Adaptive identification of coherent states

Markku P.V. Stenberg,* Kevin Pack, and Frank K. Wilhelm

Theoretical Physics, Saarland University, 66123 Saarbrücken, Germany

We present methods for efficient characterization of an optical coherent state $|\alpha\rangle$. We choose measurement settings adaptively and stochastically, based on data while it is collected. Our algorithm divides the estimation into two distinct steps: (i) before the first detection of a vacuum state, the probability of choosing a measurement setting is proportional to detecting vacuum with the setting, which makes using too similar measurement settings twice unlikely; and (ii) after the first detection of vacuum, we focus measurements in the region where vacuum is most likely to be detected. In step (i) [(ii)] the detection of vacuum (a photon) has a significantly larger effect on the shape of the posterior probability distribution of α. Compared to nonadaptive schemes, our method makes the number of measurement shots required to achieve a certain level of accuracy smaller approximately by a factor proportional to the area describing the initial uncertainty of α in phase space. While this algorithm is not directly robust against readout errors, we make it such by introducing repeated measurements in step (i).

I. INTRODUCTION

A fundamental task in quantum optics is the reconstruction of the state of the light field, the complete description of which is contained in the density matrix. Close to the boundary between quantum and classical regions the density matrix is conveniently studied in phase space representation, e.g., through the Wigner function [1], since this allows the investigation of the transition between the two regimes. There is a wealth of techniques [2–5] to measure the Wigner function, most of which require many copies of the state and many measurement shots. Producing many copies, however, is not always possible or efficient. In systems just crossing the classical-to-quantum boundary such as nanomechanical resonators [6, 7], for example, data can be so noisy and prone to drift that the required number of experimental runs with nominally identical parameters is not possible.

Often, this laborious task of full quantum state tomography is also asking a too general or too unspecific question. In many cases, it is sufficient to approximate the state by estimating a few parameters characterizing it. This is the case, *e.g.*, in the simple and *prima facie* classical task of measuring both the quadrature amplitude $|\alpha|$ and phase $\arg(\alpha)$ of a weak ac signal, for example, in the microwave range. These signals are represented as coherent quantum states $|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} \alpha^n |n\rangle$, with $|n\rangle$ photon number eigenstate with n photons. Estimating coherent states is an important stepping stone for the characterization of more complicated quantum states of light because many such states, e.g., a Schrödinger cat state $|\alpha\rangle + |-\alpha\rangle$ [8], a so-called voodoo cat state $e^{-i\frac{\pi}{3}}|e^{-i\frac{\pi}{3}}\alpha\rangle + e^{i\frac{\pi}{3}}|e^{i\frac{\pi}{3}}\alpha\rangle - |-\alpha\rangle$ [9], and a compass state $|\alpha\rangle + i|i\alpha\rangle - |-\alpha\rangle - i|-i\alpha\rangle$ [10], can be presented as

superpositions of coherent states. Direct applications of quadrature measurements include, for example, cosmic microwave background detection [11] and the search for dark matter axions [12].

Estimation of an optical phase alone, at fixed coherent state amplitude [13, 14], is the basis of many metrological applications, e.g., in magnetometry [15–17], detection of gravitational waves [18–20], and clock synchronization [21]. Phase estimation is also an indispensable component of several algorithms in quantum information processing [22]. Since it is difficult to define the concept of phase measurement for a single mode only, the usual approach is to consider two-mode measurements in an interferometer (see Fig. 1 in [23]). The ultimate limit set by quantum mechanics to the precision of phase measurements is due to complementarity between photon number and phase. This translates to a so-called Heisenberg limit, phase variance scaling as $\sim N^{-2}$ with N the total number of photons that pass through the interferometer. Note that inverse phase variance describes the Fisher information (variance of the score) [24] of the phase estimate.

Optimal measurements for phase estimation have been identified theoretically in [25–30] but not realized experimentally; it is not possible to perform them with photodetections at the output of the interferometer. Input states and measurements with an error scaling close to optimum (but with a different prefactor) have been proposed optimizing adaptively the next measurement in the series of measurements [23, 31] (local optimization). Adaptive measurements have also been designed attempting to optimize the whole series of measurements [32–35] (global optimization). The input state to the interferometer considered in [23, 26, 27, 31, 33, 35], however, is not separable between photon number eigenstates in the output arms of the interferometer and its creation is currently an open question. Beating the standard quantum limit has, however, been demonstrated using simpler entangled input states [36–39]. Here, we are go-

---

*markku.stenberg@iki.fi

ing beyond bare phase estimation in addressing simultaneous adaptive estimation of both phase and amplitude of α. In terms of density operators, we thus estimate the state within a family $\rho_\alpha = |\alpha\rangle\langle\alpha|, \ \alpha \in \mathbb{C}$.

The most popular approach to quantum state estimation has been maximum likelihood estimation (MLE) [40]. Given measurement settings $S = \{s_1, \ldots, s_M\}$ and corresponding data $D = \{d_1, \ldots, d_M\}$, it seeks for a physical state ρ that maximizes the likelihood functional $P(D|\rho, S) = \prod_{i=1}^{M} P(d_i|\rho, s_i)$, with $P(d_i|\rho, s_i)$ the probability to obtain the measurement outcome d_i given the state ρ and measurement setting s_i. However, MLE does not deliver confidence intervals for the estimates. Moreover, a basic problem with MLE is that it tends to assign vanishing values for certain eigenvalues of ρ [41]. This is unreasonable since it is not possible to completely rule out some measurement outcomes with a finite amount of data.

More advanced approaches based on Bayesian inference do not suffer from these shortcomings. Bayesian inference techniques have been developed, e.g, for phase [14, 17, 23, 31, 42, 43], state [44, 45], and Hamiltonian [46–51] estimation. For certain one-parameter estimation problems it is possible to perform local Bayesian optimization of the measurement settings analytically [14, 23, 31, 48, 49]. For a larger number of unknown parameters, however, finding optimal measurement settings adaptively becomes generally analytically intractable. To perform the Bayesian updates numerically, a sequential Monte-Carlo approach [44, 49–55] has recently undergone a strong development.

Bayesian reasoning provides a general framework to assign a probability distribution for system parameters given certain data. The nature of the data, however, depends on the measurement settings chosen by the experimenter. In general it is more effective to choose the measurement settings adaptively so that they depend on the data that has been collected. Generally, the specific set of rules, also called a policy, according to which the measurement setting s is to be updated, has to be developed separately for each problem at hand. This is true also for the recently developed technique called self-guided quantum tomography (SGQT) [56]. SGQT searches the estimate of a quantum state by making measurements in the directions close to the estimate, approaching it as a power law as a function of adaptive iteration steps. The optimal prefactors in the related power laws (as well as certain coefficient added to the base of an exponential), however, depend on the problem and parameter region.

II. EXPERIMENTAL SETUP

For definiteness, we consider measurement of the mode using an ideal vacuum detector [57–60]. A vacuum detector provides a click if there is more than zero photons in the cavity but does not give any indication of the photon number beyond that.
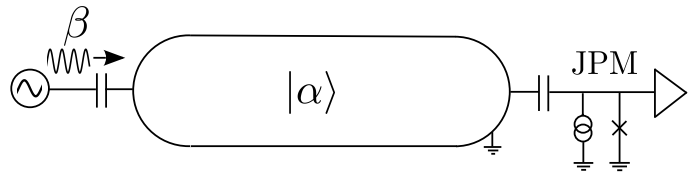

Figure 1: Measurement setup. The unknown coherent state within the cavity is denoted by $|\alpha\rangle$. The microwave source emits a pulse that displaces the coherent state by β in phase space. The cavity is capacitively coupled to a Josephson photomultiplier (JPM) that registers one of two possible measurement outcomes: The displaced state is either vacuum or a state containing photons.

Such detectors have been realized in optics in the strongly coupled regime. At microwave frequencies, the recently realized [61] Josephson photomultiplier [62, 63] has been shown theoretically [58] to reach this ideal vacuum detector limit at weak tunneling and long interaction time. In contrast to a standard Mach-Zender interferometric scheme for phase measurements, photon number resolving detectors or beam splitters are not needed, and there is no entanglement involved. This technology is also suited for the measurement of a qubit state and should allow better scalability to larger circuits than that based on superconducting amplifiers. The latter requires a strong auxiliary microwave pump tone that must be isolated from the qubit circuitry with bulky cryogenic isolators, which with the former technology can be eliminated.

In the measurement with a vacuum detector, different points in phase space may be accessed by injecting an additional drive pulse to the input signal, cf. Fig. 1. Appropriately normalized, the drive pulse $\Omega(t)$ displaces the coherent state by an amount $\beta = -\frac{i}{2}\int \Omega(t)dt$ in phase space, i.e., it turns $|\alpha\rangle$ into $|\alpha + \beta\rangle$, hence allowing one to measure the Husimi Q-function $Q(\alpha) = \frac{1}{\pi}\langle\alpha|\rho_{-\beta}|\alpha\rangle = \frac{1}{\pi}|\langle 0|\alpha + \beta\rangle|^2$, with $\rho_{-\beta} = |-\beta\rangle\langle-\beta|$.

In terms of a positive operator-valued measure (POVM), a measurement setting is thus characterised by a set of operators $\mathbb{M}_s$, with $s = \beta$ given by the pulse parameter. The set $\mathbb{M}_s$ is defined by $\mathbb{M}_s = \{M_{s,d}|d \in \{v, p\}\}$, where d indexes the possible measurement outcomes, a vacuum state or a state containing photons. Here, we only need the operators $M_{\beta,v} = |-\beta\rangle\langle-\beta|$ and $M_{\beta,p} = I - M_{\beta,v}$.

In [9] and [64], respectively, the states of a superconducting resonator and a nanomechanical resonator were characterized. The tomographic method used in these papers is analogous to the measurement model above, since it consists of displacing the resonator state by a microwave pulse and then performing a projective measurement of the qubit. In [65] a nonadaptive tomographic scheme based on semidefinite programming was presented for the characterization of NOON states in resonators coupled to qubits. This type of measurements can be contrasted to probing of the Wigner function by
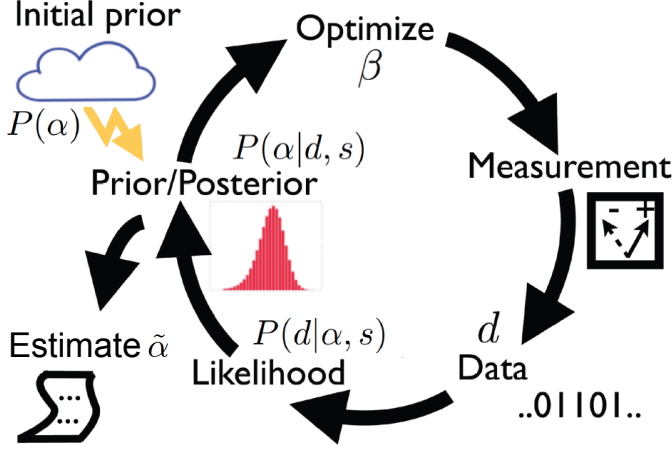
Figure 2: (Color online) Illustration of an adaptive Bayesian inference scheme.

full photon counting with number resolution [66, 67].

III. BAYESIAN INFERENCE

The basic difference between frequentist and Bayesian approaches to parameter estimation is that the latter allow one to assign an initial prior probability distribution $P(\alpha)$ to describe an unknown state $|\alpha\rangle$. It quantifies the experimenters *a priori* conception of the state and its uncertainty.

Once the initial prior is set and given a suitable policy, one can take advantage of the information contained in the prior in choosing the measurement setting s. The policies used in this paper are described in Sec. V. Bayesian inference proceeds by iteratively applying Bayes' theorem

$$P(\alpha|d, s) = \frac{P(d|\alpha, s)P(\alpha)}{P(d|s)} \tag{1}$$

as illustrated in Fig. 2. Here $P(d|\alpha, s)$, referred to as likelihood, is the probability to obtain data d (here, $d = \mathrm{v}, \mathrm{p}$) in the state $|\alpha\rangle$, given the measurement setting s. Using the notation of POVM in Sec. II, it is related to the Hermitian operators $M_{s,d}$ through Born's rule $P(d|\alpha, s) = \mathrm{Tr}(M_{s,d}\rho_\alpha)$. The normalization factor $P(d|s)$ is obtained by integrating the likelihood over all possible states $P(d|s) = \int P(d|\alpha, s)P(\alpha)d\alpha$. The probability distribution $P(\alpha|d, s)$ for $|\alpha\rangle$ given data d and measurement setting s is called the posterior. The posterior can be set as the prior for the next measurement which allows iterative application of Eq. (1)

$$
\begin{aligned}
P[\alpha|(d_{M+1}, s_{M+1}), (D, S)] = \\
\frac{P(d_{M+1}|\alpha, s_{M+1})P[\alpha|(D, S)]}{\int P(d_{M+1}|\alpha, s_{M+1})P[\alpha|(D, S)]d\alpha}, \\
(D, S) = \{(d_M, s_M)\dots, (d_1, s_1)\}.
\end{aligned}
\tag{2}
$$

Once a sufficient amount of data has been collected and the posterior is narrow enough, the estimate $\tilde{\alpha}$ is obtained from its mean value. The functional form of our likelihood function $P(d|\alpha, s)$ as well the rules to choose the measurement settings s are described in Sec. V.

IV. NUMERICAL METHOD

The sequential Monte-Carlo method [44, 52–55] delivers an efficient numerical method to perform the updates of the posterior. The posterior is approximated by keeping track of its value in N_p moving grid points, or "particles," $P[\alpha|(D, S)] \approx \sum_{n=1}^{N_\mathrm{p}} w_n \delta(\alpha - \alpha_n)$. Here α_n are the locations of the particles while w_n are their relative probabilities or weights that can be updated through Bayes' theorem,

$$\tilde{w}_n^{(m+1)} \leftarrow P(d_{m+1}|\alpha_n, s_n)w_n^{(m)}, \tag{3}$$

$$w_n^{(m+1)} \leftarrow \frac{\tilde{w}_n^{(m+1)}}{\sum_{n=1}^{N_\mathrm{p}} \tilde{w}_n^{(m+1)}}. \tag{4}$$

Here, $w_n^{(m)}$ are the weights evaluated after the mth measurement shot. Equation (4) ensures the norming $\sum_{n=1}^{N_\mathrm{p}} w_n = 1$ and the conservation of probability.

In the following, the key quantities are the mean $\tilde{\alpha}$ and the covariance matrix $\mathrm{Cov}(\alpha)$ of α over the posterior. Numerically, these are readily approximated through

$$\tilde{\alpha} = \int P(\alpha|D)\alpha d\alpha \approx \sum_{n=1}^{N_\mathrm{p}} w_n \alpha_n, \tag{5}$$

$$\mathrm{Cov}(\alpha) = \int P(\alpha|D)\alpha\alpha^T d\alpha - \tilde{\alpha}\tilde{\alpha}^T \approx \sum_{n=1}^{N_\mathrm{p}} w_n \alpha_n \alpha_n^T - \tilde{\alpha}\tilde{\alpha}^T,$$

$$\alpha = \begin{pmatrix} \mathrm{Re}(\alpha) \\ \mathrm{Im}(\alpha) \end{pmatrix}, \quad \alpha^T = [\mathrm{Re}(\alpha), \mathrm{Im}(\alpha)]. \tag{6}$$

A fixed grid would limit the achievable precision of the estimate, but we make use of an adaptive grid [44, 52–55] which makes it possible to focus the particles in the regions where the probability distribution concentrates. Here, N_p locations α_n are first chosen following the discrete probability distribution $\{w_n\}_{n=1}^{N_\mathrm{p}}$. The particles are then assigned new locations α'_n by sampling from the normal distribution

$$\alpha'_n \sim \mathcal{N}[\mu_n, (1 - a^2)\mathrm{Cov}(\alpha)], \tag{7}$$

with the mean

$$\mu_i = a\alpha_n + (1 - a)\tilde{\alpha} \tag{8}$$

and the covariance matrix $(1 - a^2)\mathrm{Cov}(\alpha)$. Here, a is a parameter that we set to $a = 0.999\,95$. Finally all the weights are set to $w_n = \frac{1}{N_\mathrm{p}}$. The artificial dynamics induced by Eqs. (7) and (8) conserves by construction the covariance matrix $\mathrm{Cov}(\alpha)$.
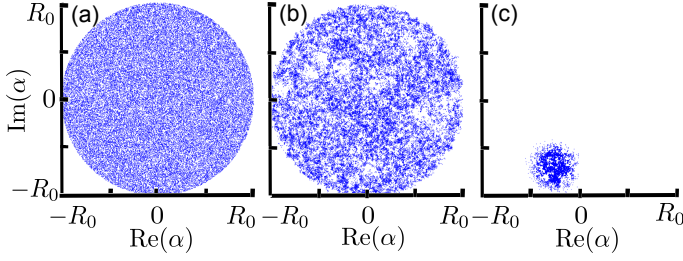
Figure 3: (Color online) Exemplary evolution of the particles (see text) describing the probability distribution of the unknown coherent state $|\alpha\rangle$ in phase space. (a) Uniform initial prior within the disk $|\alpha| < R_0$. Here, $R_0 = 10$. (b) Posterior after ten measurement shots. The "holes" at positions $\{\alpha'\}$ have been created by photon detections with pulse parameters $\beta = -\alpha'$. (c) Posterior after the first detection of a vacuum state (here the 11th measurement shot). The weight of the posterior is concentrated near $\alpha = -\beta_1$, with β_1 the pulse parameter corresponding to the first vacuum detection.

In our calculations we choose $N_{\mathrm{p}} = 50\,000$. To compare different policies, we apply them on $15\,000$ simulated samples with randomly chosen true values α. We choose α from a uniform distribution on the region $|\alpha| < R_0$, with $R_0 = 10$ [see Fig. 3(a)]. Note that R_0 is related to the maximum expectation value of the photon number operator through $R_0^2 = \langle \hat{n} \rangle_{\mathrm{max}}$. The initial prior was chosen to coincide with the aforementioned probability distribution. While here, this prior exactly incorporates what is known about estimated quantities before data collection, it has to be noted that in an actual experimental situation there is no unique and objective way to assign the initial prior, but choosing it necessarily involves certain arbitrary or subjective elements.

V. POLICIES

Our measurement setting s is defined by the pulse parameter β. We consider an ideal vacuum detector with the likelihood functions for the detections of vacuum (v) and a state containing photons (p), respectively

$$P(d = \mathrm{v}|\alpha, \beta) = \mathrm{Tr}(\rho_\alpha M_{\beta,\mathrm{v}}) = e^{-x^2}, \quad x = |\alpha + \beta|$$
$$P(d = \mathrm{p}|\alpha, \beta) = \mathrm{Tr}(\rho_\alpha M_{\beta,\mathrm{p}}) = 1 - P(d = \mathrm{v}|\alpha, \beta). \quad (9)$$

In the beginning of the experiment, the first detection of a vacuum state narrows the posterior $P(\alpha|D)$ significantly more than detection of a photon. In the latter case the posterior only changes in the proximity of $\alpha \approx -\beta$, where its value considerably decreases [see Fig. 3(b)]. However, in the former case the weight of the posterior is concentrated near $\alpha \approx -\beta$, while outside this region the posterior is exponentially decreased [see Fig. 3(c)]. We therefore start the experiment by choosing the displacement pulse β randomly from a probability distribution P_β such that $P_{-\beta} = P[\beta|(D, S)]$ equals the posterior (here the argument α has to be replaced by β). This makes
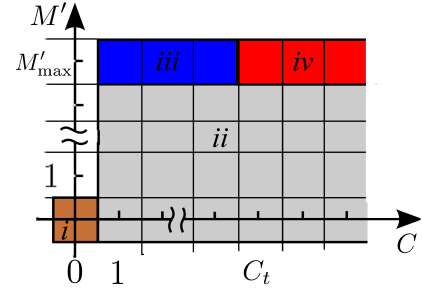

Figure 4: (Color online) Illustration of the policy of Eq. (12). The vertical and horizontal axes describe the number of vacuum detections and the number of measurement shots performed after the first vacuum detection, respectively. Different regions (i)-(iv) correspond the actions on the first to the fourth row of Eq. (12), respectively.

measurements with similar values of β unlikely. After the first detection of vacuum we adjust the support of P_β in the proximity of $\alpha \approx -\beta_1$, with β_1 the measurement setting with which vacuum is detected. Before the first vacuum detection, the measurements are thus relatively uninformative, whereas most of the vacuum detections take place within a region with a radius $\mathcal{O}(1)$ in phase space (see below). Hence, compared to nonadaptive schemes, focusing the adaptive measurements in the correct region makes the number of required measurement shots smaller approximately by a factor $\sim R_0^2$ or the area describing the initial uncertainty of α in phase space.

Specifically, we choose the measurement settings β according to the following policy [68]

$$P_\beta = \begin{cases} P[-\beta|(D, S)] & \text{if } C = 0, \\ \frac{1}{\pi r^2(C) R_\alpha^2} \text{ for } |\beta + \tilde{\alpha}| < r(C) R_\alpha & \text{if } C \geq 1, \\ 0 \text{ otherwise} & \text{if } C \geq 1. \end{cases} \quad (10)$$

Here, C is the number of measurement shots that have detected the vacuum state, $\tilde{\alpha}$ is the current mean of the posterior, and R_α describes the width of the region where the weight of the likelihood function is concentrated. More precisely, we define $R_\alpha = \sqrt{\mathrm{Tr}[\mathrm{Cov}(\alpha)] + \frac{1}{2}}$, with $\mathrm{Cov}(\alpha)$ the covariance matrix for the Bayesian probability distribution of α [see Eq. (6)]. After the first detection of vacuum, P_β is chosen to be a uniform probability distribution on a disk with a radius $r(C) R_\alpha$. We have carried out extensive numerical calculations to search an optimal $r(C)$ in the form of a power law

$$r(C) = aC^b, \quad (11)$$

with a and b constants. The pair $(a, b) \in \mathbb{R}^2$ parametrizes the space within which we search for a near-optimal policy.

Since with a single vacuum detection the posterior concentrates near $\alpha \approx -\beta$, the policy (10) is not robust against readout errors in the experiment. However, it can

be made robust against such errors by confirming that an absence of a detector click is due to vacuum state rather than a readout error. This can be achieved through repetition. In the presence of readout errors we search policies of the form (see Fig. 4)

$$\begin{cases} P_\beta = P[-\beta|(D,S)] & \text{if } C = 0, \\ \beta = \beta_1, M' \to M'+1 & \text{if } C \geq 1, M' < M'_{\max}, \\ \beta = \beta_1, M' \to 0, C \to 0 & \text{if } C_t > C \geq 1, M' = M'_{\max}, \\ P_\beta = \frac{1}{\pi r^2(C) R_\alpha^2} \text{ for } |\beta + \tilde{\alpha}| < r(C) R_\alpha, \ 0 \text{ otherwise} \\ \qquad\qquad\qquad\qquad \text{if } C \geq C_t, M' = M'_{\max}. \end{cases} \quad (12)$$

Here, the measurements are repeated $M'_{\max}$ times at the setting β_1 that indicates vacuum state (possibly a readout error). The variable M' counts the measurement shots performed after the vacuum detection. Should after $M'_{\max}$ shots the number of vacuum detections C be less than the threshold value C_t, the variables M' and C are set back to value 0. If after $M'_{\max}$ shots C is greater or equal to the threshold value C_t, the remaining measurement settings are chosen as in policy (10). Similarly as in policy (10), we look for an optimal policy with $r(C)$ of the form Eq. (11). Here, we set $M'_{\max} = 39$, $C_t = 15$. In our computations we assume that the probability of misidentifying a vacuum state as a photon state and vice versa is $P_e = 0.1$.

VI. RESULTS

In the absence of readout errors, we find that for an optimal policy, $a \ll 1$ and r only weakly depends on C (see Fig. 5). For instance, for the policy $\mathcal{P}$ that minimizes the median of the normalized squared error $2|\alpha_{\text{true}} - \tilde{\alpha}|^2/R_0^2$ after 10^5 measurement shots, we find $a = 0.04$ and $b = 0.05$. Here, the width of the plateau obtained with a low number of measurement shots depends on the degree of the initial parameter uncertainty or the radius R_0. By a crude estimate, one expects that the first detection of a vacuum state takes place after $\frac{\pi R_0^2}{\pi R_\alpha^2} \sim O(10^2)$ measurement shots and that the error is then rapidly reduced to $O(1)$, corresponding to the width of the likelihood function [cf. Fig. 3(c)]. This is consistent with the fact that approximately after 50 measurement shots the plateau shape of the curves crosses over to a rapid decrease. Up to the level where the normalized median squared error reaches the value $\frac{1}{R_0^2}$, the different curves overlap since until this point only the first line in Eq. (10) is executed, and the policy thus does not depend on r. The curve shape following the expected first vacuum detection (rapid decrease of the error) is universal. Interestingly, we find that the curves with different values of r cross, which means that the globally best policy can not be found by local optimization.

With $r = 1$, the boundary of the disk mentioned above coincides with the steepest slope of the likelihood function $P(d|\alpha)$ of Eq. (9). Such a disk contains approxi-
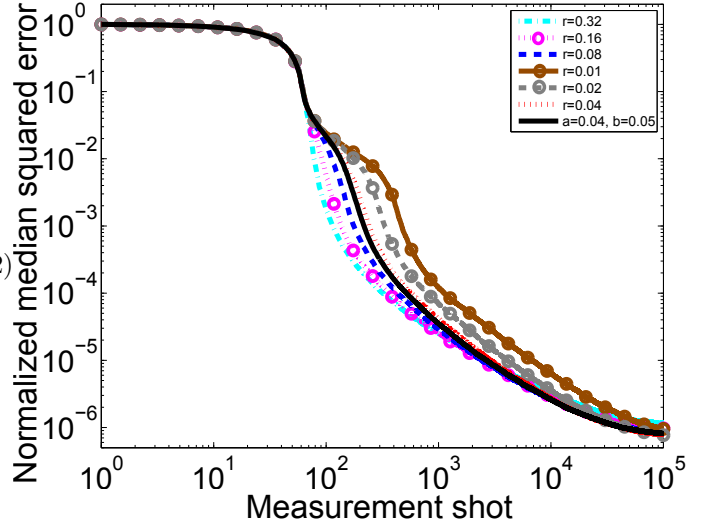

Figure 5: (Color online) Median of the normalized squared error $2|\alpha_{\text{true}} - \tilde{\alpha}|^2/R_0^2$ calculated from an ensemble of 15 000 simulated samples (see text) through the policies of Eq. (10). Different curves are for different radii r denoted in the inset. Black curve is for $r(C)$ calculated through Eq. (11).

mately 39 % of the weight of $P(d = \text{v}|\alpha)$. One might expect that policies attempting to search for the steepest slope of the likelihood function, with $a \approx 1$, would be effective, but this is not the case. The policies with smaller values of a are able to find a crude estimate faster. In the region $x \ll 1$ the likelihood of detecting a photon $P(d = \text{p}|\alpha) \approx x^2$ is quadratically small. Therefore with $a \ll 1$, once a crude estimate has been found, most measurement shots detect vacuum and confirm the estimate. However, since the relative rate of change

$$\frac{1}{P(d = \text{p}|\alpha)} \frac{\partial P(d = \text{p}|\alpha)}{\partial x} = \frac{2x}{e^{x^2} - 1} \approx \frac{2}{x} \quad (13)$$

increases with decreasing x, the rare detections of photons can effectively make a distinction between different possible values of α in the region $x \ll 1$. The relative rate of change above describes how much, due to Bayes' theorem (1), a detection of a photon changes the relative posterior probabilities of two possible values, α and α', when $|\alpha - \alpha'| = |\Delta x|$ is fixed. Putting all together, measurement settings with $x \ll 1$ are, somewhat unexpectedly, more effective than those with larger values of x.

Even though for optimal policies we have $a \ll 1$, the optimal choice is not $r = 0$. Indeed, the policy at $r = 0$ corresponds to choosing $-\beta$ equal to the mean of the posterior, somewhat similarly with a simple, relatively ineffective, policy in the context of bare phase estimation where the control phase is chosen to coincide with the mean of the posterior (see Eq. (6.2) in [31]).

Policies where the measurement strategy is changed after a certain number of measurement shots have been developed for phase [69] and Hamiltonian [48] estimation.

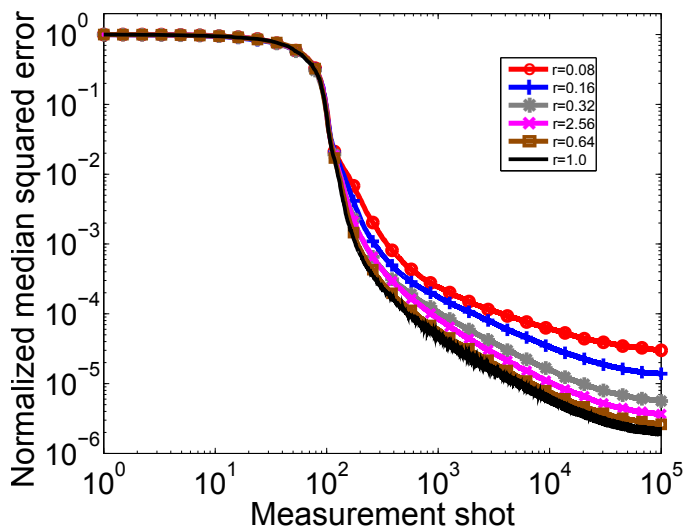to eliminate the outliers with an acceptable overhead.

Figure 6: (Color online) Median of the normalized squared error $2|\alpha_{\text{true}} - \tilde{\alpha}|^2/R_0^2$ calculated from an ensemble of 15 000 simulated samples (see text) through the policies of Eq. (12) when the probability of a readout error is $P_e = 0.1$. Different curves are for different radii r denoted in the inset.

On the second line of Eq. (10), rather than on the number of all the measurement shots, we expect a possible dependence on the number of shots that take place after the first detection of a vacuum state. We therefore count in Eq. (10) the number of shots in which a vacuum state is detected.

In the presence of readout errors, we find that the optimal value for a is larger than in the absence of these errors and the settings are therefore more spread around their mean. The dependence of r on C should still be weak so that $|b| \ll 1$ (see Fig. 6). For the policy $\mathcal{P}_e$ that minimizes the relative median squared error after 10^5 measurement shots, we find $a = 1.0$ and $b = 0$.

For each simulated ensemble we obtain some samples that we refer to as "outliers" for which the error significantly exceeds the median and the width of the posterior probability distribution. Our policies can be made robust against such outliers through repetition as follows. After 10 000 measurement shots we set the prior back to the initial prior. We thereafter perform another 10 000 measurement shots. We then compare the estimates after 10 000 and 20 000 measurement shots. If their difference is smaller than a set threshold, we conclude that we have found a correct estimate, otherwise we start a new search of the estimate. For the new search we choose a prior that again coincides with the original prior. Tables I and II summarize the performance of the policies $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{P}}_e$ in the absence and presence of readout errors, respectively. These correspond to the policies $\mathcal{P}$ and $\mathcal{P}_e$ supplemented with the outlier correction scheme. Outliers are defined as the samples with the squared error $|\tilde{\alpha} - \alpha_{\text{true}}|$ larger than threshold $\mathcal{E}^2$. The outlier correction scheme appears

Table I: Number of outliers per 10 000 simulated samples with $P_e = 0$ for the policy $\tilde{\mathcal{P}}$ (see text). Rows correspond to the number of outliers with normalized squared error larger than $\mathcal{E}^2$ after a given number of measurement shots (indicated by the columns).

$\mathcal{E}^2$/Shots	2×10^4	4×10^4	6×10^4	8×10^4	1.2×10^5	1.4×10^5
10^{-5}	8410	2062	504	118	5	0
10^{-4}	2218	173	3	0	0	0
10^{-3}	207	105	2	0	0	0

Table II: Number of outliers per 10 000 simulated samples with $P_e = 0.1$ for the policy $\tilde{\mathcal{P}}_e$ (see text). Rows correspond to the number of outliers with normalized squared error larger than $\mathcal{E}^2$ after a given number of measurement shots (indicated by the columns).

$\mathcal{E}^2$/Shots	2×10^4	4×10^4	6×10^4	8×10^4	1.4×10^5	2.2×10^5
10^{-5}	9283	3474	1293	496	27	0
10^{-4}	5075	70	1	0	0	0
10^{-3}	825	42	0	0	0	0

VII. DISCUSSION

Based on Bayesian inference, we have delivered powerful methods for adaptive characterization of coherent states. For larger photon numbers $\langle \hat{n} \rangle \gg 1$, the adaptive schemes discussed here reduce the number of measurement shots required to achieve a certain level of accuracy by a factor proportional to the area describing the initial uncertainty of α in phase space. In terms of measurement shots, we expect that efficiency can be quite generally improved by several orders of magnitude which motivates making experiments adaptively. For more complicated quantum states of light in a superposition $|\Psi\rangle = \sum_{n=1}^{N_s} |\alpha_n\rangle$, our results are readily applicable for estimating $\{|\alpha_n\rangle\}_{n=1}^{N_s}$. For full identification of $|\Psi\rangle$, our policies have to be supplemented with methods to obtain relative weights $\{|a_n|^2\}_{n=1}^{N_s}$ of different components as well as their relative phases. Our work thus constitutes a building block that opens up an avenue for efficient estimation of multicomponent Schrödinger-cat states of light.

[1] E. Wigner, Phys. Rev. **40**, 749 (1932).
[2] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).
[3] L. G. Lutterbach and L. Davidovich, Phys. Rev. Lett. **78**, 2547 (1997).
[4] A. I. Lvovsky and S. A. Babichev, Phys. Rev. A **66**, 011801 (2002).
[5] P. Bertet, A. Auffeves, P. Maioli, S. Osnaghi, T. Meunier, M. Brune, J. M. Raimond, and S. Haroche, Phys. Rev. Lett. **89**, 200402 (2002).
[6] A. N. Cleland and M. R. Geller, Phys. Rev. Lett. **93**, 070501 (2004).
[7] J. Bochmann, A. Vainsencher, D. D. Awschalom, and A. N. Cleland, Nature Phys. **9**, 712 (2013).
[8] S. Deléglise, I. Dotsenko, C. Sayrin, J. Bernu, M. Brune, J.-M. Raimond, and S. Haroche, Nature (London) **455**, 510 (2008).
[9] M. Hofheinz *et al.*, Nature (London) **459**, 546 (2009).
[10] G. Kirchmair, B. Vlastakis, Zaki Leghtas, S. E. Nigg, H. Paik, E. Ginossar, M. Mirrahimi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Nature (London) **495**, 205 (2013).
[11] P. K. Day, H. G. DeDuc, B. A. Mazin, A. Vayonakis, and J. Zmuidzinas, Nature (London) **425**, 817 (2003).
[12] R. Bradley *et al.*, Rev. Mod. Phys. **75**, 777 (2003).
[13] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).
[14] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, Phys. Rev. A **80**, 052114 (2009).
[15] G. Waldherr *et al.*, Nat. Nanotechnol. **7**, 105 (2012).
[16] N. M. Nusran, M. U. Momeen, and M. V. G. Dutt, Nat. Nanotechnol. **7**, 109 (2012).
[17] A. J. F. Hayes and D. W. Berry, Phys. Rev. A **89**, 013838 (2014).
[18] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).
[19] A. Abramovici *et al.*, Science **256**, 325 (1992).
[20] K. Goda *et al.*, Nature Phys. **4**, 472 (2008).
[21] M. de Burgh and S. D. Bartlett, Phys. Rev. A **72**, 042301 (2005).
[22] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, (Cambridge University Press, Cambridge, UK, 2000).
[23] D. W. Berry and H. M. Wiseman, Phys. Rev. Lett. **85**, 5098 (2000).
[24] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer, New York, 1998).
[25] A. S. Holevo, Rep. Math. Phys. **16**, 385 (1979).
[26] B. C. Sanders and G. J. Milburn, Phys. Rev. Lett. **75**, 2944 (1995).
[27] B. C. Sanders, G. J. Milburn, and Z. Zhang, J. Mod. Opt. **44**, 1309 (1997).
[28] A. Luis and J. Peřina, Phys. Rev. A **54**, 4564 (1996).
[29] V. Bužek, R. Derka, and S. Massar, Phys. Rev. Lett. **82**, 2207 (1999).
[30] H. Imai and M. Hayashi, New J. Phys. **11**, 043034 (2009).
[31] D. W. Berry, H. M. Wiseman, and J. K. Breslin, Phys. Rev. A **63**, 053804 (2001).
[32] A. Fujiwara, J. Phys. A: Math. Theor. **39**, 12489 (2006).
[33] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. **104**, 063603 (2010).
[34] M. Hayashi, Commun. Math. Phys. **304**, 689 (2011).
[35] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. **107**, 233601 (2011).
[36] B. L. Higgins *et al.*, Nature (London) **450**, 393 (2007).
[37] T. Nagata, R. Okamoto, J. L. O'Brien, K. Sasaki, and S. Takeuchi, Science **316**, 726 (2007).
[38] R. Okamoto, H. F. Hofmann, T. Nagata, J. L. O'Brien, K. Sasaki, and S. Takeuchi, New J. Phys. **10**, 073033 (2008).
[39] J. A. Jones, S. D. Karlen, J. Fitzsimmons, A. Ardavan, S. C. Benjamin, G. A. D. Briggs, and J. J. L. Morton, Science **324**, 1166 (2009).
[40] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).
[41] R. Bloume-Kohout, New J. Phys. **12**, 043034 (2010).
[42] H. M. Wiseman, Phys. Rev. Lett. **75**, 4587 (1995).
[43] M. A. Armen, J. K. Au, J. K. Stockton, A. C. Doherty, and H. Mabuchi, Phys. Rev. Lett. **89**, 133602 (2002).
[44] F. Huszár and N. M. T. Houlsby, Phys. Rev. A **85**, 052120 (2012).
[45] K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. T. Houlsby, F. Huszár, and S. P. Kulik, Phys. Rev. A **87**, 062122 (2013).
[46] S. G. Schirmer and D. K. L. Oi, Phys. Rev. A **80**, 022333 (2009).
[47] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, Phys. Rev. A **84**, 052315 (2011).
[48] C. Ferrie, C. E. Granade, and D. G. Cory, Quant. Inf. Proc. **12**, 611 (2013).
[49] N. Wiebe, C. E. Granade, C. Ferrie, and D. G. Cory, Phys. Rev. Lett. **112**, 190501 (2014).
[50] N. Wiebe, C. E. Granade, C. Ferrie, and D. G. Cory, Phys. Rev. A **89**, 042314 (2014).
[51] M. P. V. Stenberg, Y. R. Sanders, and F. K. Wilhelm, Phys. Rev. Lett. **113**, 210404 (2014).
[52] M. West, J. Roy Stat. Soc. B **55**, 409 (1993).
[53] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, Radar and Signal Processing IEE Proc. F **140**, 107 (1993).
[54] J. Liu and M. West, in *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. Freitas, and N. Gordon (Springer, New York, 2001).
[55] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, New J. Phys. **14**, 103013 (2012).
[56] C. Ferrie, Phys. Rev. Lett. **113**, 190404 (2014).
[57] J. Sperling, W. Vogel, and G. S. Agarwal, Phys. Rev. A **85**, 023820 (2012).
[58] L. C. G. Govia, E. J. Pritchett, S. T. Merkel, D. Pineau, and F. K. Wilhelm, Phys. Rev. A **86**, 032311 (2012).
[59] L. C. G. Govia, E. J. Pritchett, and F. K. Wilhelm, New J. Phys. **16**, 045011 (2014).
[60] D. K. L. Oi, V. Potoček, and J. Jeffers, Phys. Rev. Lett. **110**, 210504 (2013).
[61] Y.-F. Chen, D. Hover, S. Sendelbach, L. Maurer, S. T. Merkel, E. J. Pritchett, F. K. Wilhelm, and R. McDermott, Phys. Rev. Lett. **107**, 217401 (2011).
[62] A. Poudel, R. McDermott, and M. G. Vavilov, Phys. Rev. B **86**, 174506 (2012).
[63] B. Peropadre, G. Romero, G. Johansson, C. M. Wilson, E. Solano, and J. J. García-Ripoll, Phys. Rev. A **84**, 063834 (2011).
[64] A. D. O'Connell *et al.*, Nature (London) **464**, 697 (2010).

[65] S. T. Merkel and F. K. Wilhelm, New J. Phys. **12**, 093036 (2010).

[66] K. Banaszek and K. Wódkiewicz, Phys. Rev. Lett. **76**, 4344 (1996).

[67] K. Banaszek, C. Radzewicz, K. Wódkiewicz, and J. S. Krasiński, Phys. Rev. A **60**, 674 (1999).

[68] It is unlikely but possible that vacuum is not detected even though a large number, M_0, of measurement shots is carried out. Due to limited numerical accuracy, this can lead to a drastic failure of the estimation if the particles do not converge close to the true value. Therefore, in case vacuum has not been detected after M_0 measurement shots, we switch the first line of Eq. (2) to a uniform probability distribution $P_\beta = \frac{1}{\pi R_0^2}$ on the region $\beta < R_0$. We set $M_0 = 600$ and $R_0 = 10$.

[69] M. W. Mitchell, Proc. SPIE **5893**, 589310 (2005).