4} & B_5 e^{i\phi_5} \\ B_2 e^{-i\phi_2} & B_4 e^{-i\phi_4} & A_3 & B_6 e^{i\phi_6} \\ B_3 e^{-i\phi_3} & B_5 e^{-i\phi_5} & B_6 e^{-i\phi_6} & A_4 \end{pmatrix} \end{array}, \tag{26}$$

where $\hat{\rho}$ is positive and Hermitian with unit trace. Henceforth, the 'bra' and 'ket' labels will be omitted from density matrices presented in this standard basis.

Perhaps the most famous examples of pure two-qubit states are the Bell states [23]:

$$\begin{aligned} |\phi^{\pm}\rangle &= \frac{1}{\sqrt{2}} \left(|HH\rangle \pm |VV\rangle\right) \\ |\psi^{\pm}\rangle &= \frac{1}{\sqrt{2}} \left(|HV\rangle \pm |VH\rangle\right). \end{aligned} \tag{27}$$

Mixed states of note include the Werner states [24],

$$\hat{\rho}_W = P|\gamma\rangle\langle\gamma| + (1-P)\frac{1}{4}I, \tag{28}$$

where $|\gamma\rangle$ is a maximally entangled state and $\frac{1}{4}I$ is the totally mixed state, and the maximally entangled mixed states (MEMS), which possess the maximal amount of entanglement for a given amount of mixture [25].

The entanglement and the mixture are two of the many quantities derived from the density matrix used to characterize a quantum state, several of which will be included here for reference:

Fidelity Fidelity is a measure of state overlap. $F(\rho_1, \rho_2) = \left(\mathrm{Tr}\left\{\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\right\}\right)^2$, which - for ρ_1 or ρ_2 pure - simplifies to $\mathrm{Tr}\{\rho_1\rho_2\}$ [19].

Tangle The concurrence and tangle are measures of the quantum-coherence properties of a quantum state [34]. For two qubits[6], concurrence is defined as follows: consider the non-Hermitian matrix $\hat{R} = \hat{\rho}\hat{\Sigma}\hat{\rho}^{\mathrm{T}}\hat{\Sigma}$ where the superscript T denotes transpose and the 'spin flip matrix' $\hat{\Sigma}$ is defined by:

$$\hat{\Sigma} \equiv \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}. \tag{29}$$

If the eigenvalues of $\hat{R}$, arranged in decreasing order, are given by $r_1 \geq r_2 \geq r_3 \geq r_4$, then the concurrence is defined by

$$C = \mathrm{Max}\left\{0, \sqrt{r_1} - \sqrt{r_2} - \sqrt{r_3} - \sqrt{r_4}\right\}. \tag{30}$$

The tangle is calculated directly from the concurrence:

$$T = C^2. \tag{31}$$

The tangle (and the concurrence) range from 0 for product states (or, more generally, any incoherent mixture of product states) to a maximum value of 1 for Bell states.

Entropy and the Linear Entropy The Von Neuman entropy quantifies the degree of mixture in a quantum state, and is given by

$$S \equiv -\mathrm{Tr}\left\{\hat{\rho}\ln\left[\hat{\rho}\right]\right\} = -\sum_i p_i \ln\left\{p_i\right\}, \tag{32}$$

where the p_i are the eigenvalues of ρ. The linear entropy [1] is a more analytically convenient form of the same quantity. The linear entropy for a two-qubit system is defined by:

[6] The analysis in this subsection applies to the two qubit case only. Measures of entanglement for mixed n-qubit systems are a subject of on-going research: see, for example, [35] for a recent survey. It may be possible to measure entanglement directly, without quantum state tomography; this possibility was investigated in [36].

$$S_L = \frac{4}{3}\left(1 - \mathrm{Tr}\left\{\hat{\rho}^2\right\}\right)$$
$$= \frac{4}{3}\left(1 - \sum_{a=1}^{4} p_a^2\right), \tag{33}$$

where p_a are the eigenvalues of ρ. S_L ranges from 0 for pure states to 1 for the completely mixed state.

Multiple Qubit Stokes Parameters

Extending the single-qubit density matrix representation (equation 9), any n-qubit state $\hat{\rho}$ may be represented as

$$\hat{\rho} = \frac{1}{2^n} \sum_{i_1, i_2, \ldots i_n = 0}^{3} S_{i_1, i_2, \ldots i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \ldots \otimes \hat{\sigma}_{i_n}. \tag{34}$$

Normalization requires that $S_{0,0,\ldots 0} = 1$, allowing $4^n - 1$ real parameters (the multiple-qubit analog of the single-qubit Stokes parameters) to identify any point in Hilbert space, just as three parameters determined the exact position of a one-qubit state in Bloch/Poincaré space. Already for two qubits, the state space is much larger, requiring 15 independent real parameters to describe it. For this reason, there is no convenient graphical picture of this space, as there was in the single-qubit case (see, however, the interesting approaches made by Zyczkowski [31]).

For multiple qubits the link between the multiple-qubit Stokes parameters [28, 30] and measurement probabilities still exists. The formalism of $\hat{\tau}$ operators also still holds for larger qubit systems, so that

$$T = \mathrm{Tr}\left\{\hat{\tau}\hat{\rho}\right\}. \tag{35}$$

For 'local' measurements (a local measurement is the tensor product of a number of single-qubit measurements: the first projecting qubit one along $\hat{\tau}_{i_1}$, the second qubit two along $\hat{\tau}_{i_2}$, , etc.), $\hat{\tau} = \hat{\tau}_{i_1} \otimes \hat{\tau}_{i_2} \otimes \ldots \otimes \hat{\tau}_{i_n}$. Combining equations 34 and 35,

$$T_{i_1, i_2, \ldots i_n} = \mathrm{Tr}\left\{\left(\hat{\tau}_{i_1} \otimes \hat{\tau}_{i_2} \otimes \ldots \otimes \hat{\tau}_{i_n}\right)\hat{\rho}\right\} \tag{36}$$
$$= \frac{1}{2^n} \sum_{j_1, j_2, \ldots j_n = 0}^{3} \mathrm{Tr}\left\{\hat{\tau}_{i_1}\hat{\sigma}_{j_1}\right\} \mathrm{Tr}\left\{\hat{\tau}_{i_2}\hat{\sigma}_{j_2}\right\} \ldots \mathrm{Tr}\left\{\hat{\tau}_{i_n}\hat{\sigma}_{j_n}\right\} S_{j_1, j_2, \ldots j_n}.$$

Recall that for single qubits,

$$T_{i=1,2,3} = P_{|\psi_i\rangle} - P_{|\psi_i^{\perp}\rangle}$$
$$T_0 = P_{|\psi\rangle} + P_{|\psi^{\perp}\rangle}, \forall\psi \tag{37}$$

Therefore, for an n-qubit system,

$$T_{i_1,i_2,\ldots i_n} = \\ (P_{|\psi_{i_1}\rangle} \pm P_{|\psi_{i_1}^{\perp}\rangle}) \otimes (P_{|\psi_{i_2}\rangle} \pm P_{|\psi_{i_2}^{\perp}\rangle}) \otimes \ldots \otimes (P_{|\psi_{i_n}\rangle} \pm P_{|\psi_{i_n}^{\perp}\rangle}), \quad (38)$$

where the plus sign is used for a 0 index and the minus sign is used for a nonzero index. For a two-qubit system where $i_1 \neq 0$ and $i_2 \neq 0$, T_{i_1,i_2} simplifies dramatically, giving

$$\begin{aligned} T_{i_1,i_2} &= (P_{|\psi_{i_1}\rangle} - P_{|\psi_{i_1}^{\perp}\rangle}) \otimes (P_{|\psi_{i_2}\rangle} - P_{|\psi_{i_2}^{\perp}\rangle}) \\ &= P_{|\psi_{i_1}\rangle|\psi_{i_2}\rangle} - P_{|\psi_{i_1}\rangle|\psi_{i_2}^{\perp}\rangle} - P_{|\psi_{i_1}^{\perp}\rangle|\psi_{i_2}\rangle} + P_{|\psi_{i_1}^{\perp}\rangle|\psi_{i_2}^{\perp}\rangle}. \end{aligned} \quad (39)$$

This relation will be crucial for rebuilding a two-qubit state from local measurements.

As before, we are not restricted to multiple-qubit Stokes parameters based on orthogonal operators. Extending equation 21 to multiple qubits, and again assuming two representations $S_{i_1,i_2,\ldots i_n} = \mathrm{Tr}\left\{(\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \ldots \otimes \hat{\sigma}_{i_n})\,\hat{\rho}\right\}$, and $T_{i_1,i_2,\ldots i_n} = \mathrm{Tr}\left\{(\hat{\tau}_{i_1} \otimes \hat{\tau}_{i_2} \otimes \ldots \otimes \hat{\tau}_{i_n})\,\hat{\rho}\right\}$,

$$T_{i_1,i_2,\ldots i_n} = \quad (40)$$
$$\frac{1}{2^n} \sum_{j_1,j_2,\ldots j_n=0}^{3} \mathrm{Tr}\left\{(\hat{\tau}_{i_1} \otimes \hat{\tau}_{i_2} \otimes \ldots \otimes \hat{\tau}_{i_n})\,(\hat{\sigma}_{j_1} \otimes \hat{\sigma}_{j_2} \otimes \ldots \otimes \hat{\sigma}_{j_n})\right\} S_{j_1,j_2,\ldots j_n}.$$

Example: Two-Qubit Polarization States

Consider the state $|HH\rangle$. Following the example in equation 20,

$$\begin{aligned} \hat{\rho}_{HH} &= |HH\rangle\langle HH| \\ &= \frac{1}{2}(\hat{\sigma}_0 + \hat{\sigma}_3) \otimes \frac{1}{2}(\hat{\sigma}_0 + \hat{\sigma}_3) \\ &= \frac{1}{4}(\hat{\sigma}_0 \otimes \hat{\sigma}_0 + \hat{\sigma}_3 \otimes \hat{\sigma}_0 + \hat{\sigma}_0 \otimes \hat{\sigma}_3 + \hat{\sigma}_3 \otimes \hat{\sigma}_3). \end{aligned} \quad (41)$$

This implies that there are exactly four non-zero two-qubit Stokes parameters: $S_{0,0}, S_{0,3}, S_{3,0}$, and $S_{3,3}$ – all of which are equal to one. (As earlier, for the special case when $\hat{\tau}_{i,j} = \hat{\sigma}_{i,j}$, we relabel the $T_{i,j}$ as $S_{i,j}$, the two-qubit Stokes parameters [28, 30].) The separable nature of this state makes it easy to calculate the two-qubit Stokes decomposition.

If instead we investigate an entangled state, $|\psi^-\rangle$, it will be necessary to calculate each two-qubit Stokes parameter from the $\hat{\sigma}$ matrices. As an example, consider $\hat{\sigma}_{3,3} \equiv \hat{\sigma}_3 \otimes \hat{\sigma}_3$, for which

$$S_{3,3} = \mathrm{Tr}\left\{\hat{\sigma}_{3,3}|\psi^-\rangle\langle\psi^-|\right\} = -1. \quad (42)$$

We could instead calculate $S_{3,3}$ directly from probability outcomes of measurements on $|\psi^-\rangle$:

$$\begin{aligned} S_{3,3} &= (P_H - P_V) \otimes (P_H - P_V) \\ &= P_{HH} - P_{HV} - P_{VH} + P_{VV} \\ &= 0 - \frac{1}{2} - \frac{1}{2} + 0 \\ &= -1. \end{aligned} \tag{43}$$

In general, a given $\hat{\tau}$ operator is not uniquely mapped to a single pair of analysis states. For example, if the analysis statse are $|\psi_1\rangle \equiv |H\rangle$ and $|\psi_2\rangle \equiv |V\rangle$, $\hat{\tau}_{1,1} \equiv \hat{\sigma}_3 \otimes \hat{\sigma}_3 = -\hat{\sigma}_3 \otimes -\hat{\sigma}_3 \equiv \hat{\tau}_{2,2}$. However, this would be an unacceptable choice for $|\psi_1\rangle$ and $|\psi_2\rangle$, as their respective density matrices are not linearly independent.

Continuing on, we measure $S_{0,3}$:

$$\begin{aligned} S_{0,3} &= (P_H + P_V) \otimes (P_H - P_V) \\ &= P_{HH} - P_{HV} + P_{VH} - P_{VV} \\ &= 0 - \frac{1}{2} + \frac{1}{2} - 0 \\ &= 0. \end{aligned} \tag{44}$$

Here the signs of the probabilities changed due to the zero index in $S_{0,3}$. These results would have been the same even if the analysis bases of the first qubit had been shifted to any other orthogonal basis, i.e., $S_{0,3} = \left(P_\psi + P_{\psi^\perp}\right) \otimes (P_H - P_V)$.

If the method above is continued for all the Stokes parameters, one concludes that

$$\begin{aligned} \hat{\rho}_{\psi^-} &= \frac{1}{2}(|HV\rangle - |VH\rangle)(\langle HV| - \langle VH|) \\ &= \frac{1}{4}(\hat{\sigma}_0 \otimes \hat{\sigma}_0 - \hat{\sigma}_1 \otimes \hat{\sigma}_1 - \hat{\sigma}_2 \otimes \hat{\sigma}_2 - \hat{\sigma}_3 \otimes \hat{\sigma}_3). \end{aligned} \tag{45}$$

2 Exact Tomography

The goal of tomography is to reconstruct the density matrix of an unknown ensemble of particles through a series of measurements. In practice, this can never be performed exactly, as an infinite number of particles would be required to eliminate statistical error. If exact measurements were taken on infinite ensembles, each measurement would yield an exact probability of success, which could then be used to reconstruct a density matrix. Though unrealistic, it is highly illustrative to examine this exact tomography before seeing the more general treatment. Hence, this section will treat all measurements as yielding exact probabilities, and ignore all sources of error in those measurements.

2.1 Single Qubit Tomography

Although reconstructive tomography of any size system follows the same general procedure, beginning with tomography of a single qubit allows the visualization of each step using the Poincaré sphere, in addition to providing a simpler mathematical introduction.

Visualization of Single Qubit Tomography

Exact single-qubit tomography requires a sequence of three linearly independent measurements. Each measurement exactly specifies one degree of freedom for the measured state, reducing the free parameters of the unknown state's possible Hilbert space by one.

As an example, consider measuring R, D, and H on the mixed state

$$\hat{\rho} = \begin{pmatrix} \frac{5}{8} & \frac{-i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{3}{8} \end{pmatrix} \tag{46}$$

Rewriting the state (using equation 9) as

$$\hat{\rho} = \frac{1}{2}\left(\hat{\sigma}_0 + \frac{1}{\sqrt{2}}\hat{\sigma}_2 + \frac{1}{4}\hat{\sigma}_3\right) \tag{47}$$

allows us to read off the normalized Stokes parameters corresponding to these measurements:

$$S_1 = 0, S_2 = \frac{1}{\sqrt{2}}, \text{ and } S_3 = \frac{1}{4}. \tag{48}$$

As always, $S_0 = 1$ due to normalization. Measuring R first, and looking to the Poincaré sphere, we see that the unknown state must lie in the $z = \frac{1}{\sqrt{2}}$ plane (as $S_2 = \frac{1}{\sqrt{2}}$). A measurement in the D basis further constrains the state to the $y = 0$ plane, resulting in a total confinement to a line parallel to and directly above the x axis. The final measurement of H pinpoints the state. This process is illustrated in figure 2a. Obviously the order of the measurements is irrelevant: it is the intersection point of three orthogonal planes that defines the location of the state.

If instead measurements are made along *non*-orthogonal axes, a very similar picture develops, as indicated in figure 2b. The first measurement always isolates the unknown state to a plane, the second to a line, and the third to a point.

Of course, in practice, the experimenter has no knowledge of the unknown state before a tomography. The set of the measured probabilities, transformed into the Stokes parameters as above, allow a state to be directly reconstructed.

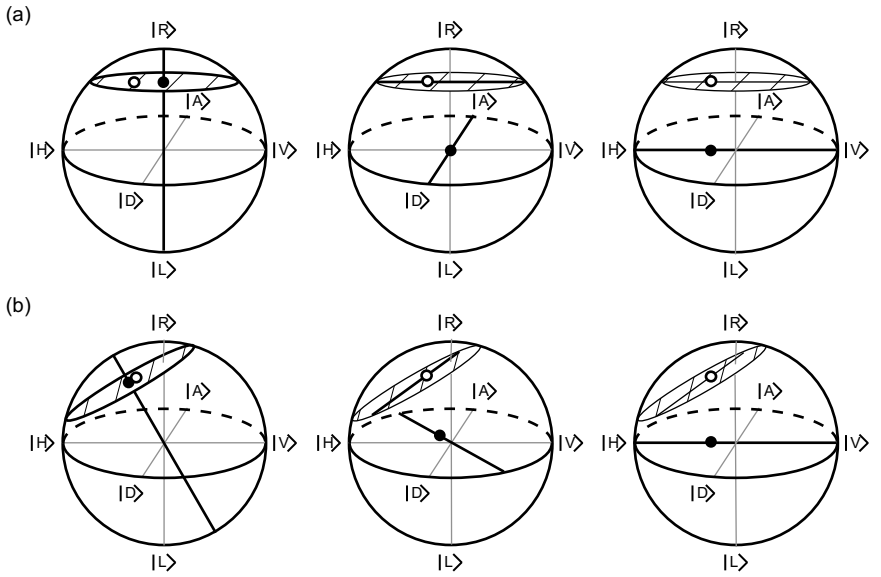


Fig. 2. A sequence of three linearly independent measurements isolates a single quantum state in Hilbert space (shown here as an open circle in the Poincaré sphere representation). The first measurement isolates the unknown state to a plane perpendicular to the measurement basis. Further measurements isolate the state to the intersections of non-parallel planes, which for the second and third measurements correspond to a line and finally a point. The black dots shown correspond to the projection of the unknown state onto the measurement axes, which determines the position of the aforementioned planes. (a) A sequence of measurements along the right-circular, diagonal, and horizontal axes. (b) A sequence of measurements on the same state taken using non-orthogonal projections: elliptical light rotated 30° from H towards R, 22.5° linear, and horizontal.

A Mathematical Look at Single Qubit Tomography

Using the tools developed in the first section of this chapter, single-qubit tomography is relatively straightforward. Recall equation 9, $\hat{\rho} = \frac{1}{2}\sum_{i=0}^{3} S_i \hat{\sigma}_i$. Considering that S_1, S_2, and S_3 completely determine the state, we need only measure them to complete the tomography. As $S_{j>0} = 2P_{|\psi\rangle} - 1$ (equation 13), three measurements respectively in the $|0\rangle$, $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and $\frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$ bases will completely specify the unknown state. If instead measurements are made in another basis, even a non-orthogonal one, they can be easily related back to the S_i parameters, and therefore the density matrix, by means of equation 21.

While this procedure is straightforward, there is one subtlety which will become important in the multiple-qubit case. Projective measurements generally refer to the measurement of a single basis state and return a single

value between zero and one. This corresponds, for example, to an electron beam passing through a Stern-Gerlach apparatus with a detector placed at one output. While a single detector and knowledge of the input particle intensity will – in the one-qubit case – completely determine a single Stokes parameter, one could collect data from both outputs of the Stern-Gerlach device. This would measure the probability of projecting not only onto the state $|\psi\rangle$, but also onto $|\psi^{\perp}\rangle$, and without needing to know the input intensity. All physical measurements on single qubits, regardless of implementation, can in principle be measured this way (though in practice measurements of some qubit systems may typically detect a population in only *one* of the states [32]). We will see below that although one detector functions as well as two in the single-qubit case, this situation will not persist into higher dimensions.

Arbitrary Measurements Using Waveplates and Polarizers

An arbitary polarization measurement and its orthogonal compliment can be realized using, in order, a quarter-wave plate, a half-waveplate, and a polarizing beam splitter. Waveplates implement unitary operations, and in the Poincaré sphere picture, act as rotations about an axis lying within the linear polarization plane (the equator) [26]. Specifically, a waveplate whose optic axis is oriented at angle θ with respect to the horizontal induces a rotation on the Poincaré sphere about an axis 2θ from horizontal, in the linear plane. The magnitude of this rotation is equal to the waveplate's retardance (90° for quarter-wave plates and 180° for half-wave plates). For the remainder of this chapter we adopt the convention that polarizing beam splitters transmit horizontally polarized light and reflect vertically polarized light.

This analysis, while framed in terms of waveplates acting on photon polarization, is directly applicable to other systems, e.g., spin-$\frac{1}{2}$ particles [11, 12, 13, 14] or two-level atoms [15, 16]. In these systems, measurements in arbitrary bases are obtained using suitably phased π- and $\frac{\pi}{2}$-pulses (externally applied electromagnetic fields) to rotate the state to be measured into the desired analysis basis.

To derive the settings for these waveplates as a function of the projection state desired, we use the Poincaré sphere (see figure 3). For any state on the surface of the sphere, a 90° rotation about a linear axis directly below it will rotate that state into a linear polarization (see figure 3b). Assume the desired projection state is

$$|\psi_P\rangle = \cos\left(\frac{\theta}{2}\right)|H\rangle + \sin\left(\frac{\theta}{2}\right)e^{i\phi}|V\rangle. \tag{49}$$

Simple coordinate transforms from spherical to cartesian coordinates reveal that a quarter-waveplate at $\theta_{QWP} = \frac{1}{2}\mathrm{acos}\left\{\sin(\theta)\tan(\phi)\right\}$ will rotate the projection state (49) into a linear state

$$|\psi'_P\rangle = \cos\left(\frac{\theta'}{2}\right)|H\rangle + \sin\left(\frac{\theta'}{2}\right)|V\rangle. \tag{50}$$

A half-waveplate at $\frac{1}{4}\theta'$ will rotate this state to $|H\rangle$.[7] The PBS will then transmit the projected state and reflect its orthogonal compliment.

Of course, these calculations assume that waveplates with retardances equal to exactly π or $\frac{\pi}{2}$ are used (or Rabi pulses producing perfect phase differences). Imperfect yet well characterized waveplates will lead to measurements in slightly different yet known bases. This will still yield an accurate tomography, but first these results must be transformed from a non-orthogonal basis into the canonical Stokes parameters using equation 21. Later, the maximum likelihood technique (see section 3) will provide a different but equally effective way to accomodate for imperfect measurements.

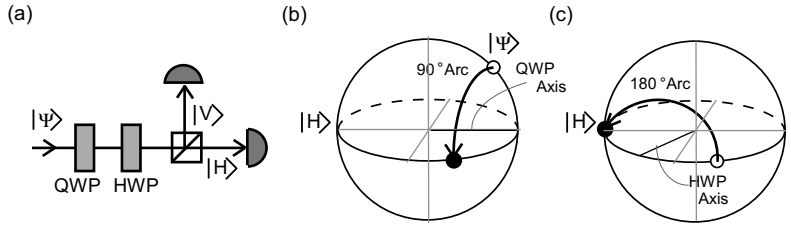


Fig. 3. A quarter-waveplate (QWP), half-waveplate (HWP), and polarizing beam splitter (PBS) are used to make an arbitrary polarization measurement. Both a diagram of the experimental apparatus (a) and the step-by-step evolution of the state on the Poincaré sphere are shown. (b) The quarter-waveplate rotates the projection state (the state we are projecting into, *not* the incoming unknown state) into the linear polarization plane (the equator). (c) The half-waveplate rotates this linear state to horizontal. The PBS transmits the projection state (now $|H\rangle$) and reflects its orthogonal compliment (now $|V\rangle$), which can then both be measured.

Exact Tomography of Multiple Qubits Using $2n$ Detectors

Tomography of multiple qubits, though an extension of the single-qubit technique, becomes more complicated and depends on the experimental apparatus used. The simplest, fastest, and most intuitive version of this tomography uses an array of $2n$ detectors, which project every incoming n-qubit state into one of 2^n basis states. This is the generalization of simultaneously measuring both outputs in the single qubit case. These detectors must measure in n-fold coincidence, and for the purposes of exact tomography it is assumed they have no errors and operate on an infinite ensemble of states. It should be emphasized that these additional detectors are not some 'trick', effectively

[7] $\theta' = \text{acos}\{\sin(\theta)\tan(\phi)\} - \text{acos}\{\cot(\theta)\cot(\phi)\}$. In practice, care must be taken that consistent conventions are used (e.g., right vs. left circular polarization), and it may be easier to calculate this angle directly from waveplate operators and the initial state.

masking a number of sequential settings of n detectors. If only n detectors are used, then over the course of a tomography most states within the input ensemble will never be measured. For example, consider measuring the projection of an unknown state into the $|00\rangle$ basis using two detectors. While this will give a number of counts, unmeasured coincidences will be routed into the $|01\rangle, |10\rangle$, and $|11\rangle$ modes. The information of how many coincidences are routed to which mode will be lost, unless another two detectors are in place in the '1' modes to measure it.

The primary advantage to using $2n$ detectors is that every setting of the analysis system (every group of the projector and its orthogonal compliments) generates exactly enough information to define a single multiple-qubit Stokes vector. Expanding out the probabilities that a multiple-qubit Stokes vector (which for now we will limit to those with only non-zero indices) is based on,

$$\begin{aligned} S_{i_1,i_2,\ldots i_n} &= \left(P_{\psi_1} - P_{\psi_1^\perp}\right) \otimes \left(P_{\psi_2} - P_{\psi_2^\perp}\right) \otimes \ldots \otimes \left(P_{\psi_n} - P_{\psi_n^\perp}\right) \\ &= P_{\psi_1,\psi_2,\ldots\psi_n} - P_{\psi_1,\psi_2,\ldots\psi_n^\perp} - \ldots \pm P_{\psi_1^\perp,\psi_2^\perp,\ldots\psi_n^\perp}, \end{aligned} \tag{51}$$

where the sign of each term on the last line is determined by the parity of the number of orthogonal ($\perp$) terms. (If instead we had included zero indices, each zero index in S would correspond to a plus sign in the first line of 51.)

These probabilities are precisely those measured by a single setting of the entire analysis system followed by a $2n$ detector array. Returning to our primary decomposition of the density matrix (equation 34),

$$\hat{\rho} = \frac{1}{2^n} \sum_{i_1,i_2,\ldots i_n=0}^{3} S_{i_1,i_2,\ldots i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \ldots \otimes \hat{\sigma}_{i_n}, \tag{52}$$

we once again need only determine all of the multiple-qubit Stokes parameters to exactly characterize the density matrix. At first glance this might seem to imply that we need to use $4^n - 1$ settings of the analysis system, in order to find all of the multiple-qubit Stokes parameters save $S_{0,0,\ldots 0}$, which is always one.

While this is certainly sufficient to solve for $\hat{\rho}$, many of these measurements are redundant. In order to choose the smallest possible number of settings, note that the probabilities that constitute some multiple-qubit Stokes parameters overlap exactly with the probabilities for other multiple-qubit Stokes parameters. Specifically, any multiple-qubit Stokes parameter with at least one 0 subscript is derived from a set of probabilities that at least one other multiple-qubit Stokes vector (with no 0 subscripts) is also derived from. As an example, consider that

$$S_{0,3} = P_{|00\rangle} - P_{|01\rangle} + P_{|10\rangle} - P_{|11\rangle}, \tag{53}$$

while

$$S_{3,3} = P_{|00\rangle} - P_{|01\rangle} - P_{|10\rangle} + P_{|11\rangle}. \tag{54}$$

The same analysis settings will provide enough information to determine both values. This dependent relationship between multiple-qubit Stokes vectors is true in general, as can be seen by returning to equation 51. Each non-zero subscript for S contributes a term to the tensor product on the right that looks like $\left(P_{\psi_i} - P_{\psi_i^\perp}\right)$. Had there been zero subscripts, however, they each would have contributed a $\left(P_{\psi_i} + P_{\psi_i^\perp}\right)$ term, which would have been totally redundant with any other term that has a non-zero term in the same index. This reduces the minimum number of analysis settings to 3^n, a huge improvement in multiple qubit systems (e.g., 9 vs. 15 settings for 2-qubit tomography, 81 vs. 255 for 4-qubit tomography, etc.). Note that, as discussed earlier, this benefit is only possible if one employs $2n$ detectors, leading to a total of 6^n measurements.[8]

Because equation 41 can be used to transform any set of non-orthogonal multiple-qubit Stokes parameters into the canonical form, orthogonal measurement sets need not be used. One advantage of the option to use non-orthogonal measurement sets is that an orthogonal set may not be experimentally acheivable, for instance, due to waveplate imperfections.

Example: A Complete Ideal 2-Qubit Tomography of Photon Pairs

Consider measuring a state with nine settings of the apparatus and four detectors, for a total of 36 measurement results. The results for this example are compiled below, with each row representing a single setting of the apparatus, and therefore a single two-qubit Stokes parameter. All measured probabilities are non-negative, but the results below have been given a minus sign if they are negated when summed into their respective two-qubit Stokes parameter.

$$
\begin{aligned}
S_{1,1} &= \underset{\frac{1}{3}}{+P_{DD}} \ \underset{\frac{1}{6}}{-P_{DA}} \ \underset{\frac{1}{6}}{-P_{AD}} \ \underset{\frac{1}{3}}{+P_{AA}} &&= \frac{1}{3} \\
S_{1,2} &= \underset{\frac{1}{4}}{+P_{DR}} \ \underset{\frac{1}{4}}{-P_{DL}} \ \underset{\frac{1}{4}}{-P_{AR}} \ \underset{\frac{1}{4}}{+P_{AL}} &&= 0 \\
S_{1,3} &= \underset{\frac{1}{4}}{+P_{DH}} \ \underset{\frac{1}{4}}{-P_{DV}} \ \underset{\frac{1}{4}}{-P_{AH}} \ \underset{\frac{1}{4}}{+P_{AV}} &&= 0 \\
S_{2,1} &= \underset{\frac{1}{4}}{+P_{RD}} \ \underset{\frac{1}{4}}{-P_{RA}} \ \underset{\frac{1}{4}}{-P_{LD}} \ \underset{\frac{1}{4}}{+P_{LA}} &&= 0 \\
S_{2,2} &= \underset{\frac{1}{6}}{+P_{RR}} \ \underset{\frac{1}{3}}{-P_{RL}} \ \underset{\frac{1}{3}}{-P_{LR}} \ \underset{\frac{1}{6}}{+P_{LL}} &&= -\frac{1}{3} \\
S_{2,3} &= \underset{\frac{1}{4}}{+P_{RH}} \ \underset{\frac{1}{4}}{-P_{RV}} \ \underset{\frac{1}{4}}{-P_{LH}} \ \underset{\frac{1}{4}}{+P_{LV}} &&= 0 \\
S_{3,1} &= \underset{\frac{1}{4}}{+P_{HD}} \ \underset{\frac{1}{4}}{-P_{HA}} \ \underset{\frac{1}{4}}{-P_{VD}} \ \underset{\frac{1}{4}}{+P_{VA}} &&= 0
\end{aligned}
$$

[8] These measurements, even though they result from the minimum number of analysis settings for $2n$ detectors, are overcomplete. A density matrix has only $4^n - 1$ free parameters, which implies that only $4^n - 1$ measurements are necessary to specify it (see n-detector tomography). Because the overcomplete set of 6^n measurements is not linearly independent, they can be reduced to a $4^n - 1$ element subset and still completely specifiy an unknown state.

$$\begin{aligned} S_{3,2} &= \underset{\frac{1}{4}}{+P_{HR}} \ \underset{\frac{1}{4}}{-P_{HL}} \ \underset{\frac{1}{4}}{-P_{VR}} \ \underset{\frac{1}{4}}{+P_{VL}} = 0 \\ S_{3,3} &= \underset{\frac{1}{3}}{+P_{HH}} \ \underset{\frac{1}{6}}{-P_{HV}} \ \underset{\frac{1}{6}}{-P_{VH}} \ \underset{\frac{1}{3}}{+P_{VV}} = \frac{1}{3} \end{aligned} \tag{55}$$

Measurements are taken at each of these nine settings, directly determining the above nine two-qubit Stokes Parameters. The six remaining required parameters, listed below, are dependent upon the same measurements.

$$\begin{aligned} S_{0,1} &= \underset{\frac{1}{3}}{+P_{DD}} \ \underset{\frac{1}{6}}{-P_{DA}} \ \underset{\frac{1}{6}}{+P_{AD}} \ \underset{\frac{1}{3}}{-P_{AA}} = 0 \\ S_{0,2} &= \underset{\frac{1}{6}}{+P_{RR}} \ \underset{\frac{1}{3}}{-P_{LR}} \ \underset{\frac{1}{3}}{+P_{RL}} \ \underset{\frac{1}{6}}{-P_{LL}} = 0 \\ S_{0,3} &= \underset{\frac{1}{3}}{+P_{HH}} \ \underset{\frac{1}{6}}{-P_{HV}} \ \underset{\frac{1}{6}}{+P_{VH}} \ \underset{\frac{1}{3}}{-P_{VV}} = 0 \\ S_{1,0} &= \underset{\frac{1}{3}}{+P_{DD}} \ \underset{\frac{1}{6}}{+P_{DA}} \ \underset{\frac{1}{6}}{-P_{AD}} \ \underset{\frac{1}{3}}{-P_{AA}} = 0 \\ S_{2,0} &= \underset{\frac{1}{6}}{+P_{RR}} \ \underset{\frac{1}{3}}{+P_{LR}} \ \underset{\frac{1}{3}}{-P_{RL}} \ \underset{\frac{1}{6}}{-P_{LL}} = 0 \\ S_{3,0} &= \underset{\frac{1}{3}}{+P_{HH}} \ \underset{\frac{1}{6}}{+P_{HV}} \ \underset{\frac{1}{6}}{-P_{VH}} \ \underset{\frac{1}{3}}{-P_{VV}} = 0 \end{aligned} \tag{56}$$

These terms will not in general be zero. Recall (c.f. equation 41) that for $|HH\rangle$, $S_{0,3} = S_{3,0} = 1$. Of course, $S_{0,0} = 1$. Taken together, these two-qubit Stokes parameters determine the density matrix.

$$\begin{aligned} \hat{\rho} &= \frac{1}{4}\left(\hat{\sigma}_0 \otimes \hat{\sigma}_0 + \frac{1}{3}\hat{\sigma}_1 \otimes \hat{\sigma}_1 - \frac{1}{3}\hat{\sigma}_2 \otimes \hat{\sigma}_2 + \frac{1}{3}\hat{\sigma}_3 \otimes \hat{\sigma}_3\right) \\ &= \frac{1}{6}\begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} + \frac{1}{6}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned} \tag{57}$$

This is the final density matrix, a Werner State, as defined in equation 28.

Exact Tomography of Multiple Qubits Using n Detectors

While the scheme outlined above is most efficient in the sense that it requires only 3^n analysis settings for the experimental apparatus, in practice it may only be experimentally possible to make a single projective measurement at a time (e.g., because the analyzer can monitor only a single outcome). As discussed earlier, this will require $4^n - 1$ probabilities in order to define a complete set of T_i parameters. In practice, this will mean that 4^n measurements are necessary in order to normalize counts to probabilities. By making a set of single projective measurements on each qubit and only taking into account those results where a definite result is obtained (e.g., the photon was transmitted by the polarizer), it is possible to reconstruct a state using only n detectors.

First define a set of $2^n \times 2^n$ matrices which have the following properties:

$$\begin{aligned} \mathrm{Tr}\left\{\hat{\Gamma}_\nu \cdot \hat{\Gamma}_\mu\right\} &= \delta_{\nu,\mu} \\ \hat{A} &= \sum_\nu \hat{\Gamma}_\nu \mathrm{Tr}\left\{\hat{\Gamma}_\nu \cdot \hat{A}\right\} \quad \forall \hat{A}, \end{aligned} \tag{58}$$

where $\hat{A}$ is an arbitrary $2^n \times 2^n$ matrix. A convenient set of $\hat{\Gamma}$ matrices to use are the tensor producted $\hat{\sigma}$ matrices used throughout this paper:

$$\hat{\Gamma}_\nu = \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \ldots \otimes \hat{\sigma}_{i_n}, \tag{59}$$

where ν is simply a short-handed index by which to label the Γ matrices, which is simpler than writing out $i_1, i_2, \ldots i_n$. Given this notation, we substitute into equation 34 and find that

$$\hat{\rho} = \frac{1}{2^n} \sum_\nu \hat{\Gamma}_\nu S_\nu. \tag{60}$$

Now that the density matrix is represented in a useful form, it is necessary to consider exactly which measurements to use. In particular, we now wish to determine the necessary and sufficient conditions on the 4^n measurements.[9] Let $|\psi_\mu\rangle$ ($\mu = 1$ to 4^n) be the measurement bases. We can then define the probability of the μ^{th} measurement as $P_\mu = \langle\psi_\mu|\hat{\rho}|\psi_\mu\rangle$.

Combining this with equation 60,

$$P_\mu = \langle\psi_\mu|\frac{1}{2^n}\sum_{\nu=1}^{4^n} \hat{\Gamma}_\nu S_\nu|\psi_\mu\rangle = \frac{1}{2^n}\sum_{\nu=1}^{4^n} B_{\mu,\nu} S_\nu, \tag{61}$$

where the $4^n \times 4^n$ matrix $B_{\mu,\nu}$ is given by

$$B_{\mu,\nu} = \langle\psi_\mu|\hat{\Gamma}_\nu|\psi_\mu\rangle. \tag{62}$$

Immediately we find a necessary and sufficient condition for the completeness of the set of tomographic states $\{|\psi_\mu\rangle\}$: if the matrix $B_{\mu,\nu}$ is nonsingular, then eq.(61) can be inverted to give

[9] While only $4^n - 1$ measurements are necessary, in practice it is often the case that exact probabilities are not known, only numbers of counts (successful measurements), with no information about the number of counts which would have been measured by detectors in orthogonal bases. In this case an extra measurement is necessary to normalize the inferred probabilities. If instead exact probabilities are known, the final result need not actually be measured, but instead can be calculated from three of the other measurements (choose this last measurement to be the n^{th} in a complete basis of measurements, and use the normalization constraint that all probabilities in a complete basis must sum to one).

$$S_\nu = 2^n \sum_{\mu=1}^{4^n} \left(B^{-1}\right)_{\mu,\nu} P_\mu. \tag{63}$$

As before, the density matrix can be calculated directly from the S_ν values. Obviously, this reconstruction technique would also work with $2n$ detectors, as the measurements gained from 3^n analysis settings are more than sufficient to define the P_μ above.

3 Real Tomography: Errors and the Maximum Likelihood Technique

The tools developed so far allow the perfect reconstruction of a density matrix from an infinite set of ideal data. When applying this technique to any set of real measurements, the assumption of ideal probabilities must be discarded. In fact, the probabilities predicted by real results can in practice be contradictory or even physically impossible. It is therefore necessary to implement a procedure that takes these errors into account yet always returns a legitimate density matrix.

3.1 Real Tomography of a Single Qubit

Once again, the single-qubit example precedes the general treatment in order to provide intuition and a graphical picture.

Types of Errors

Errors in the measurement of a density matrix fall into three main categories: errors in the measurement basis, errors from counting statistics, and errors from experimental stability. The first problem can be addressed by increasing the accuracy of the measurement apparatus (e.g., obtaining higher tolerance waveplates, better controlling the Rabi pulses, etc.) while the second problem is reduced by performing each measurement on a larger ensemble (counting for a longer time). The final difficulty is drift which occurs over the course of the tomography. [10] This drift occurs either in the state produced or the efficiency of the detection system, and can constrain the data-collection time.

[10] These are the main sources of error that are likely to be present to some degree in *any* qubit implementation. In addition, each implementation may have its own unique errors. For example, if the waveplates used to analyze optical polarization qubits are slightly wedged, and the detectors slightly non-uniform, this can lead to a troubling systematic error whereby the detector efficiency depends on the analysis settings. Other errors can be compensated, such as when accidental coincidence counts due to background light are subtracted from measured count rates. Here we neglect such system-specific difficulties.

Figure 4a shows what a basis error looks like on the Poincaré sphere and how that error affects the ability to isolate a state in Poincaré space. Accidentally measuring in a different basis, slightly different than the intended one, introduces a different amount of error depending on the state being measured. All possible measurement axes must pass through the center of the Poincaré sphere, corresponding to the fact that the totally mixed state has a probability of $\frac{1}{2}$ of being measured in any basis. Some Gaussian distribution of error around the intended axis looks like two cones meeting point to point at the center of the sphere, traced out by all of the nearby measurement axes. When translated into knowledge of the state, this transforms what would have been a single plane (perpendicular to the measurement axis) into a disk, thick at the edges of the sphere and thinning to a single point at the origin. This picture indicates that a basis error is more pronounced when measuring a pure state, but actually has no effect when measuring a totally mixed state (because all bases give the same answer).

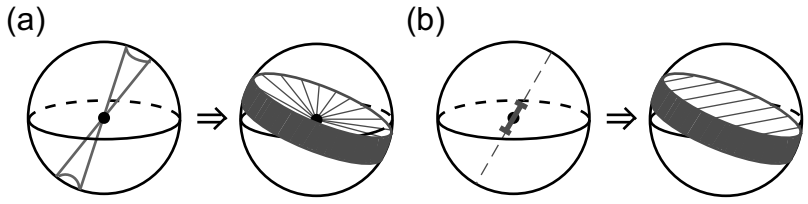


Fig. 4. Graphical representation of errors in a single-qubit tomography. (a) Basis errors. Errors in the setting of measurement apparatus can result in an accurate measurement being taken in an unintended basis. Shown graphically is the effect that an uncertainty in the measurement basis can have on the reconstruction of a state. Instead of a single axis on the Poincaré sphere, the possible measurement axes form uncertainty cones touching at the center, since all possible measurement axes pass through the origin. This uncertainty in axis is then translated into an uncertainty in the state (shown on the right). Instead of isolating the state to a plane, all possible measurement axes trace out a volume with large uncertainty near the surface of the sphere and low uncertainty near the center. (b) Counting errors. Even if the measurement basis is exactly known, only a limited number of qubits can be measured to gain an estimate of a state's projection onto this axis (taken directly from the probability of a successful measurement). This uncertainty results in an unknown state being isolated to a one-dimensional gaussian (approximately) in three-dimensional space, rather than to a plane.

Figure 4b shows the same analysis of errors in counting statistics. Any real measurement can only be carried out on a limited size ensemble. Though the details of the statistics will be dealt with later, the detection events are accurately described by a Poissonian distribution, which for large numbers of counts is well approximated by a Gaussian distribution. This will cause the resultant knowledge about the unknown state to change from a plane (in the

exact case) to a thick disk (uniformally thick for pure and mixed states), a one-dimensional Gaussian distribution plotted in three-dimensional space.

The final source of error, that of drift, can occur on three distinct timescales. Quickly varying systems experience drift which cycles many times within the span of a single measurement. Rather than producing errors, these types of drift result in the measurement of a mixed state (which describes a probabilistic mixture of several states, exactly the situation when a state drifts). Very slowly varying systems also do not introduce errors, as the timescale of the drift exceeds the timescale of the measurements (during the measurement process, the state is essentially constant). The final source of drift occurs on the timescale of the measurement, allowing a totally different state to be measured from one analysis setting to the next. This error is difficult to characterize, as it is difficult to know what the correct answer for a tomography would be, much less how far the predicted value strayed from it.

Drift in state *intensity* (e.g., the rate of photons produced) can be compensated for by employing to a $2n$ detector system. Because every member of the ensemble is measured in a complete basis, this gives an exact intensity for each set of 2^n measurements, eliminating the need for the assumption that each measurement run is over the same size ensemble. (Recall that a *measurement* gives the probability to project the unknown ρ into a single state $|\psi\rangle$. When using $2n$ detectors, therefore, each analysis setting will yield 2^n measurement results.)

Detector efficiency drift also presents a problem, with similar effects to state drift. Special care must be taken to address this problem when using a $2n$ detector scheme, as measurements taken with each pair of detectors may appear to give different values, when in fact a particular detector pair is simply more efficient than another. In order to compensate for this, the relative efficiencies of each detector pair must be measured, and all counts taken renormalized using these efficiencies.

Combining Real Measurements: State Estimation

After all sources of error are taken into account, a single measurement results in a distribution over all possible states describing the experimenter's knowledge of the unknown state. This distribution represents the likelihood that a particular state would give the measured results, relative to another state. When independent measurements are combined, these distributions are multiplied, and ideally the knowledge of the unknown state is restricted to a small ball in Poincaré space, approximately equal to a three-dimensional Gaussian. This type of state isolation occurs regardless of which measurements are taken, as long as they are linearly independent. State isolation is shown graphically in figure 5 for a set of orthogonal measurements.

In contrast to the ideal case in the previous section, for which the accuracy of a reconstructed state did not depend on whether orthogonal or

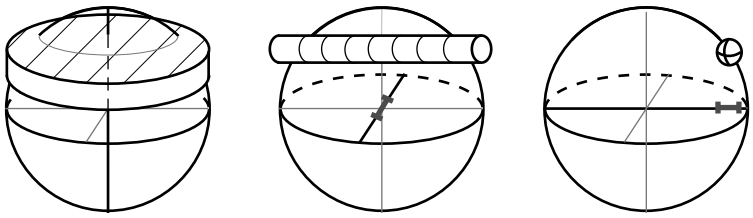


Fig. 5. Isolation of a quantum state through inexact measurements. Although a series of real measurements (those with uncertainties) will never be able to exactly isolate an unknown quantum state, they can isolate it to a region of Hilbert space that is far more likely than any other region to contain the unknown state. Consider a series of three measurements, each containing counting errors, along orthogonal axes. From left to right, the area of Hilbert space containing the unknown state is truncated from a one-dimensional Gaussian probability distribution (the disk in the left figure) to a two-dimensional Gaussian (the cylinder in the middle figure) and finally to a three-dimensional Gaussian (the ball in the right figure). This results in an 'error ball' which approximates the position of the unknown state. The global maximum, however, can often be outside allowed Hilbert space (outside the Poincaré sphere), which is one reason a maximum likelihood technique must be used to search over only allowed quantum states.

non-orthogonal measurements were made, with real measurements the advantage of orthogonal measurements becomes clear. Unlike in figure 5, non-orthogonal measurements result in a non-symmetric error ball, increasing the error in state estimation in one direction in Hilbert space. In practice, making the right measurements for the right amount of time to make this error ball spherical can be acheived by adaptive tomography [37], but this is beyond the scope of this work.

Even after tomography returns a distribution of likelihood over Poincaré space, one final problem remains. It is very possible, especially with low counts or with the measurement of very pure states, that state estimation will return an illegal state. For example, in figure 5, the measurements seem to place the error ball just on the edge of the sphere and slightly outside it. As all legal states have a radius of less than or equal to one in Poincaré space, it is necessary to find a way to return the most likely *legitimate* state reconstructed from a set of measurements.

The Maximum Likelihood Technique

The problem of reconstructing illegal density matrices is resolved by selecting the legitimate state most likely to have returned the measured counts [28, 29]. In practice, analytically calculating this maximally likely state is prohibitively difficult, and a numerical search is necessary. Three elements are required: a manifestly legal parametrization of a density matrix, a like-

lihood function which can be maximized, and a technique for numerically finding this maximum over a search of the density matrix's parameters.

The Stokes parameters are an unacceptable parametrization for this search, as there are clearly legitimate combinations of these parameters which result in an illegal state. In this context, a legitimate state refers to a non-negative definite Hermitian density matrix of trace one. The property of non-negative definiteness for any matrix $\hat{\mathcal{G}}$ is written mathematically as

$$\langle\psi|\hat{\mathcal{G}}|\psi\rangle \geq 0 \quad \forall|\psi\rangle. \tag{64}$$

Any matrix that can be written in the form $\hat{\mathcal{G}} = \hat{T}^\dagger\hat{T}$ must be non-negative definite. To see that this is the case, substitute into eq.(64):

$$\langle\psi|\hat{T}^\dagger\hat{T}|\psi\rangle = \langle\psi'|\psi'\rangle \geq 0, \tag{65}$$

where we have defined $|\psi'\rangle = \hat{T}|\psi\rangle$. Furthermore $(\hat{T}^\dagger\hat{T})^\dagger = \hat{T}^\dagger(\hat{T}^\dagger)^\dagger = \hat{T}^\dagger\hat{T}$, i.e., $\hat{\mathcal{G}} = \hat{T}^\dagger\hat{T}$ must be Hermitian. To ensure normalization, one can simply divide by the trace. Thus the matrix $\hat{g}$ given by the formula

$$\hat{g} = \hat{T}^\dagger\hat{T}/\mathrm{Tr}\{\hat{T}^\dagger\hat{T}\} \tag{66}$$

has all three of the mathematical properties required for density matrices.

For the one-qubit system, we have a 2×2 density matrix with 3 independent real parameters (although we will search over 4 in order to fit the intensity of the data). Since it will be useful to be able to invert relation (66), it is convenient to choose a tri-diagonal form for $\hat{T}$:

$$\hat{T}(t) = \begin{pmatrix} t_1 & 0 \\ t_3 + it_4 & t_2 \end{pmatrix}. \tag{67}$$

The manifestly 'physical' density matrix $\hat{\rho}_p$ is then given by the formula

$$\hat{\rho}_p(t) = \hat{T}^\dagger(t)\hat{T}(t)/\mathrm{Tr}\{\hat{T}^\dagger(t)\hat{T}(t)\}, \tag{68}$$

where t is shorthand for all t_i.

This satisfies the first criterion for a successful maximum likelihood search, by providing an explicitly physical parametrization for $\hat{\rho}$. The second criterion, a likelihood function, will in general depend on the specific measurement apparatus used and the physical implementation of the qubit (as these will determine the statistical distributions of counts, and therefore their relative weightings).

If we assume both Gaussian counting statistics and that each of our measurements is taken for the same amount of time, then we can provide a suitable likelihood function. (Basis errors are neglected on the assumption that they are symmetric about a known central basis, and so will not affect the outcome, though they *will* affect the error on that outcome.)

Let n_ν be the result for the ν^{th} measurement, out of a total of Ξ measurements. For the two techniques presented in the previous section applied to measurement of a single-qubit, Ξ would equal four for one detector and six for two detectors (with two detectors there are *three* analysis settings but *six* measurements). The expected values for these measurements on an unknown state $\hat{\rho}$ are given by $\bar{n}_\nu = \mathcal{N}\langle\psi_\nu|\hat{\rho}|\psi_\nu\rangle$. Here $\mathcal{N}$ is an unknown normalization parameter defined by $\frac{n_\nu}{\mathcal{N}} = P_\nu$, corresponding to the total size per measurement of the ensemble. (It is not always possible to know the size of a measured ensemble, and so the counts rather than the probabilities are used in the likelihood function. In addition, the Gaussian statistical distribution is over counts, not probabilities.) Given these definitions, the probability of obtaining the observed experimental counts n_ν from the density matrix $\hat{\rho}$ is

$$P\left(n_1, n_2, \ldots n_\Xi\right) = \frac{1}{Norm} \prod_\nu \exp\left[-\frac{\left(\bar{n}_\nu - n_\nu\right)^2}{2\hat{\sigma}_\nu^2}\right], \tag{69}$$

where $\hat{\sigma}_\nu$ is the standard deviation of the ν^{th} measurement (given approximately by $\sqrt{\bar{n}_\nu}$) and *Norm* is the normalization constant. For our candidate physical density matrix $\hat{\rho}_p$ the number of counts expected for the ν-th measurement is

$$\bar{n}_\nu\left(t_1, t_2, \ldots t_{n^2}\right) = \mathcal{N}\langle\psi_\nu|\hat{\rho}_p\left(t_1, t_2, \ldots t_{n^2}\right)|\psi_\nu\rangle. \tag{70}$$

Thus the likelihood that the matrix $\hat{\rho}_p\left(t_1, t_2, \ldots t_{n^2}\right)$ could produce the measured data $\{n_1, n_2, \ldots n_\Xi\}$ is

$$P\left(n_1, n_2, \ldots n_\Xi\right) = \frac{1}{Norm} \prod_\nu \exp\left[-\frac{\left(\mathcal{N}\langle\psi_\nu|\hat{\rho}_p\left(t_1, t_2, \ldots t_{n^2}\right)|\psi_\nu\rangle - n_\nu\right)^2}{2\mathcal{N}\langle\psi_\nu|\hat{\rho}_p\left(t_1, t_2, \ldots t_{n^2}\right)|\psi_\nu\rangle}\right]. \tag{71}$$

Note that here we assume that $\mathcal{N}$ is the same for each measurement. In practice this may not necessarily be the case due to drift (in either source intensity or detector efficiencies) or differing measurement conditions.

Rather than find the maximum value of $P\left(t_1, t_2, \ldots t_{n^2}\right)$, it is somewhat simpler to find the maximum of its logarithm (which is mathematically equivalent). In addition, because $\mathcal{N}$ is unknown, we absorb it into the $\hat{T}$ matrix, by setting

$$t_i' = \mathcal{N} t_i. \tag{72}$$

Thus the optimization problem reduces to finding the *minimum* of the following function:

$$\mathcal{L}\left(t_1', t_2', \ldots t_{n^2}'\right) = \sum_\nu \frac{\left(\langle\psi_\nu|\hat{\rho}_p\left(t_1', t_2', \ldots t_{n^2}'\right)|\psi_\nu\rangle - n_\nu\right)^2}{2\langle\psi_\nu|\hat{\rho}_p\left(t_1', t_2', \ldots t_{n^2}'\right)|\psi_\nu\rangle}. \tag{73}$$

The final piece in the maximum likelihood technique is an optimization routine, of which there are many available. The authors used the `Matlab`

function `fminunc` for the examples given in this chapter.[11] After a minimum is found, $\hat{\rho}$ can be reconstructed from the values of t'.

Note that the maximum likelihood technique easily adapts to measurements in non-orthogonal bases (e.g., due to imperfect yet well chacterized waveplates) and overcomplete measurements (taking more measurements than is necessary). In the first case the set of $|\psi\rangle$ is simply non-orthogonal with no effect on the above equations; in the second case the sum in equation 73 is extended beyond the minimum Ξ.

An Example Using Photons

Photon pairs generated via spontaneous parametric downconversion from a nonlinear crystal can be used to generate single photon states. Measuring a photon in one arm collapses the state of its partner to a single qubit Fock state [33]. An ensemble of these photons can be used to implement the maximum likelihood technique. See figure 6 for the measured counts and the final density matrix. While these counts would have resulted in an illegal density matrix using the simple linear reconstruction of section 2, the maximum likelihood technique returns the most likely legal state to have given this data.

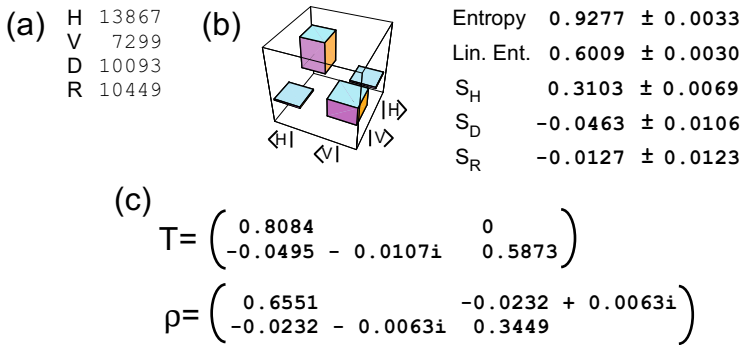


Fig. 6. Example results for the maximum likelihood technique, taken from single-qubit data. The data was taken using downconverted photons from a nonlinear crystal, using one half of a photon pair as a trigger. (a) The measured counts in four linear independent bases. (b) The likelihood function is minimized for these counts, yielding the listed density matrix, shown graphically. The entropy, linear entropy, and the three normalized Stokes parameters are shown. The errors on these values were calculated using a Monte Carlo simulation of the data in (a). (c) Numerical results. Both the $\hat{T}$ matrix, calculated using the maximum likelihood technique. and the resulting density matrix $\hat{\rho}$ are shown,

[11] For freely available code and further examples, see:
`http://www.physics.uiuc.edu/research/QuantumPhotonics/Tomography/`

3.2 The Multiple Qubit Maximum Likelihood Technique

Extending the maximum likelihood technique into the multiple-qubit regime is surprisingly straightforward. Virtually all of the mathematics from the previous section is still applicable, with minor changes. In fact, after rewriting equation 68 to be more general:

$$\hat{T}(t) = \begin{pmatrix} t_1 & 0 & \dots & 0 \\ t_{2^n+1} + it_{2^n+2} & t_2 & \dots & 0 \\ \dots & \dots & \dots & 0 \\ t_{4^n-1} + it_{4^n} & t_{4^n-3} + it_{4^n-2} & t_{4^n-5} + it_{4^n-4} & t_{2^n} \end{pmatrix}, \tag{74}$$

all other equations in the previous section are correct, and the procedure for isolating the reconstructed density matrix is exactly the same.

A 2-Qubit Example in Photons

In the 2-qubit case, equation 74 becomes

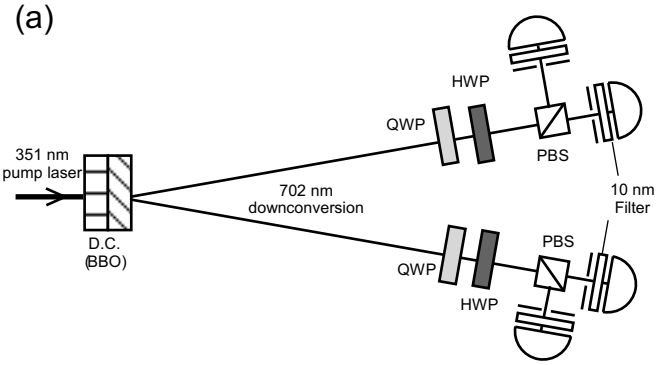
$$\hat{T}(t) = \begin{pmatrix} t_1 & 0 & 0 & 0 \\ t_5 + it_6 & t_2 & 0 & 0 \\ t_{11} + it_{12} & t_7 + it_8 & t_3 & 0 \\ t_{15} + it_{16} & t_{13} + it_{14} & t_9 + it_{10} & t_4 \end{pmatrix}. \tag{75}$$

Figure 7 shows a set of real data and its analysis. This data represents two-qubit polarization states produced via spontaneous parametric downconversion from two adjacent nonlinear crystals. The resulting uncertainty as to which crystal generated the pair leads to a nearly ideal maximally entangled state [27]. The data shown was measured using the four-detector technique; 9 analysis settings were required, for a total of 36 measurement results.

4 Error Analysis

Error analysis of reconstructed density matrices is in practice a non-trivial process. The traditional method of error analysis involves analytically solving for the error in each measurement due to each source of error, then propagating these errors through a calculation of any derived quantity. In the photon case, for example, errors in counting statistics and waveplate settings are analyzed in some detail in reference [28], giving errors in both density matrices and commonly derived quantities, such as the tangle and the linear entropy. In practice, however, these errors appear to be too large: We have experimentally repeated some of our measurements many times, and observed a spread in the value of derived quantities which is approximately an order of magnitude smaller that the spread predicted from an analytic calculation of the uncertainty. Thus it is worthwhile to discuss alternate methods of error analysis.

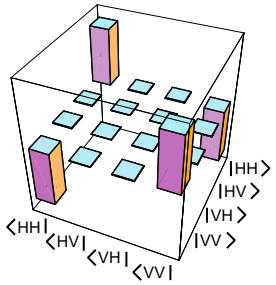
(a)



(b)

HH 40230	HV 150	VH 220	VV 39094
HD 19970	HA 20533	VD 18542	VA 20050
HR 21988	HL 18910	VR 18003	VL 21054
DH 20556	DV 19706	AH 20010	AV 18629
DD 333	DA 39918	AD 38011	AA 440
DR 20475	DL 19975	AR 18931	AL 19544
RH 19900	RV 19649	LH 20761	LV 19330
RD 17423	RA 21755	LD 21207	LA 18804
RR 38717	RL 401	LR 483	LL 39382

(c)



Tangle	0.9519 ± 0.0013
Entropy	0.1033 ± 0.0026
Lin. Ent.	0.0318 ± 0.0008
Fidelity	0.9868 ± 0.0001

$$\begin{pmatrix} 0.509 & -0.006 - 0.012i & 0.008 + 0.011i & -0.489 + 0.017i \\ -0.006 + 0.012i & 0.002 & -0.000 + 0.002i & 0.004 - 0.008i \\ 0.009 - 0.011i & -0.000 - 0.002i & 0.003 & -0.007 + 0.014i \\ -0.489 - 0.017i & 0.004 + 0.008i & -0.007 - 0.014i & 0.486 \end{pmatrix}$$

Fig. 7. (a) Experimental setup for producing two-qubit polarization states from spontaneous parameteric downconversion. (b) Experimental data for a near-Bell state ($\tilde{9}9\%$ fidelity with $|\phi^-\rangle = \frac{1}{\sqrt{2}}(|HH\rangle - |VV\rangle)$). Shown are the single photon counts from a complete tomography using four detectors. Nine analysis settings yields the 36 measurement results shown. These counts are adjusted for accidentals (dark counts) and renormalized for differences in detector efficiencies (see text for details). (c) The density matrix is shown in both numeric and graphical form, along with several quantities derived from that matrix. The errors shown were calculated using a Monte Carlo simulation of the data from (b).

One promising numerical method is the 'Monte Carlo' technique, whereby additional numerically simulated data is used to provide a statistical distribution over any derived quantity. Once an error distribution is understood over a single measurement (e.g., Gaussian for waveplate setting errors or Poissonian over count statistics), a set of 'simulated' results can be generated. These results are simulated using the known error distributions in such a way as to produce a full set of numerically generated data which could feasibly have come from the same system. Many of these sets of data are numerically generated (at the measured counts level), and each set is used to calculate a density matrix via the maximum likelihood technique. This set of density matrices is used to calculate the standard error on any quantity implicit in or derived from the density matrix.

As an example, consider the application of the Monte Carlo technique to the downconversion results from figure 7. Two polarization encoded qubits are generated within ensembles that obey Poissonian statistics, and these ensembles are used to generate a density matrix using the maximum likelihood technique. In order to find the error on a quantity derived from this density matrix (e.g., the tangle), 36 new measurement results are numerically generated, each drawn randomly from a Poissonian distribution with mean equal to the original number of counts. These 36 numerically generated results are then fed into the maximum likelihood technique, in order to generate a new density matrix, from which, e.g., the tangle may be calculated. This process is repeated many times, generating both many density matrices and a distribution of tangle values, from which the error in the initial tangle may be determined. In practice, additional sets of simulated data must be generated until the error on the quantity of interest converges to a single value. For the data in figures 6 and 7, a total of 100 simulations were used.

Clearly, the problem of error analysis in state tomography is an area of continuing research – one of many. The development of adaptive tomography techniques could allow both specific measurements and the data collection times to be tailored in order to optimize for each state to be measured [37]. In addition, because the number of measurements necessary to perform tomography grows exponentially with the number of qubits, it will eventually be necessary to partially characterize states with fewer measurements. Finally, each distinct qubit implementation provides a myriad of unique challenges. Nevertheless, we hope the discussions presented here will be useful for characterizing quantum systems in a broad spectrum of qubit realizations.

The authors would like to thank Evan Jeffrey, Nick Peters, Aaron Vandevender, and Tzu-Cheih Wei for helpful conversations and invaluable assistance in the preparation of this manuscript. The development for this work was supported over several years by grants from the National Science Foundation, the National Security Agency, the Advanced Research and Development Activity, the DCI Postdoctoral Program, and most recently, the Army Re-

search Office-sponsored MURI: 'Center for Photonic Quantum Information Systems'.

References

1. A. G. White, D. F. V. James, P. H. Eberhard and P. G. Kwiat: Phys. Rev. Lett. **83**, 3103 (1999)
2. K. Sanaka, K Kawahara, and T. Kuga: Physical Review Letters **86** 5620 (2001)
3. A. Mair, A. Vaziri, G. Weihs, A. Zeilinger: Nature **412**, 3123 (2001)
4. Y. Nambu, K. Usami, Y. Tsuda, K. Matsumoto, and K. Nakamura: Phys. Rev. A **66**, 033816 (2002)
5. G. Giorgi, G. Di Nepi, P. Mataloni, and F De Martini: Laser Phys. **13** 350 (2003)
6. T. Yamamoto, M. Koashi, S.K. Özdemir, and N. Imoto: Nature **421** 343 (2003)
7. A.V. Sergienko, G. Di Giuseppe, M. Atatüre, et al: Entangled-Photon State Engineering. In: Proceedings of the Sixth International Conference on Quantum Communication, Measurement and Computing (QCMC), ed by J.H. Shapiro and O. Hirota (Rinton, Princeton, 2003) pp 147–152
8. T. B. Pittman, M. J. Fitch, B. C. Jacobs, and J. D. Franson: /quant-ph/0303095 (2003)
9. J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, and D. Branning: "Demonstration of an all-optical controlled-NOT gate" to appear in Nature, (2003)
10. I. Marcikic, H. de Riedmatten, W. Tittel, H. Zbinden, N. Gisin: Nature **421**, 509 (2003)
11. D. G. Cory, A. F. Fahmy, and T. F. Havel: Proc Natl. Acad. Sci. U.S.A. **94**, 1634 (1997)
12. J. A. Jones, M. Mosca, and R. H. Hansen: Nature **392**, 344 (1997)
13. Y. S. Weinstein, M. A. Pravia, E. M. Fortunato, S. Lloyd, and D. G. Cory: Physical Review Letters **86**, 1889 (2001)
14. R. Laflamme et al.: quant-ph/0207172
15. C. Monroe: Nature **416** 238 (2002)
16. R. Blatt et al.: Nature **422**, 408 (2003)
17. J. Lehner, U. Leonhardt, and H. Paul: Phys. Rev. A **53**, 2727 (1996)
18. P. G. Kwiat and B. Englert, *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity*, Chapter 15, Cambridge Univ. Press (2004)
19. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge Univ. Press (2000)
20. S. Gasiorowitz, *Quantum Physics*, John Wiley and Sons, Inc. (1996)
21. G. C. Stokes, *Trans. Cambr. Phil. Soc.* **9**, 399 (1852)
22. N. Peters, J. Altepeter, E. Jeffrey, D. Branning, and P. G. Kwiat: Quantum Information and Computation **3**, 503 (2003)
23. J. S. Bell, Physics **1**, 195 (1965)
24. R. Werner: Phys. Rev. A **40**, 4277 (1989)
25. W. J. Munro, D. F. V. James, A. G. White, P. G. Kwiat: Phys. Rev. A **64**, 030302 (2001)
26. M. Born and E. Wolf, *Principles of Optics*, Pergamon Press (1987)

27. P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P. H. Eberhard: Phys. Rev. A **60**, R773 (1999)
28. D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White: Phys. Rev. A **64**, 052312 (2001)
29. Z. Hradil and J. Rehacek: Fortschr. Phys. **49**, 1083 (2001)
30. A. F. Abouraddy, A. V. Sergienko, B. E. A. Saleh, and M. C. Teich: Optics Communications **201**, 93 (2002)
31. K. Zyczkowski: quant-ph/0006068, quant-ph/0108064
32. D. Kielpinski, V. Meyer, M. A. Rowe, C. A. Sackett, W. M. Itano, C. Monroe, D. J. Wineland: Science **291**, 1013 (2001)
33. C. K. Hong and L. Mandel: Phys. Rev. Lett. **56**, 58 (1986)
34. W. K. Wootters: Phys. Rev. Lett. **80**, 2245 (1998); V. Coffman, J. Kundu, W. K. Wootters: Phys. Rev. A **61**, 052306 (2000)
35. B. M. Terhal: quant-ph/0101032
36. J. M. G. Sancho and S. F. Huelga: Phys. Rev. A **61**, 042303 (2000)
37. G. M. D'Ariano, M. G. A. Paris, M. F. Sacchi: quant-ph/0302028