Minimal qubit tomography

Jaroslav Řeháček,[1] Berthold-Georg Englert,[2] and Dagomir Kaszlikowski[2]

[1]*Department of Optics, Palacky University, 17. listopadu 50, 772 00 Olomouc, Czech Republic*
[2]*Department of Physics, National University of Singapore, Singapore 117542, Singapore*

We present, and analyze thoroughly, a highly symmetric and efficient scheme for the determination of a single-qubit state, such as the polarization properties of photons emitted by a single-photon source. In our scheme there are only four measured probabilities, just enough for the determination of the three parameters that specify the qubit state, whereas the standard procedure would measure six probabilities.

I. INTRODUCTION

Experiments that exploit the polarization degree of freedom of single photons, detected one by one, have become an almost routine matter in recent years. In particular, a whole class of experiments that demonstrate the technical feasibility of quantum cryptography, or quantum key distribution, use the photon polarization as the carrier of the quantum bit, or *qubit*. Other experiments make use of a spatial degree of freedom, essentially the path qubit of a two-path interferometer, which is sometimes translated into the alternative of early or late arrival for the sake of easier transmission.

In applications like these, as well as many others, one must be able to characterize the qubit source and the transmission channel. For this purpose a complete determination of the state of the qubit is required, both as it is emitted from the source and as it arrives after transmission. To be able to perform the regular on-the-fly calibration of the setup, so as to compensate for the unavoidable drifts, one needs an efficient diagnostics that does not consume more qubits than really necessary.

The standard procedure measures three orthogonal components of the relevant qubit analog of Pauli's spin vector operator, so that *six* probabilities are estimated for the determination of the *three* real parameters that specify the qubit state. But clearly, *four* measured probabilities should suffice to establish the values of three parameters. Indeed, such minimal schemes for state determination are possible, and it is the objective of this paper to present and analyze one such scheme, a highly symmetric one.

In Sec. II we briefly review the standard six-output measurement scheme and then introduce the minimal four-output scheme, followed by remarks on state determination for qubit pairs. We then proceed to describe, in Sec. III, optical implementations for the measurement of a photon's polarization qubit—polarimeters or ellipsometers in the jargon of classical optics.

The question of how one infers a reliable estimate for the qubit state after the detection of a finite, possibly small, number of qubits is addressed in Sec. IV. After discussing the optimality of the highly symmetric four-output scheme in Sec. V and remarking on some peculiar aspects of measuring pure qubit states in Sec. VI, we analyze adaptive measurement strategies in Sec. VII and then close with a summary.

II. QUBIT TOMOGRAPHY

A. Standard six-state tomography

We describe, as usual, the binary quantum alternative of the qubit by a Pauli vector operator $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. The physical nature of the qubit is irrelevant for the present general discussion—it might just as well be the spin-$\frac{1}{2}$ degree of freedom of an electron or a pseudospin such as the path degree of freedom in a two-path interferometer or the internal degree of freedom of a two-level atom—but in the particular application that we have in mind it is the polarization degree of freedom of a photon. Then we use the convention specified by

$$\sigma_x = |h\rangle\langle v| + |v\rangle\langle h|,$$

$$\sigma_y = i(|h\rangle\langle v| - |v\rangle\langle h|),$$

$$\sigma_z = |v\rangle\langle v| - |h\rangle\langle h|, \tag{2.1}$$

where $|v\rangle$ and $|h\rangle$ are the ket vectors for vertical and horizontal polarization, respectively.

The statistical operator of the qubit emitted by a given source,

$$\rho = \frac{1}{2}(1 + \vec{s} \cdot \vec{\sigma}), \tag{2.2}$$

is parametrized by the Pauli vector $\langle\vec{s}\rangle = \langle\vec{\sigma}\rangle = \mathrm{tr}\{\vec{\sigma}\rho\}$, the expectation value of $\vec{\sigma}$. The positivity of ρ restricts the Pauli vectors to the Bloch ball, $s = |\vec{s}| \leq 1$. The experimental characterization of the source requires, therefore, a complete measurement of $\vec{s}$ with sufficient precision. Any procedure that can yield this information is an example of *qubit tomography*.

In the standard approach one measures σ_x for some qubits supplied by the source, σ_y for some others, and σ_z for yet others. Assuming an unbiased procedure—that is, for each qubit there is an equal chance for either one of the three measurements to happen—there are six possible outcomes that occur with the probabilities
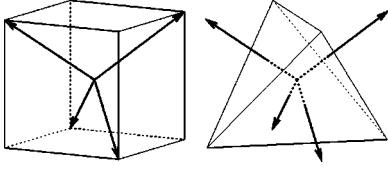
FIG. 1. Picturing the vector quartet of Eq. (2.5) as pointing from the center to four nonadjacent corners of a cube or as the vectors normal on the faces of a tetrahedron.

$$p_{\xi\pm} = \left\langle \frac{1}{6}(1 \pm \sigma_\xi) \right\rangle \equiv \langle P_{\xi\pm} \rangle \text{ for } \xi = x, y, z. \quad (2.3)$$

Each operator $P_{\xi\pm}$ is a third of a projector, and since these non-negative operators decompose the identity,

$$\sum_{\xi=x,y,z} (P_{\xi+} + P_{\xi-}) = 1, \quad (2.4)$$

they constitute the positive operator-valued measure (POVM) of this standard *six-state tomography*.

This POVM is an example of a tomographically complete set of measurements of pairwise complementary observables; namely, σ_x, σ_y, and σ_z, so that their eigenstates constitute sets of mutually unbiased bases. As Wootters and Fields have shown [1], such sets are particularly well suited for tomographic purposes, inasmuch as the statistical errors in the estimates based on a finite number of measurements are minimal. The sets themselves are not of minimal size, however, because one measures six probabilities to determine three parameters, the components of the Pauli vector $\vec{s}$. Indeed, the six probabilities of Eq. (2.3) are subject to the three constraints $p_{\xi+} + p_{\xi-} = \frac{1}{3}$, $\xi = x, y, z$, rather than to the single constraint of unit sum. A minimal POVM, by contrast, would refer to only four outcomes and their probabilities, with unit sum as the only constraint.

B. Minimal four-state tomography

We construct such a minimal POVM of high internal symmetry by first choosing four unit vectors $\vec{a}_1, \dots, \vec{a}_4$ with equal angle between each pair:

$$\vec{a}_j \cdot \vec{a}_k = \frac{4}{3}\delta_{jk} - \frac{1}{3} = \begin{cases} 1 & \text{for } j = k, \\ -1/3 & \text{for } j \neq k. \end{cases} \quad (2.5)$$

Geometrically speaking, such a quartet consists of the vectors pointing from the center of a cube to nonadjacent corners, as illustrated in Fig. 1 and exemplified by

$$\vec{a}_1 = 3^{-1/2}(1, 1, 1),$$

$$\vec{a}_2 = 3^{-1/2}(1, -1, -1),$$

$$\vec{a}_3 = 3^{-1/2}(-1, 1, -1),$$

$$\vec{a}_4 = 3^{-1/2}(-1, -1, 1). \quad (2.6)$$

Alternatively, one may picture these vectors as the normal vectors for the faces of the tetrahedron that is defined by the other four corner of the cube.

The linear dependence of the $\vec{a}_j$'s is stated by their null sum

$$\sum_{j=1}^{4} \vec{a}_j = 0 \quad (2.7)$$

and their completeness by the decomposition of the unit dyadic:

$$\frac{3}{4} \sum_{j=1}^{4} \vec{a}_j \vec{a}_j = \overset{\leftrightarrow}{1}. \quad (2.8)$$

The perfect symmetry of the tetrahedron geometry manifests itself in the simplicity of this completeness relation and the inner products of Eq. (2.5). As discussed in Sec. V B below, the tetrahedron geometry is optimal in the sense that any other vector quartet would define a less efficient scheme for four-state tomography.

Each such quartet of $\vec{a}_j$'s defines a POVM for minimal four-state tomography in accordance with

$$\sum_{j=1}^{4} P_j = 1 \text{ with } P_j \equiv \frac{1}{4}(1 + \vec{a}_j \cdot \vec{\sigma}). \quad (2.9)$$

This POVM is an example of a "symmetric informationally complete POVM" [2]. Upon measuring it and so determining the probabilities [3]

$$p_j = \langle P_j \rangle = \frac{1}{4}(1 + \vec{a}_j \cdot \vec{s}), \quad (2.10)$$

the Pauli vector is readily available,

$$\vec{s} = 3 \sum_j p_j \vec{a}_j, \quad (2.11)$$

and so are the statistical operator and its square,

$$\rho = 6 \sum_j p_j P_j - 1 = \sum_j \langle P_j \rangle (6 P_j - 1),$$

$$\rho^2 = \rho - 1 + 3 \sum_j p_j^2. \quad (2.12)$$

It follows that, in addition to being restricted to the range $0 \leq p_j \leq \frac{1}{2}$, the probabilities p_j obey the inequalities

$$\frac{1}{4} \leq \sum_j p_j^2 = \frac{3 + s^2}{12} \leq \frac{1}{3}. \quad (2.13)$$

The upper bound is reached by all pure states, $\rho = \rho^2$ and $s = 1$, the lower bound for the completely mixed state, $\rho = \frac{1}{2}$ and $s = 0$.

C. Qubit-pair tomography

The minimal property of the four-state POVM of Eq. (2.9) carries over to multiqubit states. In case of n qubits, one has 4^n joint probabilities for the $4^n - 1$ independent parameters of the $2^n \times 2^n$ matrix elements of the statistical operator, so that the count is just right. By contrast, if one were to measure n realizations of the six-state POVM of Eq. (2.4), one would

have 6^n joint probabilities which contain quite a lot of redundant information.

More specifically, consider the $n=2$ situation of a source emitting qubit pairs. Using the P_j's from above for one qubit and corresponding operators Q_k for the other, we obtain the 16 joint probabilities $\langle P_j Q_k \rangle$ by measuring the two four-state POVMs. They are the numerical ingredients in

$$\rho = \sum_{j,k} (6P_j - 1)\langle P_j Q_k \rangle (6Q_k - 1), \qquad (2.14)$$

the explicit construction of the two-qubit statistical operator. If there are no correlations in the joint probabilities, so that $\langle P_j Q_k \rangle = \langle P_j \rangle \langle Q_k \rangle$, then ρ is the product of two factors of the one-qubit form in Eq. (2.12), as it should be. The generalization of the $n=2$ expression (2.14) to $n>2$ is immediate.

Qubit-pair tomography of this kind requires that the vector quartet $\vec{b}_k$ associated with the Q_k's have a known orientation relative to the quartet $\vec{a}_j$ of the P_j's. One can determine this orientation by "quantum measurement tomography" [4]; that is, by measuring the joint probabilities $\langle P_j Q_k \rangle$ for a source with a known output [5]. In the simplest situation, for example, the source emits pairs that are perfectly anticorrelated:

$$\rho = \frac{1}{4}(1 - \vec{\sigma}^{(1)} \cdot \vec{\sigma}^{(2)}). \qquad (2.15)$$

The orthogonal dyadic $\overleftrightarrow{O}$ that turns the $\vec{a}_j$ quartet and the $\vec{b}_k$ quartet into each other,

$$\vec{b}_k = \overleftrightarrow{O} \cdot \vec{a}_k, \quad \vec{a}_j = \vec{b}_j \cdot \overleftrightarrow{O}, \qquad (2.16)$$

is then given by

$$\overleftrightarrow{O} = -9 \sum_{j,k} \vec{a}_j \langle P_j Q_k \rangle \vec{a}_k = -9 \sum_{j,k} \vec{b}_j \langle P_j Q_k \rangle \vec{b}_k, \qquad (2.17)$$

and can thus be determined experimentally.

III. ELLIPSOMETRY

A. Standard six-outcome ellipsometer

Devices for characterizing the polarization properties of a light source are called *ellipsometers*, a term that makes reference to elliptic polarization as the generic outcome of the measurement. Figure 2 shows the schematic setup of a standard six-state ellipsometer, which realizes the POVM of Eqs. (2.3) and (2.4) for the photonic polarization qubit. The input beam splitter (BS) reflects one-third of the light to a polarizing BS (PBS) that reflects vertically polarized photons and transmits horizontally polarized ones and so directs them to two photodetectors. This branch thus realizes a measurement of σ_z and accounts for P_{z+} and P_{z-} in the sum of Eq. (2.4).

Two-thirds of the light are transmitted at the input BS and are then equally split at a second BS. A photon transmitted there will be detected by either one of the two detectors behind another PBS. This PBS is rotated by 45°, so that a measurement of σ_x is realized in this branch and the term $P_{x+} + P_{x-}$ is accounted for in Eq. (2.4).
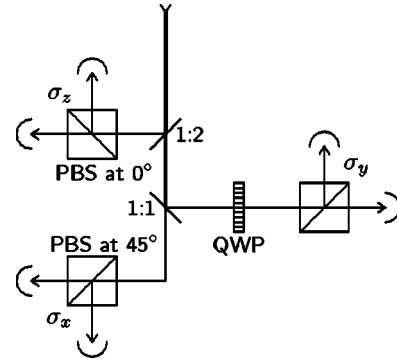

FIG. 2. Schematic setup of an ellipsometer that implements the standard six-state POVM; see text.

Finally, the photons that are reflected at the second BS pass through a quarter-wave plate (QWP) before a PBS directs them to a third pair of detectors. This branch implements the measurement of σ_y and accounts for the remaining terms in Eq. (2.4); namely, P_{y+} and P_{y-}.

B. Minimal four-outcome ellipsometer

There are a number of alternative schemes for an ellipsometer that realizes the minimal four-state POVM of Eq. (2.9). To demonstrate the case, we present one scheme here and discuss a few other schemes, which are much simpler and much more practical, elsewhere [6,7].

The principle of one minimal ellipsometer is illustrated by Fig. 3(a). It is an asymmetric four-path interferometer. At the input, one-half of the light intensity is directed into the uppermost path for reference, whereas each of the other three paths gets one-sixth of the intensity. Photons in the lower
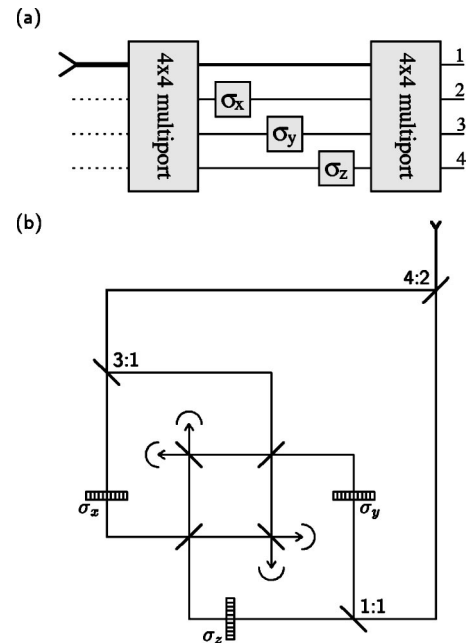

FIG. 3. (a) Principle of a four-path interferometer for minimal ellipsometry and (b) optical network for the implementation; see text.
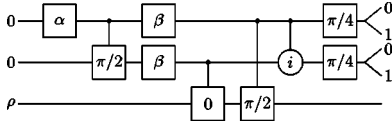
FIG. 4. Quantum computation network for minimal qubit to-mography. The bottom qubit is the one of interest; it enters in the state ρ whose properties are to be determined. The two top qubits are auxiliary qubits that enter in state 0 of the computational basis and are eventually measured. The probabilities for the four different measurement results—00, 01, 10, and 11—are the p_j's of Eq. (2.10) with the $\vec{a}_j$'s of Eq. (2.6). The network uses eight gates of the generalized Hadamard type (3.1), three of them controlled by one of the auxiliary qubits, and a controlled phase gate.

paths pass through wave plates that realize the unitary polarization transformations $\rho \rightarrow \sigma_\xi \rho \sigma_\xi$ with $\xi = x$, or y, or z, respectively. Then all four paths are recombined by a balanced beam merger, which distributes the intensity of each input evenly among the four outputs. An unpolarized photon has, therefore, a probability of 25% for being detected by a particular one of the four detectors. The probabilities for a polarized photon are the p_j's of Eq. (2.10), so that the polarization POVM of Eq. (2.9) is measured indeed.

An optical network for a four-path interferometer of this kind is shown in Fig. 3(b). The BS at the input reflects 4/6 of the intensity and transmits 2/6, and the subsequent BSs either split the beam 3:1 or 1:1; together, these three BSs implement the initial stage of Fig. 3(a). At the central stage, there are wave plates in three partial beams (labeled by σ_x, σ_y, and σ_z, respectively). And the final beam merger consists of four 1:1 BSs which direct the photons to the four photodetectors [8].

The interferometer of Fig. 3(b) has several loops. Some alternative setups have two loops [9] or need a single loop only, among them the interferometer of the experiment by Clarke *et al.* [10] and the minimal one-loop scheme of Ref. [6], and yet another setup has no loop at all [7]. We note that Clarke *et al.* did not perform ellipsometry; their experiment served a different purpose, and although their setup could be used for ellipsometry, there is no mentioning of this possible application in Ref. [10].

It is worth mentioning that such a setup can also be viewed as a quantum computation network for three qubits, one being the polarization qubit of interest, the other two qubits representing the four paths of the interferometer [11]. The network is depicted in Fig. 4. It consists of a sequence of generalized Hadamard gates,

$$-\boxed{\phi}- \; : \; \begin{pmatrix} \langle 0| \\ \langle 1| \end{pmatrix} \rightarrow \begin{pmatrix} \cos \phi & \sin \phi \\ \sin \phi & -\cos \phi \end{pmatrix} \begin{pmatrix} \langle 0| \\ \langle 1| \end{pmatrix}. \quad (3.1)$$

At the first stage, the auxiliary qubits are prepared in a superposition state that has amplitude $1/\sqrt{2}$ for $|00\rangle \triangleq |\text{vv}\rangle$ and amplitudes $1/\sqrt{6}$ for $|01\rangle, |10\rangle$, and $|11\rangle \triangleq |\text{hh}\rangle$. We achieve this by a controlled gate with $\phi = \frac{1}{2}\pi$ (this is a controlled-not gate, in fact) that is sandwiched by a gate with $\phi = \alpha$ and two gates with $\phi = \beta$, where α and β are such that $\sin(2\alpha) = \frac{1}{3}(\sqrt{3} - 1)$ and $\tan(2\beta) = \sqrt{3} + 1$.

The central stage has controlled gates acting on the qubit of interest, a gate with $\phi = 0$ for σ_z, another one with $\phi = \frac{1}{2}\pi$ for σ_x. For their product to realize $\sigma_y = i\sigma_x\sigma_z$, we provide the factor of i by a subsequent controlled phase gate, which implements the phase change $|11\rangle \rightarrow i|11\rangle$, but has no effect on $|00\rangle$, $|01\rangle$, and $|10\rangle$.

At the final stage, the two auxiliary qubits are passed through standard Hadamard gates, for which $\phi = \frac{1}{4}\pi$ in Eq. (3.1), and are then measured in the computational basis. The probabilities for getting 00, 01, 10, or 11 are p_1, p_4, p_2, and p_3 of Eqs. (2.10) and (2.6), respectively.

Conditioned on the measurement outcome, the qubit of interest emerges in the corresponding reduced state $2P_j\rho P_j/p_j$. For the optical implementation of Fig. 3(b), these final states of the polarization qubit are fictitious, however, unless the photodetectors are of a fantastic nondemolition kind: sensitive to the passage of a photon without absorbing the photon or affecting its polarization [12].

IV. COUNTING QUBITS

A. Maximum-likelihood estimator

In an actual experiment, we do not measure the probabilities of Eqs. (2.3) or (2.10), but rather relative frequencies that are statistically determined by these probabilities. The available information consists of the counts of detector clicks, n_1, n_2, n_3, n_4 for the minimal four-state tomography of Sec. II B and $n_{x\pm}$, $n_{y\pm}$, $n_{z\pm}$ for the standard six-state scheme of Sec. II A. In what follows, we focus on the novel four-detector situation.

In view of the intrinsic probabilistic nature of quantum phenomena, a given total number of N qubits does not result in a definite, predictable number of clicks for each detector. It is, therefore, clear that an observed breakup

$$N = n_1 + n_2 + n_3 + n_4 \quad (4.1)$$

is consistent with not just one Pauli vector $\vec{s}$ in Eq. (2.2), but with many. One expects that, as a rule, the obvious guess that obtains from Eq. (2.11) upon the replacement $p_j \rightarrow n_j/N$, namely,

$$\vec{S} = 3\sum_j \nu_j \vec{a}_j \quad \text{with} \quad \nu_j \equiv \frac{n_j}{N} \quad (4.2)$$

gives a reliable estimate of the true Pauli vector $\vec{s}$. But it may happen that the length of this inferred $\vec{S}$ exceeds unity, and this is in fact a typical situation if the true $\vec{s}$ is of unit length; that is, the source emits qubits in a pure state.

Let us thus proceed to show how one infers a plausible and physically correct answer on the basis of the registered data. The experimental data consist in fact not just of the total counts n_j in the breakup (4.1), but of a particular sequence of detector clicks. *Provided that* the source emits qubits in the state specified by the Pauli vector $\vec{s}$, with no statistical correlations between different qubits, the probability $\mathcal{L}(\vec{s}; n_1, \ldots, n_4)$ for getting the observed click sequence is, therefore, given by
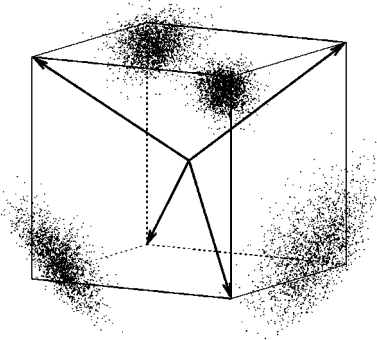
FIG. 5. The likelihood function as a probability density for the Pauli vector $\vec{s}$. The plot shows the "likelihood clouds" for simulated measurements for which the most likely Pauli vector has length $s = 0.84$ and is pointing toward one of the corners of the cube of Fig. 1. A total of 100 qubits have been detected for the bottom right corner, 200 for the bottom left corner, 400 for the top rear corner, and 800 for the top front corner. The successive shrinking of the cloud is clearly visible.

$$\mathcal{L} = \prod_{j=1}^{4} p_j(\vec{s})^{n_j}, \tag{4.3}$$

where $p_j(\vec{s})$ is the probability (2.10) that a qubit is registered by the jth detector.

Conversely, in the spirit of the Bayesian principle of statistical inference, we can regard $\mathcal{L}(\vec{s}; n_1, \ldots, n_4)$ as the likelihood that the source is characterized by $\vec{s}$, *given that* a click sequence with total detector counts $n_1, \ldots, n_4$ was observed. When many qubits have been counted, the likelihood function is sharply peaked and essentially vanishes outside the immediate vicinity of its maximum. These matters are illustrated by the four "likelihood clouds" in Fig. 5.

The maximum-likelihood (ML) estimator $\vec{S}$ picks out the most likely Pauli vector, the one for which $\mathcal{L}$ is largest:

$$\max_{\vec{s}} \mathcal{L}(\vec{s}) = \mathcal{L}(\vec{S}). \tag{4.4}$$

For the purposes of this paper, we accept this $\vec{S}$ as our plausible guess for $\vec{s}$, while being fully aware of other strategies [13–15].

An important theorem by Fisher states that ML estimators become efficient in the large-N limit [16,17]. So the performance, for large N, of an experimental tomographic setup can be quantified by the accuracy of the ML estimator. Usually an analytical expression for the ML estimator is not available and one has to solve a nonlinear operator equation for $\rho = \frac{1}{2}(1 + \vec{S} \cdot \vec{\sigma})$ to find it [18]. In the present context, however, the high symmetry of the vector quartet of Eq. (2.5) makes it possible to simplify this problem considerably.

There is a benefit in maximizing the right-hand side of Eq. (4.2) with respect to the probabilities p_j rather than with respect to $\vec{s}$. Lagrange multipliers are used to account for the two constraints. One is the unit sum of the p_j, $\Sigma_j p_j = 1$, which is the unit trace of ρ; the other is the positivity of ρ, $\rho \geq 0$, which is the upper bound in Eq. (2.13).

Without this positivity constraint, the likelihood (4.3) would be maximized by $p_j = \nu_j$, which in turn would imply the Pauli vector of Eq. (4.2). But if the actual counts violate the inequality

$$\sum_j \nu_j^2 \leq \frac{1}{3}, \tag{4.5}$$

this simple estimation fails to provide a physically meaningful result. When this happens, both constraints must be taken into account and the likelihood $\mathcal{L}(\vec{s})$ attains its maximum on the boundary of the set of qubit states—that is, for a Pauli vector of unit length [19].

The variation of the likelihood vanishes at the extremal point:

$$\delta \ln \mathcal{L} = \sum_j \frac{n_j}{p_j} \delta p_j = 0. \tag{4.6}$$

The variations δp_j are subject to the two constraints

$$\sum_j \delta p_j = 0, \quad \sum_j p_j \delta p_j = 0, \tag{4.7}$$

for which we use the Lagrange multipliers $N\lambda$ and $3N\mu$, respectively. Denoting the p_j values at the extremal point by $\tilde{p}_j$, Eq. (4.6) then implies

$$\frac{\nu_j}{\tilde{p}_j} = \lambda + 3\mu\tilde{p}_j \quad \text{for } j = 1, \ldots, 4. \tag{4.8}$$

We exploit $\Sigma_j \nu_j = 1$, $\Sigma_j \tilde{p}_j = 1$, and $\Sigma_j \tilde{p}_j^2 = \frac{1}{3}$ to establish

$$\sum_j \frac{\nu_j}{\tilde{p}_j} = 4\lambda + 3\mu \tag{4.9}$$

and

$$1 = \lambda + \mu, \tag{4.10}$$

the first by summing the four equations in (4.8), the second by summing them after multiplication with $\tilde{p}_j$.

Taken together, Eqs. (4.8)–(4.10) make up a set of six equations for the six unknowns $\tilde{p}_1, \ldots, \tilde{p}_4$, λ, and μ. We solve the quadratic equations (4.8) for $\tilde{p}_j$, or rather $\nu_j / \tilde{p}_j$,

$$\frac{\nu_j}{\tilde{p}_j} = \frac{1}{2}(\lambda + \sqrt{\lambda^2 + 12\mu\nu_j}), \tag{4.11}$$

and Eq. (4.10) for λ, and substitute these into Eq. (4.9) to arrive at a single equation for μ:

$$\mu + 2 - \frac{1}{2}\sum_j \sqrt{(1-\mu)^2 + 12\mu\nu_j} = 0. \tag{4.12}$$

There is always the solution $\mu = 0$, and thus $\lambda = 1$, but Eq. (4.11) amounts to Eq. (4.2) for these values, and therefore this solution is not acceptable, when the inequality (4.5) is not obeyed, as is the case in the present discussion. Upon recognizing that Eq. (4.12) can also be written as

$$6\mu^2\sum_j \frac{(3\nu_j-1)\nu_j}{1-\mu+6\mu\nu_j+\sqrt{(1-\mu)^2+12\mu\nu_j}}=0, \quad (4.13)$$

we can discard the unphysical solution $\mu=0$ and find the relevant solution in the range $0<\mu\leq 2$ as the root of this sum over j. The upper bound $\mu=2$ is reached when all qubits are detected by the same detector, so that one of the ν_j equals 1 and the others vanish.

In summary, we determine the ML estimator

$$\vec{S}=3\sum_j \tilde{p}_j\vec{a}_j \quad (4.14)$$

as follows. If the inequality (4.5) is obeyed by $\nu_j=n_j/N$, we take $\tilde{p}_j=\nu_j$. Otherwise, we find μ as the positive root of Eqs. (4.12) and (4.13) and then get the four $\tilde{p}_j$'s from Eq. (4.11) with $\lambda=1-\mu$. In the extremal situation of $\nu_j=\delta_{jk}$, we have $\tilde{p}_j=\frac{1}{6}+\frac{1}{3}\delta_{jk}$ and $\vec{S}=\vec{a}_k$.

B. Many detector clicks

In this procedure for finding the ML estimator, there is a crucial difference between relative frequencies ν_j that obey the inequality (4.5), and can therefore serve as probabilities, and those that violate the inequality. When the total number N of detected qubits is small, statistical fluctuations are relatively large and a violation is hardly surprising. But what is the typical situation for a large number of detector clicks; should we expect Eq. (4.5) to be obeyed or violated?

For the likelihood (4.3), only the breakup (4.1) matters, not the particular sequence of detector clicks. There are $N!/(n_1!n_2!n_3!n_4!)$ different sequences for a given breakup, so that there is a multinomial statistics for the probability of getting a particular breakup for the given probabilities $p_j(\vec{s})$ of Eq. (2.10). We denote the corresponding averages over possible breakups by overbars, as illustrated by

$$\overline{n_j}=N\overline{\nu_j}=Np_j \quad (4.15)$$

and

$$\overline{\nu_j\nu_k}=p_jp_k+\frac{1}{N}(\delta_{jk}p_k-p_jp_k). \quad (4.16)$$

Upon recalling that $\sum_j p_j^2=(3+s^2)/12$ [cf. Eq. (2.13)], the latter averages imply

$$\sum_j \overline{\nu_j^2}=\frac{1}{N}+\frac{N-1}{N}\sum_j p_j^2=\frac{1}{3}-\frac{1-s^2}{12}+\frac{9-s^2}{12N}, \quad (4.17)$$

so that, *on average*, the inequality (4.5) is violated for pure states (for which $s=1$) and is obeyed for mixed states (for which $s<1$). We thus expect that the detector counts for a pure qubit state will typically violate inequality (4.5), whereas the counts for a mixed state will tend to obey it.

As a more precise statement about this matter, we note that the fraction of detector click sequences that violate the inequality is

$$\text{prob(violation)}=\frac{1}{2}+\frac{1}{2}\text{erf}\left(\frac{\sqrt{N}}{2\kappa}\left(\sum_j \overline{\nu_j^2}-\frac{1}{3}\right)\right), \quad (4.18)$$

where erf() is the standard error function and κ is given by

$$\kappa^2=\sum_{j,k} p_jp_k(p_j-p_k)^2=2\sum_j p_j^3-2\left(\sum_j p_j^2\right)^2. \quad (4.19)$$

Equation (4.18) applies for $N\gg 1$; in order to derive it, first observe that the central limit theorem ensures that

$$\overline{\exp\left(i\sum_j \alpha_j(\nu_j-p_j)\right)}=\exp\left(-\frac{1}{2N}\sum_{j,k}\alpha_j(\delta_{jk}p_k-p_jp_k)\alpha_k\right) \quad (4.20)$$

for large N [20]. Then use this generating function for the mean values of products of the ν_j's to calculate the probability that $\sum_j \nu_j^2>\frac{1}{3}$, with consistent approximations for $N\gg 1$.

For pure states, the argument of the error function in Eq. (4.18) is $1/(3\kappa\sqrt{N})$, which is always positive, but decreases with growing N, so that click sequences that violate the inequality are more frequent than the ones obeying it, but they are not much more frequent. A remarkable exception is the situation of the Pauli vector $\vec{s}$ being exactly opposite to one of the directions of the vector quartet of Eqs. (2.5) and (2.6). Then one of the p_j's vanishes and the other three are all equal to $\frac{1}{3}$, so that $\kappa=0$ and there is a unit probability for getting a violation.

For mixed states, the argument of the error function is negative and increases in magnitude $\propto\sqrt{N}$ (more about this in Sec. V C). Accordingly, a violation of the inequality is highly improbable. In the extreme situation of the completely mixed state, $\rho=\frac{1}{2}$, all p_j's are $\frac{1}{4}$, so that $\kappa=0$ here too, and the probability for violating the inequality vanishes.

C. Asymptotic efficiency

After detecting many qubits, we expect the ML estimator $\vec{S}$ to be quite close to the true Pauli vector $\vec{s}$. If inequality (4.5) is obeyed, the average squared distance

$$\overline{(\Delta \vec{s})^2}=\overline{(\vec{S}-\vec{s})^2}=\overline{\left(3\sum_j (\nu_j-p_j)\vec{a}_j\right)^2}=\frac{12}{N}\left(1-\sum_j p_j^2\right)$$

$$=\frac{9-s^2}{N} \quad (4.21)$$

is immediately available as a consequence of Eq. (4.16).

The analysis is more involved when Eq. (4.5) is not obeyed. Let us consider once more the extreme situation of $\vec{s}$ being exactly opposite to one of the $\vec{a}_j$, such as $\vec{s}=-\vec{a}_4$, say, so that $p_4=0$ and $\nu_4=0$, and ν_1, ν_2, ν_3 differ little from $p_1=p_2=p_3=\frac{1}{3}$. To leading order in $\nu_j-\frac{1}{3}$, the solution of Eq. (4.12) is then

$$\mu=1+\frac{9}{8}\sum_{j=1}^3 \left(\nu_j-\frac{1}{3}\right)^2 \quad (4.22)$$
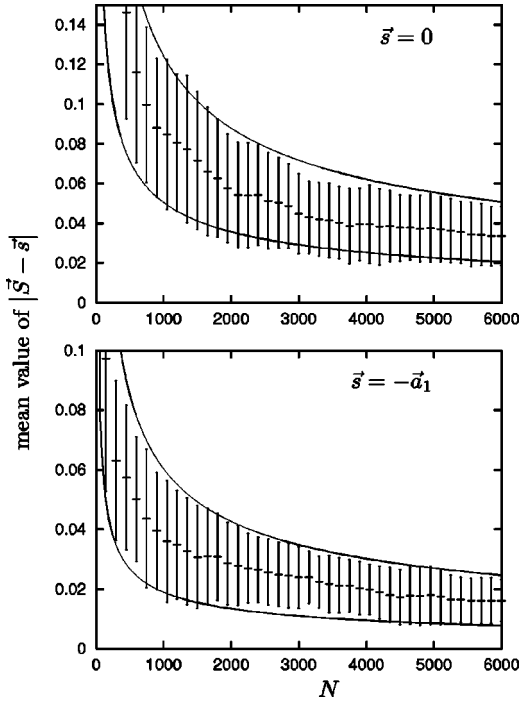
and the resulting ML estimator is

FIG. 6. Mean value of the distance from the ML estimator $\vec{S}$ to the true Pauli vector $\vec{s}$ for 40 simulated experiments with up to 6000 detected qubits per run. The top plot is for $\vec{s}=0$, the bottom plot for $\vec{s}=-\vec{a}_4$. The solid lines indicate one standard deviation to each side of the mean value, as obtained from the large-N approximations for $(\Delta\vec{s})^2$ in Eqs. (4.21) and (4.24), respectively.

$$\vec{S} = \frac{1}{2}\sum_{j=1}^{3}(1+3\nu_j)\vec{a}_j = \vec{s} + \frac{3}{2}\sum_{j=1}^{3}\left(\nu_j - \frac{1}{3}\right)\vec{a}_j. \quad (4.23)$$

In conjunction with Eq. (4.16), we thus get the large-N approximation

$$(\Delta\vec{s})^2 = 3\overline{\nu_1^2 + \nu_2^2 + \nu_3^2} - 1 = \frac{2}{N} \quad (4.24)$$

in this case of perfect antialignment between $\vec{s}$ and the quartet of $\vec{a}_j$'s. The sensitivity to small misalignments is discussed in Sec. VII C.

As demonstrated by the numerical results of Fig. 6, these asymptotic approximations are actually quite reliable for $N \gtrsim 1000$. The plots report the mean values, with statistical error bars of one standard deviation, of the distance $|\vec{S}-\vec{s}|$ as obtained from 40 simulated experiments, whereby up to 6000 qubits are detected in each run. The top plot refers to a true state that is completely mixed, so that $s=0$ in Eq. (4.21). In the bottom plot we have $\vec{s}=-\vec{a}_4$ and Eq. (4.24) applies.

By comparison, for the standard six-output ellipsometer with counts $n_{\xi\pm}$ of the respective detectors, the ML estimator is

$$\vec{S} = \left(\frac{n_{x+}-n_{x-}}{n_{x+}+n_{x-}}, \frac{n_{y+}-n_{y-}}{n_{y+}+n_{y-}}, \frac{n_{z+}-n_{z-}}{n_{z+}+n_{z-}}\right), \quad (4.25)$$

provided that the length of this $\vec{S}$ does not exceed unity. If it does, the search has to be constrained to unit vectors, much

like in the discussion of Eqs. (4.6)–(4.14). When Eq. (4.25) applies, the error corresponding to Eq. (4.21) is given by

$$(\Delta\vec{s})^2 = \frac{9-3s^2}{N}. \quad (4.26)$$

For $s>0$, this is a bit smaller than the value in Eq. (4.21), so that the six-output scheme gives a slightly more reliable estimate. But when s gets close to unity, the four-output scheme can provide better estimates because there are states for which the error is substantially smaller, as the extreme situation of Eq. (4.24) shows. For the six-output ellipsometer, the privileged pure states ρ project on an eigenstate of σ_x, or σ_y, or σ_z, so that one of the six detectors in Fig. 2 never registers a qubit. Then we have

$$(\Delta\vec{s})^2 = \frac{8}{3N}, \quad (4.27)$$

which is a bit larger than its analog (4.24). The main lesson is thus that the two schemes are of comparable asymptotic efficiency.

V. OPTIMALITY OF MINIMAL FOUR-STATE TOMOGRAPHY

A. Cramér-Rao bound

As a preparation of the discussion below on the optimality and efficiency of the minimal four-state tomography, we recall some well-known facts about state estimation. We regard the qubit state as parametrized by the three components of its Pauli vector $\vec{s}=(s_x,s_y,s_z)$, rather than by the four probabilities p_j, because the latter are constrained. Then, the error of the components of the ML estimator $\vec{S}$ can be estimated by the Cramér-Rao lower bound [17,21]

$$(\Delta s_\xi)^2 \geq (I_F^{-1})_{\xi\xi} \quad \text{with } \xi=x,y,z, \quad (5.1)$$

where I_F is the Fisher information matrix,

$$(I_F)_{\xi\zeta} = \overline{\frac{\partial\log\mathcal{L}}{\partial s_\xi}\frac{\partial\log\mathcal{L}}{\partial s_\zeta}}, \quad (5.2)$$

and the averaging is done over the data—that is, over the multinomial distribution that we exploited already in Sec. IV B.

There is no unique numerical measure for the comparison of the estimated state $\rho_{est}=\frac{1}{2}(1+\vec{S}\cdot\vec{\sigma})$ with the true state $\rho_{true}=\frac{1}{2}(1+\vec{s}\cdot\vec{\sigma})$. One can make a case for the trace-class distance

$$\mathcal{D}_{tr} = \frac{1}{2}\text{tr}\{|\rho_{est}-\rho_{true}|\} = \frac{1}{2}|\vec{S}-\vec{s}|, \quad (5.3)$$

for the Hilbert-Schmidt distance

$$D_{HS} = \frac{1}{2}[\text{tr}\{(\rho_{est}-\rho_{true})^2\}]^{1/2} = \frac{1}{2}|\vec{S}-\vec{s}|, \quad (5.4)$$

for the Uhlmann fidelity

$$\mathcal{U} = \text{tr}\{|\sqrt{\rho_{\text{est}}} \sqrt{\rho_{\text{true}}}|\}$$

$$= \frac{1}{2}(1 + \vec{S} \cdot \vec{s} + u)^{1/2} + \frac{1}{2}(1 + \vec{S} \cdot \vec{s} - u)^{1/2}, \quad (5.5)$$

with

$$u = \sqrt{(\vec{S} + \vec{s})^2 - (\vec{S} \times \vec{s})^2}, \quad (5.6)$$

and for some more. In the single-qubit situation of present interest, there is no difference between $\mathcal{D}_{\text{tr}}$ and $\mathcal{D}_{\text{HS}}$, but they are not the same for higher-dimensional systems, such as the qubit pairs of Sec. II C. We note in passing that the Uhlmann fidelity and the trace-class distance are natural pairs, inasmuch as they obey a fundamental inequality

$$\mathcal{D}_{\text{tr}}^2 + \mathcal{U}^2 \leq 1, \quad (5.7)$$

irrespective of the dimension of the Hilbert space.

As long as one is comparing pure states with each other, the actual choice between the quantitative measures of Eqs. (5.3)–(5.5) does not matter much, because then all of them are monotonic functions of the length of the difference $\vec{S} - \vec{s}$ of the two Pauli vectors. If one state is mixed, however, this is not true for the Uhlmann fidelity. For the sake of computational simplicity, we opt for the (square of twice) the Hilbert-Schmidt distance and thus quantify the measure of the estimate by the size of $(\vec{S} - \vec{s})^2$, whose statistical average is already considered in Sec. IV C. A convenient conservative estimate is given by the Cramér-Rao bound of Eq. (5.1).

According to Fisher's theorem [16] the ML estimator of ρ attains the Cramér-Rao bound for large N. Strictly speaking, this statement applies only to mixed states, because only for them are the ML estimators unbiased. For pure states, the positivity constraint plays an important role, and the Cramér-Rao bound derived for unbiased estimators tends to overestimate the error, and therefore we shall consider pure states separately below. As a rule of thumb, the Cramér-Rao bound is reliable when the lion's share of the likelihood function is contained within the Bloch ball. This can, in fact, be used as an operational definition of "mixedness." Indeed, as the analysis in Sec. IV B shows, for mixed states there is no significant fraction of the likelihood function outside the Bloch ball if N is sufficiently large.

B. Optimality of the tetrahedron geometry

The four-state tomography with the tetrahedron geometry of Eqs. (2.5)–(2.9) is the best minimal qubit-state measurement, in the sense that, among the measurements with four output channels, it provides the greatest accuracy. This can be seen as follows.

For the multinomial statistics of Sec. IV B, the Fisher information simplifies to

$$(I_F)_{\xi\zeta} = N \sum_j \frac{1}{p_j} \frac{\partial p_j}{\partial s_\xi} \frac{\partial p_j}{\partial s_\zeta}, \quad (5.8)$$

where p_j is the probabilities of detecting a qubit in the jth output channel. Consider now this variational problem: For a given input state, the average distance D is to be minimized

over all possible four-element POVMs. The functional in question is

$$D = \text{Sp}(I_F^{-1}) - \lambda \sum_j \Pi_j, \quad (5.9)$$

where $\text{Sp}()$ is the trace ("spur") of the 3×3 matrix, and a Lagrange operator Λ takes care of the constraint $\Sigma_j \Pi_j = 1$.

First, let us find the POVM that minimizes the functional (5.9) for the maximally mixed input state. In that case the extremal equations read

$$R_j \Pi_j = \Lambda \Pi_j \quad \text{for } j = 1, \dots, 4,$$

$$\Lambda = \sum_j R_j \Pi_j, \quad (5.10)$$

where

$$R_j = \text{Sp}(I_F^{-2} T_j) \quad (5.11)$$

involves the operator matrix T_j whose matrix elements are given by

$$T_{j,\xi\zeta} = \text{tr}\{\Pi_j \sigma_\xi\} \sigma_\zeta + \sigma_\xi \text{tr}\{\sigma_\zeta \Pi_j\}. \quad (5.12)$$

Note the joint appearance of 3×3 $\text{Sp}()$ traces and quantum-mechanical $\text{tr}\{\}$ traces. One verifies by inspection that any POVM of the tetrahedron geometry of Eqs. (2.5)–(2.9) satisfies these extremal equations and hence identifies the maximally mixed state with greatest accuracy.

Obviously, for biased states the optimal POVM itself becomes biased. The generalization of the extremal equations to this case is straightforward; the operators T_j will then contain one more term proportional to the true state ρ. Although an analytical solution may be difficult to obtain, one can always find the extremal measurement by an iterative procedure. As the input state is usually not known and selected in random, operators R_j should be averaged over the Bloch ball. In this way one obtains an algorithm providing the optimal sequential measurement, in the sense that the Hilbert-Schmidt distance is consistently reduced in each iteration step. Numerical results show that the tetrahedron POVM is also optimal for uniformly distributed input states.

Quite explicitly, the optimal distance for the tetrahedron POVM is

$$D_{\text{opt}} = \frac{9 - s^2}{N}, \quad (5.13)$$

in agreement with, or as a consequence of, the mean square distance of Eq. (4.21). As one would expect, the accuracy of the ellipsometer is somewhat better for pure input states than for mixed states. But, most importantly, the accuracy does not depend on the orientation of $\vec{s}$ relative to the measurement tetrahedron.

C. Orientation of the measurement tetrahedron

Their relative orientation is, however, not completely irrelevant. For example, the value of κ^2 in Eq. (4.19) clearly depends on it. More generally, we can obtain the large-N mean values of functions of $\vec{S} - \vec{s}$ from the asymptotic generating function

$$\overline{\exp[i\vec{r}\cdot(\vec{S}-\vec{s})]} = \exp\left(-\frac{1}{2N}\vec{r}\cdot\overset{\leftrightarrow}{K}\cdot\vec{r}\right), \qquad (5.14)$$

where the dyadic

$$\overset{\leftrightarrow}{K} = 9\sum_{j,k}\vec{a}_j(\delta_{jk}p_j - p_jp_k)\vec{a}_k \qquad (5.15)$$

depends on the positioning of $\vec{s}$ relative to the vector quartet of the $\vec{a}_j$'s. This is, of course, an immediate consequence of Eq. (4.20).

As an application, let us consider the asymptotic mean value of the Uhlmann fidelity of Eqs. (5.5) and (5.6). We first note that

$$\mathcal{U} = 1 - \frac{1}{8}\left[(\vec{S}-\vec{s})^2 + \frac{[\vec{s}\cdot(\vec{S}-\vec{s})]^2}{1-s^2}\right] \qquad (5.16)$$

holds when $\vec{S}-\vec{s}$ is small. Then we recall Eq. (4.21) and extract

$$\overline{[\vec{s}\cdot(\vec{S}-\vec{s})]^2} = \frac{1}{N}\vec{s}\cdot K\cdot\vec{s} = \frac{72\kappa^2}{N} \qquad (5.17)$$

from Eq. (4.25) to arrive at

$$\bar{\mathcal{U}} = 1 - \frac{9-s^2}{8N} - \frac{9\kappa^2}{N(1-s^2)}, \qquad (5.18)$$

where we meet the orientation-dependent quantity κ^2 of Eq. (4.19).

There are extremal orientations of three kinds. The value of κ^2 is largest when $\vec{s}$ is parallel to one of the vectors of the tetrahedron quartet and smallest when $\vec{s}$ is antiparallel. In addition to these four maxima and four minima of κ^2, there are also six saddle points that have $\vec{s}$ parallel to the sum of two different vectors of the quartet. The maximal and minimal values of κ^2 are given by the upper and lower signs in

$$\kappa^2 = \frac{(1\pm s)(3\mp s)s^2}{72} \quad \text{for } \vec{s}=\pm s\vec{a}_j \quad \text{(any } j\text{)}, \qquad (5.19)$$

respectively, and the value at the saddle points is

$$\kappa^2 = \frac{(3-s^2)s^2}{72} \text{ for } \vec{s}=\frac{\sqrt{3}}{2}s(\vec{a}_j+\vec{a}_k) \text{ (any } j\neq k\text{)}. \qquad (5.20)$$

Accordingly, the approach of $\bar{\mathcal{U}}$ to unity is fastest for the $\vec{s} = -s\vec{a}_j$ orientation, for which the large-N approximation

$$\bar{\mathcal{U}} = 1 - \frac{(3+s)(3+2s)}{8N(1+s)} \qquad (5.21)$$

applies. A small value of κ is also advantageous in Eq. (4.18), as it ensures a large negative argument of the error function and thus a small probability for violating the inequality (4.5).

VI. MEASURING PURE QUBIT STATES

When the measured quantum system is known to be in a pure state—which is a bold overidealization of any realistic situation—this knowledge can be exploited systematically when estimating the otherwise unknown state. A somewhat more realistic situation arises when the input state is pure but we do not know this to begin with, although this case is quite a bit artificial as well, because one can hardly assume that real sources are not affected by classical noise or that the experimental setup is totally decoupled from the environment. Put differently, it is far-fetched to assume that the experimenter will ever have the perfect control that is necessary to ensure that a source emits a pure state. Nevertheless, there is an interest in such idealized scenarios because they come up in analyses of eavesdropping attacks on schemes for quantum cryptography, where—as a matter of principle—it is assumed that the eavesdropper is only limited by the laws of physics, not by practical limitations.

For $s=1$, the Uhlmann fidelity (5.5) simplifies to

$$\mathcal{U} = \sqrt{\frac{1+\vec{s}\cdot\vec{S}}{2}} \qquad (6.1)$$

and is equivalent to

$$\mathcal{U} = \sqrt{1-\frac{1}{4}(\vec{S}-\vec{s})^2} \qquad (6.2)$$

if the estimator is a pure state as well, $S=1$. When indeed estimating pure states with pure states, the situation to be considered now, it is customary to take the average of $\mathcal{U}^2$,

$$\bar{F} = \overline{\mathcal{U}^2} = 1 - \frac{1}{4}\overline{(\vec{s}-\vec{S})^2} = 1 - \frac{1}{4}(\Delta\vec{s})^2, \qquad (6.3)$$

as the fidelity measure that judges the quality of the estimation procedure. One must keep in mind that both the true Pauli vector $\vec{s}$ and its estimator $\vec{S}$ are unit vectors here, as Eq. (6.3) applies only under this restriction. Knowing that the source generates pure states means having a lot of prior knowledge because the set of pure states is much smaller than the set of mixed states (the Bloch sphere rather than the Bloch ball), and so one can safely expect that a better, possibly much better, accuracy of the estimation can be achieved.

The average fidelity of the optimal *joint* measurement on N copies is known to obey the inequality [22,23]

$$\bar{F} \leq \frac{N+1}{N+2}, \qquad (6.4)$$

so that the corresponding error $1-\bar{F}$ will decrease as $1/N$ in the large-N limit. Any reasonable estimator should show this dependence.

In our scheme the qubits are measured individually, not jointly. Nevertheless, as a consequence of the Fisher theorem in general and of the findings in Sec. IV C in particular, the variance $(\Delta\vec{s})^2$ in Eq. (6.3) is proportional to $1/N$ in the large-N limit, and so is then $1-\bar{F}$. But we need to reconsider the asymptotics of the maximum-likelihood estimation, now

taking into account that both $\vec{s}$ and $\vec{S}$ are restricted to the Bloch sphere, so that we get the estimator $\vec{S}$ from Eqs. (4.10)–(4.13), whether inequality (4.5) is obeyed or not.

Here, too, the estimation is sensitive to the orientation of the Pauli vector $\vec{s}$ relative to the vector quartet of the measurement tetrahedron. We deal first with the case of "not antialigned"; that is, $\vec{s} \neq -\vec{a}_j$ for all $\vec{a}_j$'s. Then, $\nu_j - p_j$ and μ are of the order of $1/\sqrt{N}$ and μ is given by

$$\mu = \frac{2}{3\kappa^2} \sum_j p_j(\nu_j - p_j), \tag{6.5}$$

with $\kappa^2 > 0$ from Eq. (4.19). The resulting averages

$$\overline{(\nu_j - p_j)\mu} = \frac{2}{9N\kappa^2}(3p_j - 1)p_j,$$

$$\overline{\mu^2} = \frac{2}{9N\kappa^2}, \tag{6.6}$$

available as a consequence of Eq. (4.16), are used in

$$\bar{F} = 1 - 3\sum_j \overline{[(\nu_j - p_j) - (3p_j - 1)p_j\mu]^2} \tag{6.7}$$

to arrive at

$$1 - \bar{F} = \frac{2}{N} - \frac{2}{3N\kappa^2} \sum_j [(3p_j - 1)p_j]^2$$

$$= \frac{4}{N} - \frac{2}{9N\kappa^2}\left(27\sum_j p_j^4 - 1\right). \tag{6.8}$$

In particular, when the tetrahedron is aligned with the Pauli vector, the upper-sign case of Eq. (5.19) with $s=1$, we have $\kappa^2 = 1/18$ and $27\sum_j p_j^4 = 7/4$, so that

$$1 - \overline{F_{\uparrow\uparrow}} = \frac{1}{N} \tag{6.9}$$

for this parallel strategy ($\uparrow\uparrow$). This is only half as big as what one would get for the $s=1$ value of Eq. (4.21) and thus demonstrates the advantage of estimating the pure true state by pure-state estimators only.

There is no such advantage for the antiparallel strategy ($\uparrow\downarrow$) that has the tetrahedron antialigned with $\vec{s}$, because the argument in Sec. IV B establishes that the ML estimator is always a pure state then. Since this is now the lower-sign case of Eq. (5.19) for $s=1$, we have $\kappa^2 = 0$ and Eqs. (6.5)–(6.8) are not valid. Instead, we recall that Eqs. (4.22)–(4.24) apply and conclude that

$$1 - \overline{F_{\uparrow\downarrow}} = \frac{1}{2N} \tag{6.10}$$

holds here. The antiparallel strategy has thus half the average error of the aligned strategy.

On the other hand, for a generic orientation of the input state there will always be an uncertainty in the length of the input state vector and therefore the fidelity will approach unity at a much slower rate proportional to $1/\sqrt{N}$. Only if one knows beforehand that the input state is pure can the

$1/N$ rate be achieved even for general orientations of the tetrahedron. Let us illustrate this last remark by the example of the parallel strategy. For the parallel orientation, only about half of the measurement outcomes will violate inequality (4.5); the others would not (see Sec. IV B). Knowing that the input state is pure one can, however, ignore this inequality and always solve Eqs. (4.8)–(4.10) for the input-state unit Pauli vector.

VII. ADAPTIVE STRATEGIES

A. Premeasurement strategy

The good performance of the antiparallel strategy hints at a very simple adaptive procedure that provides the fast $1/N$ asymptotic behavior without any prior knowledge about the purity of the input state: Let us split the input ensemble into two halves. After the first $N/2$ particles have been registered and the direction of the input Pauli vector estimated, the experimenter adopts the antiparallel strategy for measuring the rest of the ensemble. Notice that in this simple adaptive scenario, the first half of the particles are used for a premeasurement and serve only for adjusting the measurement apparatus, while it is the second half which provide the actual estimate of the input state.

Let θ denote the angle between the Pauli vector $\vec{s}$ of the input state and the Pauli vector $\vec{s}_1$ estimated from the first half of the ensemble. This angle θ can be estimated with an accuracy of $1 - \cos\theta \propto 1/N$ in the first stage. This means that, in the second stage, the mean probability of detecting a particle in the channel antiparallel to $\vec{s}_1$ will be proportional to $1/N$. No matter how large is N, only a few particles will be detected in this channel.

The maximal uncertainty in the length of the Pauli vector is then easily calculated with the aid of

$$S^2 = \frac{12}{N^2} \sum_j n_j^2 - 3 \tag{7.1}$$

for the estimator of Eq. (4.2). Let us set $n_1 = \delta$ with δ a small number independent of N and look for the minimal length of the estimated Pauli vector compatible with the given δ. Since the right-hand side of Eq. (7.1) is a concave function of n_2, n_3, and n_4, it is minimal when all of them are equal to each other, $n_2 = n_3 = n_4 = (N - \delta)/3$, with the consequence

$$S_{min} = 1 - \frac{4\delta}{N} \tag{7.2}$$

for large N. This guarantees that in the second stage both the orientation and length of the Pauli vector, and so the fidelity, will be determined with an error proportional to $1/N$. Numerical simulations show that for large N (up to $N = 10^5$) the mean error $1 - \bar{F}$ of this simple protocol is about twice as large as, and thus worse than, that of the optimal joint POVM measurement, in which all N qubits are measured together. Let us emphasize that while our protocol would provide this performance for any input pure state, it also provides a meaningful estimate for any input mixed state, of course, with a larger uncertainty proportional to $1/\sqrt{N}$. In

contrast to that, the optimal joint measurement that attains the ultimate limit of Eq. (6.4) would not work for mixed input states.

B. Self-learning strategies

It is known that sequential measurements when combined with self-learning adaptive strategies can come close to the quantum estimation limit [24–26]. Their improvement on the conventional sequential measurement depends on the purity of the input state [24]. Adaptive techniques are more sensitive to pure states than to states with a lot of classical noise.

The optical network of Fig. 3(b) can easily be adapted to a self-learning procedure and so can other optical implementations. After each detection the current information about the input state can be evaluated, and the operations σ_x, σ_y, and σ_z acting on the next particle inside the interferometer can be modified by a common unitary transformation. This is economically achieved by performing the required unitary transformation on the approaching photon before it enters the interferometer proper.

From the discussion in Sec. VI one might get the impression that a particularly good adaptive strategy would be to always keep one of the measured half-projectors antiparallel to the current estimate. Matters are, however, not so simple.

Let us illustrate the difference between the adaptive and nonadaptive procedures at the extreme example of measuring only two qubits. In the *nonadaptive* case, everything is as discussed above. In particular, the probability that the first qubit is detected by the jth detector and the second by the kth detector is $p_j p_k$ with the p_j's of Eq. (2.10) and, in accordance with Sec. IV, the ML estimator is $\vec{S} = \vec{a}_j$ if $j = k$ and $\vec{S} = \sqrt{3/4}(\vec{a}_j + \vec{a}_k)$ if $j \neq k$. Upon averaging $(\vec{S} - \vec{s})^2$ first over all measurement results for a given input Pauli vector $\vec{s}$ and then over all possible inputs, we thus get

$$\overline{(\vec{S} - \vec{s})^2} = \frac{5 - \sqrt{3}}{3} = 1.089 \qquad (7.3)$$

as the figure of merit.

In the *adaptive case*, the tetrahedron is realigned for the second qubit after the first qubit has been detected. Since the ML estimator $\vec{S}$ obtained after the detection of the first qubit will coincide with one of the tetrahedron vectors, the antialigning of the tetrahedron for the second qubit amounts to the replacement $\vec{a}_j \rightarrow -\vec{a}_j$. Now the probability that the first qubit is detected by the jth detector and the second by the kth detector is $\mathcal{L} = p_j(\frac{1}{2} - p_k)$. The resulting ML estimator is then given by $\vec{S} = \sqrt{3/8}(\vec{a}_j - \vec{a}_k)$, and upon averaging over all measurement results and all input Pauli vectors we get

$$\overline{(\vec{S} - \vec{s})^2} = \frac{11 - \sqrt{24}}{6} = 1.017, \qquad (7.4)$$

which is markedly smaller than the nonadaptive value in Eq. (7.3). It is also smaller than the value for the adaptive strategy with parallel alignment because that is identical with the nonadaptive procedure when only two qubits are detected.

We note that the averages in Eqs. (7.3) and (7.4) are taken over pure input states. If one averages over all input states,
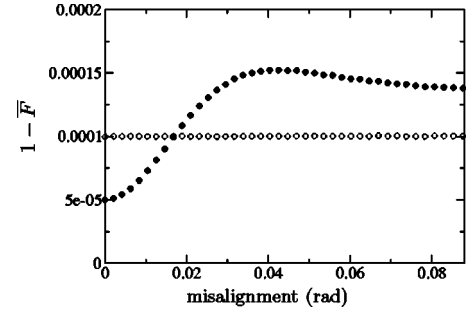

FIG. 7. Mean estimation errors of the POVM tetrahedrons that, on the Bloch sphere, differ from the exact parallel (squares) and antiparallel (circles) orientations by the known angle given on the abscissa.

pure and mixed, one obtains the respective numbers $(7 - \sqrt{3})/5 = 1.054$ and $(7 - \sqrt{6})/5 = 0.910$, again with a clear advantage for the antialigning adaptive strategy.

C. Numerical simulations

Although these numbers speak clearly in favor of the antialigning adaptive scheme, one should, however, keep in mind that they apply only for the exactly aligned or antialigned settings of the apparatus. But in such an adaptive scheme, after the first particle is observed, the uncertainty of the input state Pauli vector orientation is still quite large, which may result in a significant misalignment in the second adaptive step. In fact, the antialigning strategy has a greater sensitivity to such misalignments. This is illustrated by the simulation data shown in Fig. 7, where the estimation errors of both strategies are shown in dependence on the misalignments of the apparatuses for the chosen input intensity of 10^4 particles. Each point has been obtained by averaging over 5×10^5 ML estimates.

As expected, the antialigning adaptive strategy performs better if no misalignment is present. However, even a small misalignment (of the order of 1° in this case) is enough to wash out this advantage. For even larger misalignments the aligned setting provides much better performance [27]. Different sensitivities of both measurement strategies to this kind of error might be of quite some importance for the potential applications of the minimal ellipsometer in quantum communication protocols and quantum cryptography.

One can understand this extreme sensitivity of the antialigned setting and why it becomes immediately worse than the aligned setting by taking a second look at Eqs. (6.8) and (6.10). The $\kappa^2 = 0$ result (6.10) is *not* the $\kappa^2 \rightarrow 0$ limit of the $\kappa^2 > 0$ result (6.8). In fact we have (see Fig. 8)

$$1 - \bar{F} \rightarrow \frac{4}{3N} \quad \text{as } \kappa^2 \rightarrow 0 \quad \text{in Eq. (6.8),} \qquad (7.5)$$

which is larger than the error of $\overline{F_{\uparrow\uparrow}}$ in Eq. (6.9). Therefore, the slightest misalignment takes us from the $1/(2N)$ error of Eq. (6.10) to this $4/(3N)$ because Eq. (7.5) applies to tiny nonzero values of κ^2, while Eq. (6.10) holds only if $\kappa^2 = 0$ exactly.

This observation also resolves the apparent contradiction between the general upper bound of Eq. (6.4) and the large-
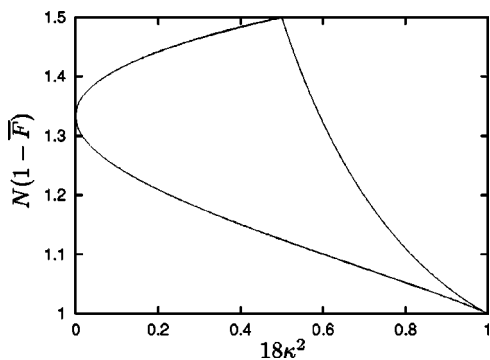
FIG. 8. For each value of κ^2 between $\kappa^2=0$ (perfect antialignment) and $\kappa^2=1/18$ (perfect alignment), the possible values of the coefficient of the $1/N$ term in Eq. (6.8) are in the area bounded by the two curves. The smallest coefficient obtains for the case of perfect alignment, when Eq. (6.9) applies.

N error for perfect antialignment in Eq. (6.10), which does not respect that upper bound. Nevertheless, this example is not a valid counterexample, because it refers to an absurdly artificial situation: The experimenter has perfect *a priori* knowledge of the state to be measured and has perfect control over his measurement apparatus, such as to ensure the perfect antialignment to which Eq. (6.10) applies. In other words, *when* it applies, there is no need for a state estimation to begin with.

Having thus compared the performances of the two extreme strategies, we now calculate the mean fidelity of the following adaptive measurement: After detecting each new qubit the information about the input state is updated and a new ML estimate is calculated. Then one of the measured half-projectors is aligned along this current estimate. These two steps are repeated until all input particles are used up. Figure 9 shows mean fidelities and errors that were obtained by averaging over 200 000 randomly selected pure input states. The quantum limit, Eq. (6.4), is shown for comparison. It is evident that this bound can be attained only for large N, while the most pronounced difference is seen for moderately sized ensembles. Such a behavior is typical for all sequential self-learning estimation strategies.

Finally, let us compare the efficiency of the parallel adaptive strategy with a very simple sequential measurement
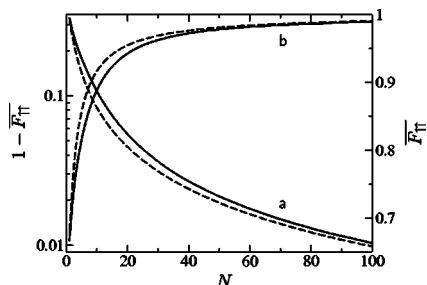

FIG. 9. Mean error (curves a, decreasing) and mean fidelity (curves b, increasing) of the minimal qubit tomography as a function of N, the size of the measured ensemble. Solid lines: the proposed network of Figs. 3 and 4 with the adaptive parallel strategy described in the text. Dashed lines: the quantum limit of Eq. (6.4). Notice that the proposed optical network attains the quantum limit asymptotically for large N.
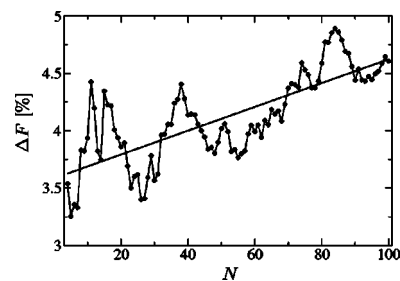

FIG. 10. Relative difference ΔF, in percent, between the errors of the random and parallel adaptive strategies (black dots) and least-squares linear fit (straight line). The lines connecting the dots guide the eye but have no further significance. The irregularities stem from the limited number of input states (200 000) used for the averaging.

where the orientation of the measured half-projectors is chosen at random in each step (see Fig. 10). As expected, the adaptive strategy is better and its benefit grows with increasing size of the measured ensemble.

VIII. SUMMARY

We have presented a minimal measurement scheme for single-qubit tomography that has no more than the necessary number of four outputs. The scheme is conceptually simple, highly symmetric and optimal among all four-output schemes, and can be realized with the present technology for the polarization qubit of photons emitted by a single-photon source. As a demonstration, we designed a simple, but not simplest, optical network.

Our thorough analysis showed that the scheme is efficient in the sense that it enables one to estimate the qubit state reliably without first detecting an enormous number of qubits—a few thousand are sufficient for most practical applications; a few hundred may be enough if extreme precision is not required. The efficiency can be increased by adaptive procedures in which the apparatus is adjusted in accordance with the current estimate for the qubit state.

Since the four-output setup provides optimal complete tomography with the minimal number of output channels, it is particularly well suited as a detection device for certain quantum communication protocols such as tomographic quantum cryptography [28]. Indeed, there are protocols for quantum key distribution that exploit the tetrahedron quartet of states [29], among them a highly efficient tomographic protocol [30].

ACKNOWLEDGMENTS

[1] W. K. Wootters and B. D. Fields Ann. Phys. (N.Y.) **191**, 363 (1989).

[2] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, J. Math. Phys. **45**, 2171 (2004).

[3] Incidentally, these probabilities are closely related to Feynman's "negative probabilities" [R. P. Feynman, in *Quantum Implications: Essays in Honour of David Bohm*, edited by B. Hiley and D. Peat (Routledge London, 1987)], which are the expectation values of $\frac{1}{2}(1+\sqrt{3}\vec{a}_j\cdot\vec{s})$. In fact, they are not probabilities, but rather the natural, discrete qubit analog of Wigner's phase-space function for continuous degrees of freedom; see W. K. Wootters, Ann. Phys. (N.Y.) **176**, 1 (1987) and D. Galetti, and A. F. R. De Toledo Piza, Physica A **149**, 267 (1988), for further details about these matters in general and, e.g., A. Luis, and J. Peřina, J. Phys. A **31**, 1423 (1998), for some particular aspects.

[4] In *state tomography* one must know the relevant properties of the measurement device; in *measurement tomography* one must know the state that is measured.

[5] We thank Ch. Kurtsiefer, for asking whether this can be done.

[6] B.-G. Englert, K. M. Tin, C. G. Goh, and H. K. Ng, Laser Phys. (to be published), e-print physics/0409015.

[7] B.-G. Englert, C. G. Goh, Ch. Kurtsiefer, A. Lamas Linares, H. K. Ng, and K. M. Tin (unpublished).

[8] It may be worth mentioning that, in principle if not in practice, any single-photon POVM can be implemented with an optical multiport in a rather transparent fashion. The setup of Fig. 3(b) is a four-port: the photodetectors register the photons exiting at the four output ports; one of the two entry ports at the 4:2 beam splitter is used, whereas no use is made of the two other entry ports at the 3:1 and 1:1 beam splitters. The setup of Ref. [9] is also a four-port, whereas the setups in Refs. [6,7] are two ports.

[9] See, for example, J. M. Renes, Ph. D. thesis, University of New Mexico, 2004, Fig. 6.5.

[10] R. B. M. Clarke, V. M. Kendon, A. Chefles, S. M. Barnett, E. Riis, and M. Sasaki, Phys. Rev. A **64**, 012303 (2001); see also A. Chefles, in *Quantum State Estimation* [13].

[11] Networks with only one auxiliary qubit are also possible. An example is given by T. Decker, D. Janzing, and T. Beth, Int. J. Quantum Inf. **2**, 353 (2004).

[12] The qubit of interest is always measured itself in the simpler networks of Ref. [11] and also in the practical schemes of Clarke *et al.* [10], the minimal single-loop setup of Ref. [6], and the no-loop setup of Ref. [7].

[13] *Quantum State Estimation*, edited by M. Paris and J. Řeháček, Lecture Notes in Physics Vol. 649 (Springer, Berlin, 2004).

[14] For example, one could take the likelihood of Eq. (4.3) as the weight function for the calculation of the mean of all Pauli vectors $\vec{s}$, thus averaging over the part of the cloud in Fig. 5 that is inside the Bloch sphere, and then accept this average as the best guess. Consult the classic text by Helstrom [15] for a general introduction to quantum-state estimation, or turn to Ref. [13] for a very recent account.

[15] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[16] R. A. Fisher, Proc. Cambridge Philos. Soc. **22**, 700 (1925).

[17] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946).

[18] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).

[19] By construction, the procedure of [14] would always give a physically acceptable estimate.

[20] A detailed analysis of how large N must be for such asymptotic large-N approximations to be valid has been performed by S. L. Braunstein, J. Phys. A **25**, 3813 (1992).

[21] C. R. Rao, Bull. Calcutta Math. Soc. **37**, 81 (1945).

[22] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).

[23] R. Derka, V. Bužek, and A. K. Ekert, Phys. Rev. Lett. **80**, 1571 (1998).

[24] D. G. Fischer and M. Freyberger, Phys. Lett. A **273**, 293 (2000).

[25] D. G. Fischer, S. H. Kienle, and M. Freyberger, Phys. Rev. A **61**, 032306 (2000).

[26] T. Hannemann, D. Reiß, C. Balzer, W. Neuhauser, P. E. Toschek, and C. Wunderlich, Phys. Rev. A **65**, 050303 (2002).

[27] Further support in favor of the aligned adaptive scheme is given by the figures of merit obtained for the two-qubits-only situation of Sec. VII B by the procedure of [14]. The numbers corresponding to those in Eqs. (7.3) and (7.4) are $11/21$ and $21/40$, when all input states are averaged over, as well as $19/24$ and $4/5$, when the average is over pure states only. In both cases, the aligned scheme does slightly better than the antialigned scheme.

[28] Y. C. Liang, D. Kaszlikowski, B.-G. Englert, L. C. Kwek, and C. H. Oh, Phys. Rev. A **68**, 022324 (2003).

[29] J. M. Renes, e-print quant-ph/0402135.

[30] B.-G. Englert, D. Kaszlikowski, H. K. Ng, W. K. Chua, and J. Anders (unpublished).