

Experimental demonstration of adaptive quantum state estimation for single photonic qubitsRyo Okamoto,^{1,2,3} Satoshi Oyama,^{2,3} Koichi Yamagata,^{4,5} Akio Fujiwara,^{4,*} and Shigeki Takeuchi^{1,2,3,†}¹*Department of Electronic Science and Engineering, Kyoto University, Kyoto 615-8510, Japan*²*Research Institute for Electronic Science, Hokkaido University, Sapporo 060-0812, Japan*³*The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan*⁴*Department of Mathematics, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043, Japan*⁵*Department of Information and System Engineering, Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan*

(Received 23 August 2016; published 18 August 2017)

We herein report the experimental demonstration of adaptive quantum state estimation for totally unknown photonic qubits. Similar to our previous study [R. Okamoto *et al.*, *Phys. Rev. Lett.* **109**, 130404 (2012)], the measurement configuration is updated using the results of each photon detection event so that our method does not require prior knowledge of the total number of samples. The experimental results obtained herein demonstrate both strong consistency and asymptotic efficiency through several rigorous statistical tests. Furthermore, we demonstrate that the experimentally obtained distribution of the states estimated using an adaptive quantum state estimation is significantly different from that obtained by conventional state tomography and agrees well with theoretical predictions.

DOI: [10.1103/PhysRevA.96.022124](https://doi.org/10.1103/PhysRevA.96.022124)**I. INTRODUCTION**

Estimating an unknown quantum state is one of the most fundamental and important tasks in the fields of quantum information [1–3], communication [4], and measurement [5–7]. The ultimate goal is to estimate the true value of the parameter that specifies the unknown state with the smallest uncertainty, such as the quantum Cramér-Rao bound [8,9], for a given limited number of samples. However, in general, it is impossible to achieve such a theoretical limit when the measurement configuration is fixed. As a solution to this problem, Nagaoka [10,11] advocated an adaptive quantum state estimation (AQSE) procedure. Later, the strong consistency and the asymptotic efficiency of AQSE were mathematically proven by Fujiwara [12,13] and were also experimentally verified by the present authors for one-dimensional parameter problems using photons [14].

Recently, a few experimental demonstrations of adaptive measurement for photonic qubits were reported [15,16]. However, these experiments used the two-step adaptive strategy in which the preliminary measurement of part of the whole ensemble was used to determine the measurement configuration for the remainder of the ensemble. Such a method may have two problems. First, this method cannot, in principle, achieve the theoretical limit exactly. Second, prior knowledge of the total number of samples is required to design the overall estimation procedure.

Adaptive schemes using the Bayesian estimation algorithm were also reported [17,18]. However, their strategy did not have an objective figure of merit to be minimized; instead, it involved the minimization of the Shannon entropy of the posterior distribution calculated from the arbitrarily fixed initial prior distribution and the history of the measurement. Such a method may have two problems. First, there is no

guarantee that an objective figure of merit, such as the mean infidelity, will converge to its theoretical limit predicted by the standard quantum point estimation theory [9,19]. In fact, the performance reported in [18] remained below the theoretical limit. Second, the time-consuming minimization procedure of the Shannon entropy is inevitable at every step of the measurement, which may cause difficulties in realizing the estimation scheme by experiment.

In the present paper, we report an experimental demonstration of a multiparameter AQSE for totally unknown photonic qubits. Similar to the one-parameter AQSE presented in [14], the measurement configuration is updated using the results of each photon detection event so that our method does not require prior knowledge of the total number of samples (Fig. 1). The experimental results obtained herein demonstrate both strong consistency and asymptotic efficiency through several rigorous statistical tests. Furthermore, we show that the experimentally obtained distribution of the states estimated using AQSE is significantly different from that obtained by conventional state tomography [20,21] and agrees well with theoretical predictions. These results are important for understanding the fundamental characteristics of quantum estimation procedures.

II. QUANTUM ESTIMATION THEORY

A d -dimensional quantum statistical model is a family $\mathcal{S} = \{\rho_\theta\}_\theta$ of density operators on a Hilbert space \mathcal{H} that is smoothly parametrized by a d -dimensional real parameter $\theta \in \mathbb{R}^d$. Suppose that the state of the physical system at hand belongs to model \mathcal{S} , but we do not know which is the true state. A quantum parameter estimation problem [8,9] involves seeking the best strategy for estimating the true value of θ that specifies the true state. An estimator for the parameter θ is given by pair $(M, \check{\theta})$, where M is a positive-operator valued measure (POVM) for which the outcomes take values on some set \mathcal{X} , and $\check{\theta} : \mathcal{X} \rightarrow \mathbb{R}^d : x \mapsto \check{\theta}(x)$ is a map that gives an estimated value $\check{\theta}(x)$ from the observed data x .

*fujiwara@math.sci.osaka-u.ac.jp

†takeuchi@kuee.kyoto-u.ac.jp

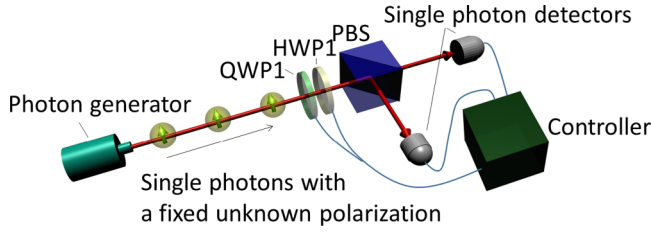


FIG. 1. Schematic diagram of adaptive quantum state estimation for single photonic qubits. Single photons with a fixed unknown polarization are emitted from a photon generator. The polarization is analyzed by QWP1, HWP1, and a polarizing beam splitter (PBS). The controller sets QWP1 and HWP1 to an angle calculated based on the photon measurement results.

It is known that the quantum Cramér-Rao bound, which gives the ultimate limit of the estimation precision when $d = 1$, is not always achievable when $d \geq 2$. Therefore, it is customary to seek the (locally unbiased) estimator $(M, \check{\theta})$ that minimizes the weighted trace $\text{Tr}\{W_\theta V_\theta[M, \check{\theta}]\}$ of the covariance matrix $V_\theta[M, \check{\theta}]$ for the estimator, given a weight (positive-definite symmetric) matrix W_θ , which may depend on θ .

In the present paper, we restrict our attention to estimating the three-dimensional parameter $\theta = (x, y, z)$ of a qubit state

$$\rho_{(x,y,z)} = \frac{1}{2}(I + x\sigma_x + y\sigma_y + z\sigma_z)$$

on $\mathcal{H} = \mathbb{C}^2$ having the domain

$$B = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \leq 1\}$$

called the Bloch ball. Although far from optimal [19], conventional state tomography is an example of an estimator for the parameter θ , where the number of photons is counted in each direction of the fixed coordinate axes [20,21]. In fact, the ultimate limit of the estimation precision is given by

$$\min_{(M, \check{\theta})} \text{Tr}\{J_\theta V_\theta[M, \check{\theta}]\} = (\text{Tr}\sqrt{\sqrt{J_\theta^{-1} W_\theta} \sqrt{J_\theta^{-1}}})^2, \quad (1)$$

where J_θ is the symmetric logarithmic derivative (SLD) Fisher information matrix. The bound (1) is sometimes referred to as the Hayashi-Gill-Massar bound [19,22,23].

An estimator that attains the bound (1) is given in [19]; in particular, it depends on the weight W_θ . It is then natural to ask what weight we should adopt. Recall that, due to the quantum Cramér-Rao inequality, the SLD Fisher information matrix J_θ is regarded as a measure of the statistical inhomogeneity of the quantum state space. Taking this into account, we adopt J_θ itself as the weight W_θ in the present study to compensate for the statistical inhomogeneity [24]. Consequently, the bound (1) with this weight becomes independent of θ , taking the following constant value:

$$\min_{(M, \check{\theta})} \text{Tr}\{J_\theta V_\theta[M, \check{\theta}]\} = 9. \quad (2)$$

In comparison, the weighted trace of the covariance matrix for simple tomography (see the Appendix for the definition) is given by

$$\text{Tr}\{J_\theta V_\theta[M_{\text{tomo}}, \check{\theta}_{\text{tomo}}]\} = 9 + \frac{6(x^2 y^2 + y^2 z^2 + z^2 x^2)}{1 - (x^2 + y^2 + z^2)}. \quad (3)$$

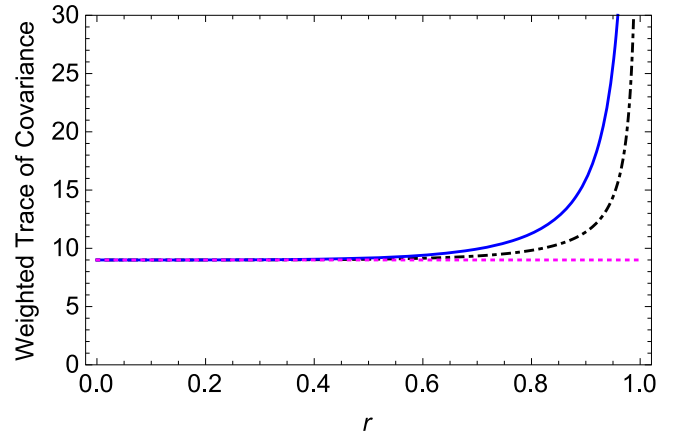


FIG. 2. Weighted traces of covariance matrices when $\theta = \frac{r}{\sqrt{3}}(1, 1, 1)$ with $0 \leq r < 1$. The dashed purple line, the blue curve, and the dot-dashed black curve indicate the results obtained by the optimal estimator (2), simple tomography (3), and totally randomized tomography (4), respectively.

Clearly, (3) is greater than (2) unless θ lies on an x, y , or z axis. Note that, due to the law of large numbers, the weighted trace of the covariance matrix for the standard maximum likelihood (ML) tomography (see the Appendix) is asymptotically identical to (3).

For further comparison, we consider a totally randomized tomography in which the measurement axes are chosen at random according to the Haar measure of the rotation group $\text{SO}(3)$. The weighted trace of the covariance matrix for this randomized tomography is given by

$$\text{Tr}\{J_\theta V_\theta[M_{\text{rand}}, \check{\theta}_{\text{rand}}]\} = \frac{r^3}{(1-r^2)(\text{arctanh } r - r)} + \frac{4r^3}{r - (1-r^2)\text{arctanh } r}, \quad (4)$$

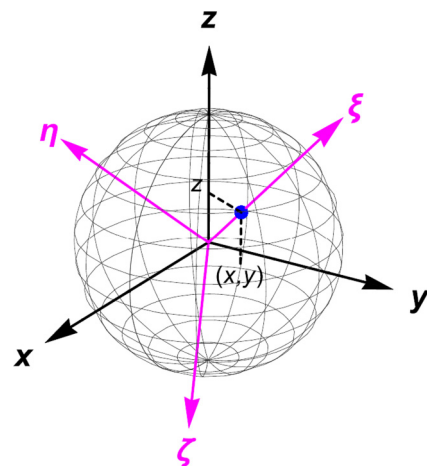


FIG. 3. An optimal measurement that attains the bound (2) is a uniform mixture of three projective measurements Π_ξ, Π_η , and Π_ζ corresponding to a set of mutually orthogonal axes ξ, η , and ζ in \mathbb{R}^3 in which the ξ axis passes through the true state $\theta = (x, y, z)$.

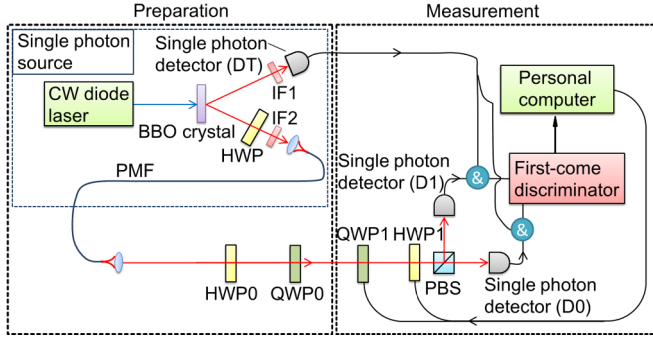


FIG. 4. Schematic diagram of the experimental setup.

where $r = \sqrt{x^2 + y^2 + z^2}$. Note that (4) depends only on the purity r : this is because the POVM M_{rand} is rotationally covariant.

The differences among (2), (3), and (4) for $\theta = \frac{r}{\sqrt{3}}(1, 1, 1)$ with $0 \leq r < 1$ are demonstrated in Fig. 2. The quantities (3) and (4) exhibit similar behavior; they diverge as the purity r approaches 1. This makes a striking contrast to the theoretical bound (2), which is independent of r . In the present paper, therefore, we adopt simple tomography and ML tomography as references in comparing the performance of estimators.

An estimator that attains the bound (2) is explicitly given by a uniform mixture (i.e., randomization)

$$M(\theta) := \frac{1}{3}(\Pi_\xi \oplus \Pi_\eta \oplus \Pi_\zeta) \quad (5)$$

of projective measurements Π_ξ , Π_η , and Π_ζ corresponding to a suitably chosen set of orthogonal axes ξ, η , and ζ , respectively, in \mathbb{R}^3 so that the ξ axis passes through the true state $\theta = (x, y, z)$ (see Fig. 3).

The reader may find that there is an obvious difficulty in realizing the optimal measurement (5) because information about the unknown value of the parameter θ is required to specify the ξ axis. This is in striking contrast to simple and/or ML tomography, which is represented by the uniform mixture

$$M_{\text{tomo}} = \frac{1}{3}(\Pi_x \oplus \Pi_y \oplus \Pi_z)$$

of projective measurements Π_x, Π_y , and Π_z corresponding to the “fixed” orthogonal axes x, y , and z in \mathbb{R}^3 .

This paradoxical difficulty can be avoided by employing an adaptive quantum state estimation (AQSE) scheme [12–14], in which successive measurements are performed to update the temporary estimate $\hat{\theta}_{n-1}$, as well as the temporary experimental setup $M(\hat{\theta}_{n-1})$, to a new estimate $\hat{\theta}_n$ based on the n th measurement and the maximum likelihood method. It has been proven [12,13] that the sequence $\hat{\theta}_n$ almost certainly converges to the true value θ^* of the parameter θ (referred to as *strong consistency*) and that the distribution of $\sqrt{n}(\hat{\theta}_n - \theta^*)$ converges to the normal distribution $N(0, V_{\theta^*})$, where $V_{\theta^*} := 3(I - |\theta^*\rangle\langle\theta^*|)$ is the covariance matrix that corresponds to the optimal measurement (5) at θ^* (referred to as *asymptotic efficiency*).

III. EXPERIMENT

Figure 4 shows the experimental setup. Single photons at 780 nm are generated from a heralded single-photon source, consisting of a cw diode pump laser (wavelength: 402 nm) and a 3-mm-long BBO crystal (Type I). A pair consisting of a signal photon (780 nm) and a trigger photon (830 nm) is created via spontaneous parametric down conversion. The detector (DT, SPCM-AQR, PerkinElmer) after an interference filter (IF1, center wavelength: 830 nm) outputs an electric pulse (width: 30 ns) when it detects a trigger photon and the electric pulse heralds the generation of a signal photon, which is coupled to a polarization maintaining fiber (PMF) after an interference filter (IF2, center wavelength: 780 nm, width: 4 nm). The HWP before the IF2 is used to change the purity of the input state. The purity can be controlled by tilting the angle of linear polarization to the slow or fast axis of the PMF due to birefringence. The polarization direction of the photon was set using HWP0 and QWP0.

The input photon state was analyzed by HWP1, QWP1, and a polarizing beam splitter (PBS). After passing through the PBS, photons are guided to single-photon detectors (D0 and D1, SPCM-AQR, PerkinElmer) at each PBS output port. The outputs of single-photon detectors are gated by the rise

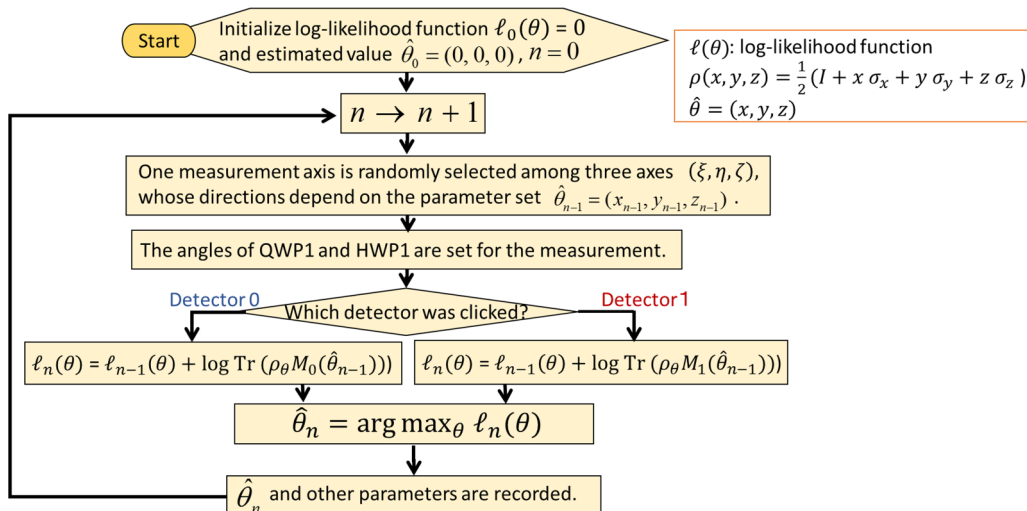


FIG. 5. Flowchart of AQSE used in the experiment.

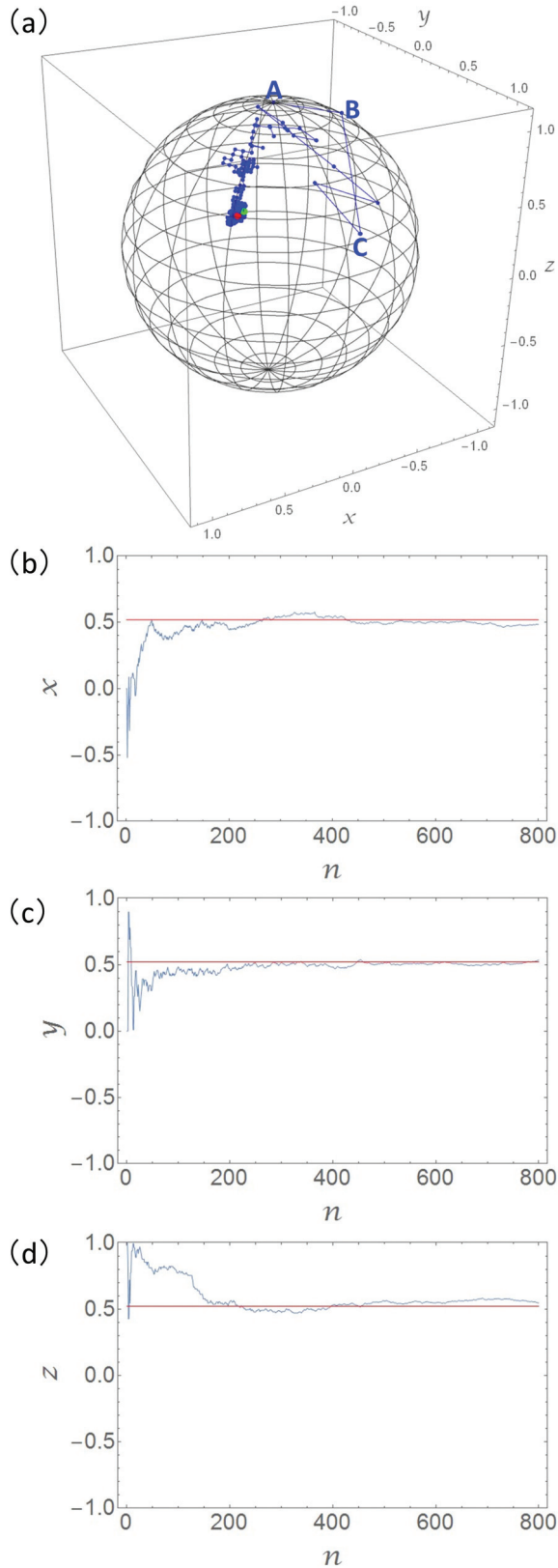


FIG. 6. Single measured trajectory for a true state of $\theta^* = (0.52, 0.52, 0.52)$. (a) Trajectory in the Bloch ball. The red dot indicates the true state and the green dot indicates the estimated state at $n = 798$. The first three estimated states are labeled A, B, and C, respectively. (b)–(d) Estimated states projected to x , y , and z axes. The red horizontal lines indicate the true state.

of the heralding signal and are connected to a first-come discriminator, consisting of a custom electric circuit. When the discriminator receives the first signal from one of the detectors (D0 or D1) after the measurement for the $(n - 1)$ th photon starts, the discriminator indicates which detector has been clicked. A minimum pulse interval of 2.5 ns can be discriminated. Note that the discriminator ignores the case in which it receives the pulses from both detectors within 2.5 ns.

The angle of HWP1 and QWP1 for measuring the n th photon is determined by the log-likelihood function at stage n chosen from among the 100 points that divide the domain $(-1, 1)$ of each parameter x , y , and z into equal parts. When the change of the HWP1 and QWP1 angles is completed, the measurement for the next (n th) photon will be started.

The procedure of AQSE is summarized in Fig. 5. In the present study, the procedure was carried out for up to 798 input photons ($n = 798$). We repeated the adaptive estimation sequence 200 times.

Figure 6 shows a single measured trajectory for a true state of $\theta^* = (0.52, 0.52, 0.52)$ [red dot in Fig. 6(a) and red horizontal lines in Figs. 6(b) through 6(d)]. Figure 6(a) is the trajectory in the Bloch ball, and Figs. 6(b) through 6(d) are the trajectories projected to x , y , and z axes, respectively. The state for $n = 0$ was initially set to be $(0, 0, 0)$ (not shown in Fig. 6). The trajectory starts at state A for $n = 1$ in Fig. 6(a) and then approaches the true state.

IV. RESULTS AND DISCUSSIONS

As mentioned in the previous section, the temporary estimate $\hat{\theta}_{n-1}$, which determines the measurement $M(\hat{\theta}_{n-1})$ at the next stage n , is taken at a lattice point of side $1/50$. In the following data analysis, however, we do not necessarily use these rough estimates. Instead, we can recalculate the true maximizer of the likelihood function that is reconstructed from the history $\{M(\hat{\theta}_n)\}_n$ of POVMs actually used. In what follows, all estimates $\hat{\theta}_n$ for $\theta^* = (0.52, 0.52, 0.52)$ refer to the maximum likelihood estimators recalculated in this way (cf., Appendix).

Let us first verify the strong consistency for the sequence $\hat{\theta}_n$. Figure 7 shows the first 10 trajectories of $\hat{\theta}_n = (\hat{x}_n, \hat{y}_n, \hat{z}_n)$ with respect to the number n of photons when the true state is set to be $\theta^* = (0.52, 0.52, 0.52)$. The curves correspond to independent runs of adaptive estimation. Evidently, each curve approaches the true value θ^* , which agrees with the mathematical result that $\hat{\theta}_n \rightarrow \theta^*$ almost certainly as $n \rightarrow \infty$.

We next carry out statistical tests for the null hypothesis that the sequence $\hat{\theta}_n$ follows a normal distribution for large n . More concretely, we investigate whether the random variable $\sqrt{n}(\hat{\theta}_n - \bar{\theta})$ follows the normal distribution $N(0, V_{\bar{\theta}})$, where $\bar{\theta}$ is the sample average of the estimated values $\hat{\theta}_n$ over 200 independent trials, and $V_{\bar{\theta}} := 3(I - |\bar{\theta}\rangle\langle\bar{\theta}|)$. Table I summarizes the results of some typical goodness-of-fit tests implemented on MATHEMATICA 10.3 for the data at $n = 798$ with $\bar{\theta} = (0.501928, 0.499555, 0.542737)$. Each P-value evaluates the probability of obtaining the experimental data under the null hypothesis $H_0 : \sqrt{n}(\hat{\theta}_n - \bar{\theta}) \sim N(0, V_{\bar{\theta}})$. If the P-value turns out to be less than a previously chosen significance level α ,

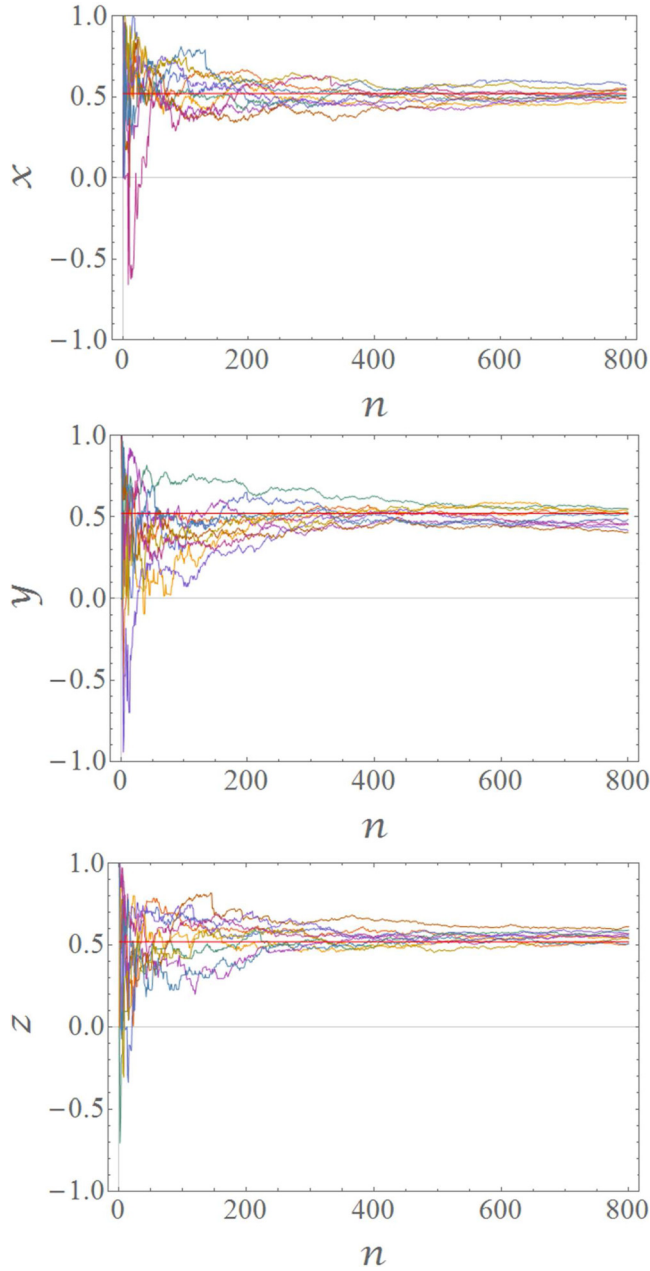


FIG. 7. Trajectories of estimated Stokes parameters $\hat{\theta}_n = (\hat{x}_n, \hat{y}_n, \hat{z}_n)$ plotted with respect to the number n of photons for the first 10 repetitions. The true value of the parameter is set to be $\theta^* = (0.52, 0.52, 0.52)$, as indicated by the red horizontal lines.

say 10%, we judge that the experimental data do not follow the null hypothesis H_0 , and the hypothesis H_0 is rejected under significance level α . Now, Table I shows that the experimental data follow the null hypothesis H_0 surprisingly well.

Next, let us compare AQSE with simple and/or ML tomography. Figure 8(a) shows the 200 estimated states at $n = 798$ with AQSE. The estimated states are distributed around the true state with the typical pancake-like distribution [15,19] of AQSE. Figure 8(b) shows the 200 states estimated at $n = 798$ with simple tomography. The distribution for simple tomography, which includes unphysical states

TABLE I. Goodness-of-fit tests.

Test	P-value	Result
Anderson-Darling	99.0079%	accept
Cramér-von Mises	99.5538%	accept
Kolmogorov-Smirnov	99.4815%	accept
Pearson χ^2	83.9265%	accept

falling outside the Bloch ball, is spherical. ML tomography avoids those unphysical estimates as shown in Fig. 8(c). Note that this refinement is equivalent to projecting the unphysical estimates in Fig. 8(b) onto the Bloch ball with respect to the Fisher metric [21]. Consequently, estimated values obtained by ML tomography are hemispherically distributed.

To demonstrate quantitatively the difference between AQSE and simple and/or ML tomography, the weighted traces of sample covariances obtained by 200 independent trials for

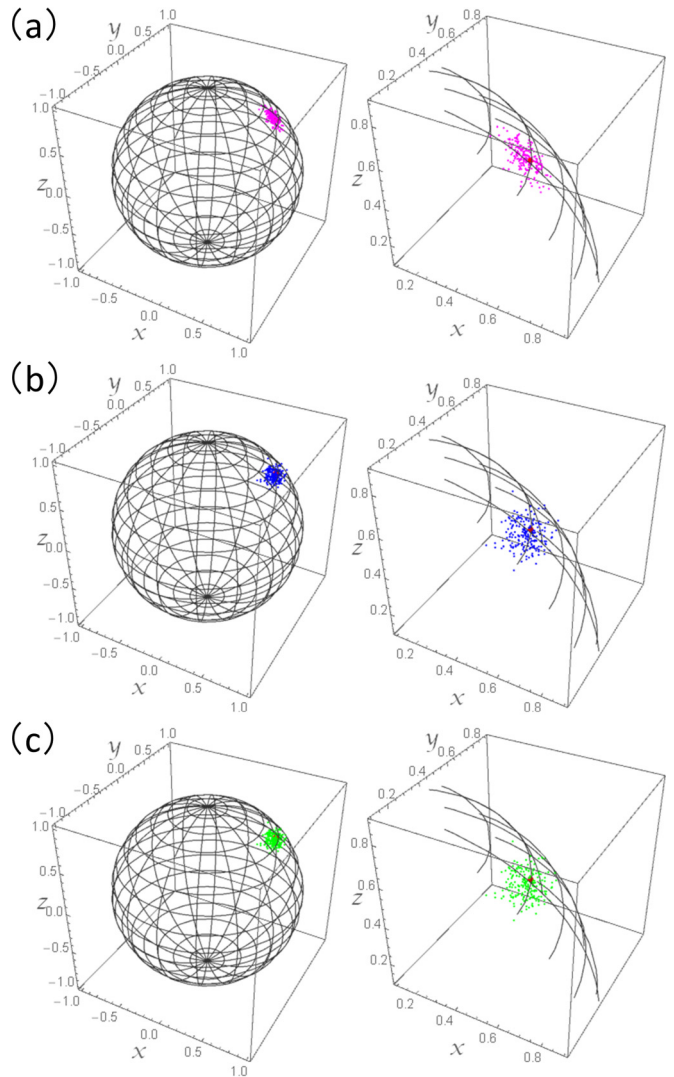


FIG. 8. Empirical distributions at $n = 798$ for (a) AQSE, (b) simple tomography, and (c) ML tomography when $\theta^* = (0.52, 0.52, 0.52)$. The true θ^* value is shown by the red points.

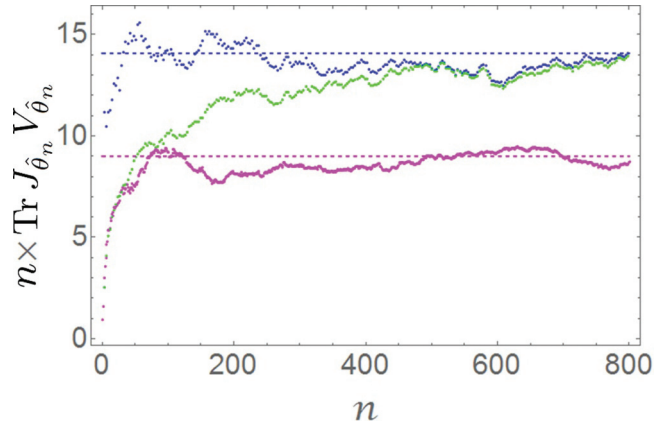


FIG. 9. Weighted trace $\text{Tr } J_{\hat{\theta}_n} V_{\hat{\theta}_n}$ of sample covariance $V_{\hat{\theta}_n}$ obtained by 200 independent trials, multiplied by n , when $\theta^* = (0.52, 0.52, 0.52)$. The purple, blue, and green dots indicate the data obtained by AQSE, simple tomography, and ML tomography, respectively. The purple and blue dashed horizontal lines indicate the theoretical values of (2) and (3) evaluated from the sample average $\bar{\theta}$ at $n = 798$.

$\theta^* = (0.52, 0.52, 0.52)$ are plotted in Fig. 9. The purple, blue, and green dots indicate the data for AQSE, simple tomography, and ML tomography, respectively. Furthermore, the purple and blue dashed horizontal lines indicate the theoretical limiting values given by (2) and (3), these last being evaluated based on the sample average $\bar{\theta} = (0.515752, 0.491917, 0.509737)$ at $n = 798$. In AQSE (purple), the figure of merit $\text{Tr } J_{\hat{\theta}_n} V_{\hat{\theta}_n}$ quickly approaches and evolves around the theoretical value. The evolution of simple tomography (blue) is similar. In contrast, the figure of merit for ML tomography (green) shows a slow uphill and eventually merges with the curve obtained by simple tomography. This odd behavior is due to the maximum likelihood data processing, which forces the unphysical estimates to be projected onto the Bloch ball, yielding a false shrinkage of variance. Put differently, the seemingly small values of $\text{Tr } J_{\hat{\theta}_n} V_{\hat{\theta}_n}$ for ML tomography at an early stage of estimation are illusive. The figure of merit eventually converges to the performance of simple tomography because the values estimated by simple tomography for large n are likely to fall inside the unit ball, as the law of large numbers asserts.

The superiority of AQSE compared to tomography becomes much clearer if the true state is taken in the vicinity of the surface of the Bloch ball, that is, if the true state is (almost) pure. Figure 10 shows the 200 estimated states at $n = 798$ when the true state is set to be $\theta^* = (0.577, 0.577, 0.577)$. For this true state, the estimated states of AQSE [Fig. 10(a)] have a much thinner pancake-like distribution, as compared to Fig. 8(a). Simple tomography [Fig. 10(b)] still has a spherical distribution around θ^* and since the center point is near the surface now, approximately half of the estimated states are unphysical and are to be projected onto the Bloch ball by ML tomography [Fig. 10(c)]. In this way, the number of unphysical estimates increases as the purity approaches 1, and thus the effect of the maximum likelihood data processing is sensitive to the purity.

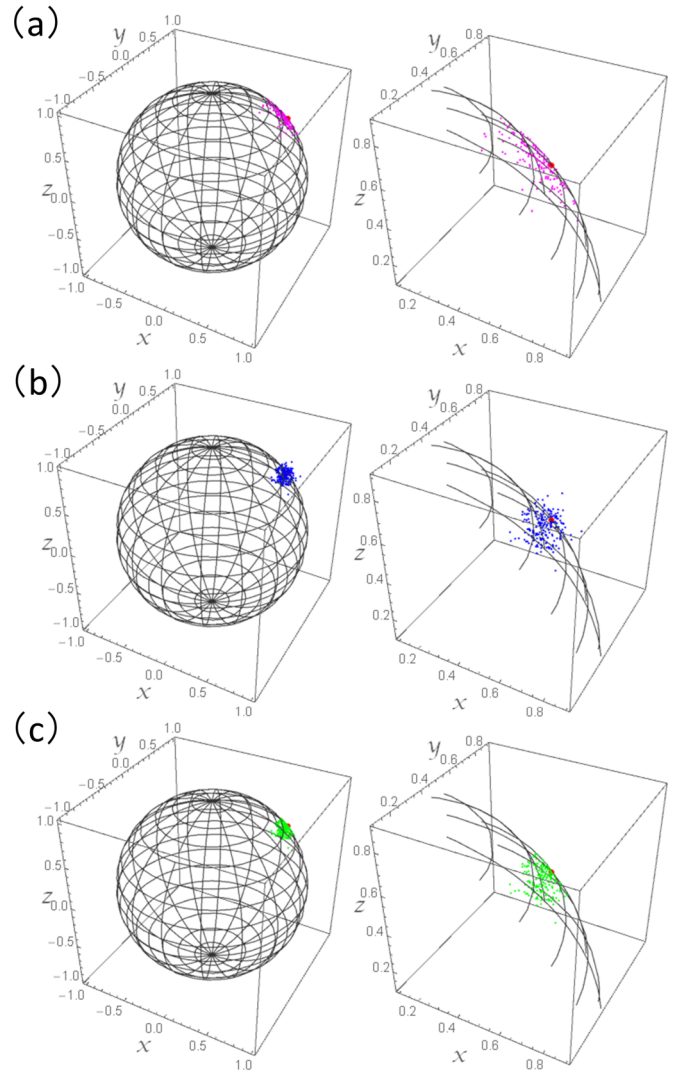


FIG. 10. Empirical distributions at $n = 798$ for (a) AQSE, (b) simple tomography, and (c) ML tomography when $\theta^* = (0.577, 0.577, 0.577)$. The true θ^* value is shown by the red points.

Let us observe this tendency from a different viewpoint. Figure 11 shows the evolution of the weighted trace of sample covariance for $\theta^* = (0.577, 0.577, 0.577)$. The purple, blue, and green dots indicate the data obtained by AQSE, simple tomography, and ML tomography, respectively. Furthermore, the purple and blue dashed horizontal lines indicate the corresponding theoretical values for AQSE and simple tomography, respectively, the latter being evaluated based on the sample average $\bar{\theta} = (0.562556, 0.535301, 0.565639)$ at $n = 798$. The figures of merit for AQSE and simple tomography quickly approach the corresponding theoretical values, whereas the figure of merit for ML tomography exhibits a very slow uphill, which would eventually merge with the curve of simple tomography. The convergence is very slow, as compared with Fig. 9 because the true state is almost pure now. Stated another way, the efficiency of AQSE compared to simple and/or ML tomography becomes more evident as the true state approaches the boundary of the Bloch ball. Thus, the performance of AQSE is clearly superior to that of conventional state tomography.

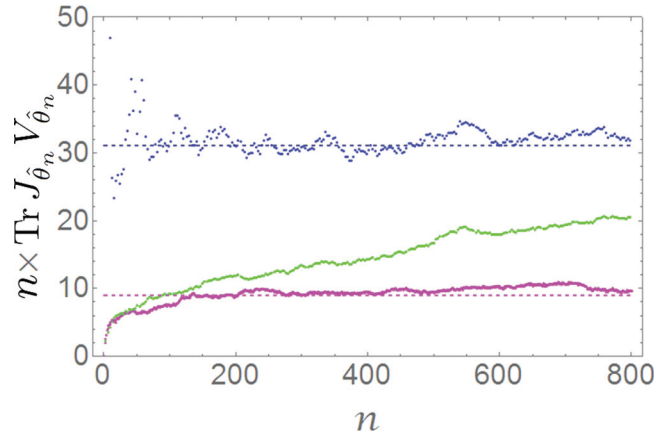


FIG. 11. Weighted trace $\text{Tr } J_{\hat{\theta}_n} V_{\hat{\theta}_n}$ of sample covariance $V_{\hat{\theta}_n}$ obtained by 200 independent trials, multiplied by n , when $\theta^* = (0.577, 0.577, 0.577)$. The purple, blue, and green dots indicate the data obtained by AQSE, simple tomography, and ML tomography, respectively. The purple and blue dashed horizontal lines indicate the theoretical values of (2) and (3) evaluated from the sample average $\bar{\theta}$ at $n = 798$.

V. CONCLUSION

In conclusion, we report the experimental demonstration of AQSE for totally unknown photonic qubits. The experimental results obtained herein reveal both strong consistency and asymptotic efficiency through several rigorous statistical tests. Furthermore, we show that the experimentally obtained distribution of the states estimated using AQSE was significantly different from that obtained by conventional state tomography and agreed well with theoretical predictions. These results are important for clarifying the fundamental characteristics of quantum estimation procedures. AQSE will find a wide range of applications from astronomy to molecular biology, whenever precise measurements for quantum objects, such as photons, spins, atoms, and superconducting artificial atoms, are needed.

ACKNOWLEDGMENTS

The authors thank Professor Holger F. Hofmann for helpful comments. The present study was supported, in part, by MEXT/JSPS KAKENHI Grants No. JP26220712, No. JP21102007, and No. JP22340019, JST-CREST No. JPMJCR1674, Special Coordination Funds for Promoting Science and Technology.

APPENDIX: COMPARISON BETWEEN CONVENTIONAL STATE TOMOGRAPHY AND AQSE

Suppose that, at the i th step of estimation, a POVM $M_i = \{M_i[\cdot]\}$ was chosen and an outcome d_i was obtained. The

maximum likelihood estimate that takes values on a parameter domain Θ at stage n is given by

$$\hat{\theta}_n := \underset{\theta \in \Theta}{\text{argmax}} \ell_n(\theta | M, d),$$

where

$$\ell_n(\theta | M, d) := \sum_{i=1}^n \log \text{Tr} \{\rho_\theta M_i[d_i]\}$$

is the log-likelihood function determined from the history

$$(M, d) = \{(M_1, d_1), (M_2, d_2), \dots, (M_n, d_n)\}$$

of experiments.

In the conventional qubit state tomography for estimating the Stokes parameter $\theta = (x, y, z)$, each POVM M_i is chosen from among the projective measurements Π_x , Π_y , and Π_z corresponding to the fixed orthogonal axes x , y , and z in \mathbb{R}^3 . When the domain Θ of the parameter is taken to be the cubic region $[-1, 1]^3$, we call the corresponding estimation scheme the *simple tomography*, and denote the estimate symbolically as

$$\hat{\theta}_n^{\text{simple}} := \underset{\theta \in [-1, 1]^3}{\text{argmax}} \ell_n(\theta | M^{\text{fix}}, d).$$

Here M^{fix} indicates that the measurement axes are fixed. The simple tomography, also called the linear tomography in [20], amounts to computing the empirical distribution from the data [21].

Note that the estimate $\hat{\theta}_n^{\text{simple}}$ may fall outside the Bloch ball, giving an unphysical estimate. This drawback is circumvented by the standard *maximum likelihood tomography*, in which the parameter domain Θ is taken to be the Bloch ball $B = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 \leq 1\}$, so that

$$\hat{\theta}_n^{\text{ML}} := \underset{\theta \in B}{\text{argmax}} \ell_n(\theta | M^{\text{fix}}, d). \quad (\text{A1})$$

The relationship between $\hat{\theta}_n^{\text{simple}}$ and $\hat{\theta}_n^{\text{ML}}$ has been scrutinized in [21].

In AQSE, on the other hand, the n th measurement M_n is chosen to be the best one at the previous estimate $\hat{\theta}_{n-1}$; thus the n th estimate is symbolically written as

$$\hat{\theta}_n^{\text{AQSE}} := \underset{\theta \in B}{\text{argmax}} \ell_n(\theta | M^{\text{adapt}}, d). \quad (\text{A2})$$

Here M^{adapt} indicates that the measurement axes are chosen in an adaptive manner. It should be emphasized that the only difference between ML tomography (A1) and AQSE (A2) is whether each M_n is determined based on information of the history $\{(M_i, d_i)\}_{1 \leq i \leq n-1}$ of experiments or not.

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).

[2] J. L. O'Brien, *Science* **318**, 1567 (2007).

[3] R. Okamoto, J. L. O'Brien, H. F. Hofmann, T. Nagata, K. Sasaki, and S. Takeuchi, *Science* **323**, 483 (2009).

[4] H. J. Kimble, *Nature (London)* **453**, 1023 (2008).

- [5] T. Nagata, R. Okamoto, J. L. O'Brien, K. Sasaki, and S. Takeuchi, *Science* **316**, 726 (2007).
- [6] T. Ono, R. Okamoto, and S. Takeuchi, *Nat. Commun.* **4**, 2426 (2013).
- [7] R. Okamoto, S. Takeuchi, and K. Sasaki, *Phys. Rev. A* **74**, 011801(R) (2006).
- [8] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
- [9] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
- [10] H. Nagaoka, An asymptotically efficient estimator for a one-dimensional parametric model of quantum statistical operators, *Proceedings of the IEEE International Symposium on Information Theory* (IEEE, 1988), p. 198.
- [11] H. Nagaoka, On the parameter estimation problem for quantum statistical models, *Proceedings of the 12th Symposium on Information Theory and Its Applications* (SITA, 1989), pp. 577–582; in *Asymptotic Theory of Quantum Statistical Inference*, edited by M. Hayashi (World Scientific, Singapore, 2005), pp. 125–132 (reprinted).
- [12] A. Fujiwara, *J. Phys. A: Math. Gen.* **39**, 12489 (2006).
- [13] A. Fujiwara, *J. Phys. A: Math. Theor.* **44**, 079501 (2011).
- [14] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi, *Phys. Rev. Lett.* **109**, 130404 (2012).
- [15] D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, and R. Blume-Kohout, and A. M. Steinberg, *Phys. Rev. Lett.* **111**, 183601 (2013).
- [16] Z. Hou, H. Zhu, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, *npj Quantum Information* **2**, 16001 (2015).
- [17] D. G. Fischer, S. H. Kienle, and M. Freyberger, *Phys. Rev. A* **61**, 032306 (2000).
- [18] K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. T. Houlby, F. Huszár, and S. P. Kulik, *Phys. Rev. A* **87**, 062122 (2013).
- [19] K. Yamagata, *Int. J. Quant. Inform.* **09**, 1167 (2011).
- [20] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, *Phys. Rev. A* **64**, 052312 (2001).
- [21] A. Fujiwara and K. Yamagata, [arXiv:1608.07983](https://arxiv.org/abs/1608.07983).
- [22] M. Hayashi, in *Quantum Communication, Computing, and Measurement*, edited by O. Hirota *et al.* (Plenum, New York, 1997), pp. 99–108.
- [23] R. D. Gill and S. Massar, *Phys. Rev. A* **61**, 042312 (2000).
- [24] Note that the SLD-weighted trace of the covariance matrix is asymptotically almost equivalent to the infidelity as a figure of merit. In fact, in the Taylor expansion of the infidelity around the true value of the parameters, the expectations of the zeroth- and the first-order terms vanish, and the expectation of the second-order term is identical, up to the factor of 1/4, to the SLD-weighted trace of the covariance matrix.