Markov chain Monte Carlo estimation of quantum states

James DiGuglielmo,[1,*] Chris Messenger,[1] Jaromír Fiurášek,[2] Boris Hage,[1] Aiko Samblowski,[1]
Tabea Schmidt,[1] and Roman Schnabel[1]

[1]*Institut für Gravitationsphysik, Leibniz Universität Hannover and Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institute),*
Callinstrasse 38, 30167 Hannover, Germany
[2]*Department of Optics, Palacky University, 17. listopadu 50, 77200 Olomouc, Czech Republic*

We apply a Bayesian data analysis scheme known as the Markov chain Monte Carlo to the tomographic reconstruction of quantum states. This method yields a vector, known as the Markov chain, which contains the full statistical information concerning all reconstruction parameters including their statistical correlations with no *a priori* assumptions as to the form of the distribution from which it has been obtained. From this vector we can derive, e.g., the marginal distributions and uncertainties of all model parameters, and also of other quantities such as the purity of the reconstructed state. We demonstrate the utility of this scheme by reconstructing the Wigner function of phase-diffused squeezed states. These states possess non-Gaussian statistics and therefore represent a nontrivial case of tomographic reconstruction. We compare our results to those obtained through pure maximum-likelihood and Fisher information approaches.

PACS number(s): 03.65.Wj, 42.50.Dv, 03.67.−a

I. INTRODUCTION

The tomographic reconstruction of quantum states represents an important laboratory tool in both quantum optics and quantum information alike. It can be applied, for example, to the reconstruction of the dynamical interaction between quantum systems in a technique known as *quantum process tomography*. This latter technique is important in quantum computing, where the characterization of quantum gates is essential to the overall quantum circuit [1–3]. Tomography has even been used as part of the optimization of advanced interferometry as was done in the preparation of frequency dependent squeezing [4].

Since the theoretical discovery by Vogel and Risken [5] that the Wigner function can be reconstructed from homodyne detector data, a number of reconstruction schemes have been developed ranging from direct inversion of the tomographic data by means of the filtered-back projection method [6] to statistical methods such as maximum-likelihood estimation [7–12]. An important feature of maximum-likelihood methods is the guaranteed positive "semidefiniteness" of the reconstructed state. The result of a maximum-likelihood reconstruction method is either a density matrix [10,11] or a set of parameters [13,14] which have maximized the likelihood functional given a model of the measurement apparatus and of the parametrized state. A full analysis of the experimental data, however, should also answer important questions regarding error bars on the estimation of the state parameters, possible correlations among the parameters, and error propagation when using the reconstructed state for further calculations of quantities such as the purity of the state or amount of entanglement.

Several methods have been proposed in the literature to put error bars on the reconstructed quantum states. For linear reconstruction techniques based on the averaging of sampling functions over the experimental data [15], one can calculate the statistical uncertainties of the estimated quantities by evaluating variances of the linear estimators [16]. Another possible approach is to numerically simulate the whole measurement and reconstruction process many times assuming that the reconstructed state is the true state that is being measured upon. The error bars are then calculated from the resulting ensemble of reconstructed states [17]. Finally, uncertainties on estimates obtained by maximum-likelihood method can be determined by evaluating the Fisher information matrix [18]. This latter approach essentially relies on approximation of the likelihood function by a Gaussian and becomes exact only asymptotically in the limit of a very large number of experimental data.

In the present paper we show that the uncertainties in quantum state estimation can be consistently determined by using a general and statistically well motivated Bayesian analysis scheme known as the Markov chain Monte Carlo (MCMC). The method is based on the implementation of a Markov chain to search the parameter space resulting in a set of *samples from the joint posterior probability density distribution* on the unknown parameters of the model. This technique produces several important results. First, it yields the Markov chain containing all of the relevant statistical information about the parameter space. Second, one can extract a set of marginalized probability density distributions for each parameter quantifying the degree of uncertainty on their estimation. Third, the resulting chain can be used in further calculations, where one can produce probability density distributions on quantities such as of the purity or amount of entanglement of the reconstructed state.

This paper is divided into the following sections: in Sec. II the necessary concepts required from Bayesian data analysis are introduced and applied to the case of quantum state estimation. In Sec. III, the quantum likelihood function for phase-diffused squeezed states is derived. In Sec. IV the Markov chain Monte Carlo algorithm is introduced, and finally in Sec. V both the technical details of the experimental

―――――――
*james.diguglielmo@aei.mpg.de

realization as well as the results of the reconstruction are presented.

II. BAYESIAN DATA ANALYSIS

Bayes' theorem prescribes the rule to invert the relationship between the experimental data already observed and the parametrized model which could have generated the measured data set. The theorem reads

$$p(\vec{\lambda}|D,I) = \frac{p(D|\vec{\lambda},I)p(\vec{\lambda}|I)}{p(D|I)}, \tag{1}$$

where we use D to represent our data, I to represent our prior information or model of the experiment, and $\vec{\lambda}$ to represent a vector of model parameters. The quantity $p(\vec{\lambda}|D,I)$ is known as the *posterior distribution*, $p(D|\vec{\lambda},I)$ is known as the *likelihood function*, $p(\vec{\lambda}|I)$ is the *prior distribution*, and $p(D|I)$ is a normalization factor. The standard application of Bayesian analysis is to calculate the posterior distribution given a parametrized model of the sought after signal, the measured data, and prior probability distributions on the values of the model parameters. These three elements are brought together through Eq. (1).

Let us now construct the likelihood function for the quantum estimation problem. A general measurement on a quantum system can be described by the so-called positive operator-valued measure (POVM). Each possible measurement outcome j is associated with a POVM element Π_j which is a positive semidefinite operator. The probability of outcome j can be calculated as $P_j = \mathrm{Tr}[\Pi_j \rho(\vec{\lambda})]$, where $\rho(\vec{\lambda})$ denotes the density matrix of the measured quantum system that depends on the model parameters $\vec{\lambda}$. Since the total probability of some outcome is one, the POVM elements sum up to identity the operator, $\Sigma_j \Pi_j = \mathbb{1}$. This generic framework in particular encompasses a tomographic reconstruction of the state $\rho(\vec{\lambda})$ that consists of several different measurements M with possible outcomes indexed by l_M. Then $j=(M,l_M)$ becomes a multi-index indicating both the measurement setting and the measurement outcome for a given setting. Let n_j denote the observed number of measurement outcome j and $N = \Sigma_j n_j$ represents the total amount of collected data. The likelihood function $\mathcal{L}$ is the probability of observation of a particular set $\{n_j\}$ for a given $\vec{\lambda}$. It follows that $\mathcal{L}$ is given by a multinomial distribution and reads

$$\mathcal{L} = N! \prod_j^N \frac{P_j^{n_j}}{n_j!}. \tag{2}$$

In terms of the constituents of Bayes' theorem, the theoretical probabilities $\{P_j\}$ are functions of the parameters which are to be determined. The measured numbers of counts $\{n_j\}$ correspond to the data.

III. LIKELIHOOD FUNCTION FOR PHASE-DIFFUSED SQUEEZED STATES

The phase-diffused squeezed states were first introduced within the context of continuous-variable squeezing purifica-

tion by [19–21]. They arise when squeezed states are transmitted over dephasing quantum channels such as optical fibers affected by thermal fluctuations. These states are characterized by non-Gaussian statistics and therefore represent a nontrivial case for quantum state tomography.

The Wigner function for phase-diffused squeezed state is given by

$$W(x,p) = \frac{1}{2\pi\sqrt{V_x V_p}} \int_{-\infty}^{\infty} \exp\left[-\left(\frac{x_\phi^2}{2V_x} + \frac{p_\phi^2}{2V_p} \right) \right] \Phi(\phi) d\phi, \tag{3}$$

where $x_\phi = x\cos\phi + p\sin\phi$, $p_\phi = p\cos\phi - x\sin\phi$ with x and p as the standard position and phase quadratures, and ϕ represent the random-phase shifts distributed according to some probability distribution $\Phi(\phi)$. We use V_x to represent the variance of the squeezed quadrature and V_p for the variance of the antisqueezed quadrature. We normalize the variances such that for vacuum state we have $V_x = V_p = 1$ and the state is squeezed in x quadrature if $V_x < 1$. In the experiment we measure several different rotated quadratures x_θ, where θ defines a specific measurement setting. The theoretical homodyne probability density distribution $p(x_\theta)$ can be calculated from Wigner function as a marginal distribution. Integration of $W(x,p)$ over the conjugate quadrature p_θ yields, after some algebra

$$p(x_\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\widetilde{V}(\phi)}} \exp\left[-\frac{x_\theta^2}{2\widetilde{V}(\phi)} \right] \Phi(\phi - \theta) d\phi, \tag{4}$$

where $\widetilde{V}(\phi) = V_x\cos^2\phi + V_p\sin^2\phi$.

The data from each measurement is binned into L bins whose lower boundaries are defined by $Q_{\theta,l}$. The outer bins extend to infinity, and we set $Q_{\theta,1} = -\infty$ and $Q_{\theta,L+1} = \infty$. The corresponding theoretical probability $P_{\theta,l}$ is given by integration of probability density (4) over the bin,

$$P_{\theta,l}(\vec{\lambda}) = \int_{Q_{\theta,l}}^{Q_{\theta,l+1}} p(x_\theta) dx_\theta. \tag{5}$$

In our experiment, the results from two quadrature measurements were formed into histograms each containing a total of $L = 70$ bins. From the perspective of direct data inversion, this corresponds to an overdetermined system because we need to estimate only three real parameters, cf. below.

The POVM elements describing such binned homodyne detection can be expressed as

$$\Pi_{\theta,l} = \int_{Q_{\theta,l}}^{Q_{\theta,l+1}} |x_\theta\rangle\langle x_\theta| dx_\theta, \tag{6}$$

where $|x_\theta\rangle$ is an eigenstate of quadrature operator x_θ. Note that, by definition, the sum of theoretical probabilities over all bins is equal to one,

$$\sum_l P_{\theta,l}(\vec{\lambda}) = 1. \tag{7}$$

This is a mathematical expression of the fact that the homodyne detection always yields some outcome and, after each measurement, one of $n_{\theta,l}$ is increased by one. Put in a different way, the homodyne detection is described by a complete POVM (6) whose elements $\Pi_{\theta,l}$ satisfy the condition $\Sigma_l \Pi_{\theta,l} = \mathbb{1}$.

Assuming the phase noise distribution, $\Phi(\phi)$, is a zero mean Gaussian, the state can be completely characterized by just three parameters $\vec{\lambda} = \{V_x, V_p, V_\phi\}$ where V_ϕ is the variance of the random-phase shifts. The quantum log-likelihood function for the phase-diffused squeezed states is finally obtained by taking the natural logarithm of Eq. (2) giving us

$$\Lambda = \sum_{\theta,l} n_{\theta,l} \ln[P_{\theta,l}(\vec{\lambda})], \tag{8}$$

where, for simplicity, we ignore all terms that do not depend on the parameter values.

IV. MARKOV CHAIN MONTE CARLO ALGORITHM

The goal of our Bayesian reconstruction scheme is the calculation of marginalized posterior distributions on our model parameters $\vec{\lambda} = \{V_x, V_p, V_\phi\}$. To this end, the MCMC method can be used to generate samples drawn from the posterior distribution $p(\vec{\lambda}|D, I)$. Since this distribution is unknown one cannot directly sample from it and instead we sample from the distribution that is the product of the likelihood and the prior. The prior is chosen by considering the possible values of the parameters to be determined. Since the parameters to be determined in this case are variances, their values must be greater than zero. In order to assume relative ignorance in the value the parameters could take, we use a prior which only requires the variances to be positive and satisfy the Heisenberg uncertainty relation, $V_x V_p \geq 1$. Since the likelihood in this analysis is a sharply peaked function, the choice of uniform priors has negligible effect on the numerical results of the MCMC [22]. In this case Bayes' theorem, Eq. (1), tells us that the likelihood function is proportional to the posterior distribution and, since this is a function that we can compute for a given parameter space location, we can use a standard sampling algorithm such as the Metropolis-Hastings sampler to draw from it. We describe our implementation of this sampler in the Appendix.

V. EXPERIMENTAL IMPLEMENTATION

A. Description of the experiment

Figure 1 shows the experimental setup that was used to prepare the phase-diffused squeezed states. The full details of the setup are provided in [20] and will be summarized here. The squeezing source was an optical parametric amplifier (OPA) constructed from a type I noncritically phase-matched $MgO:LiNbO_3$ crystal inside a standing-wave resonator, similar to the design that previously has been used in [23]. The OPA was pumped with 50 mW of green light at 532 nm
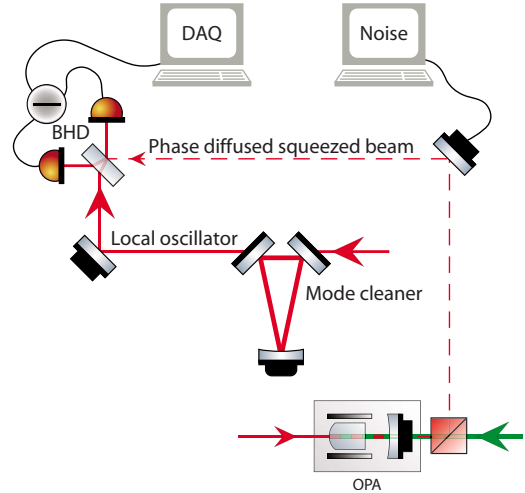

FIG. 1. (Color online) Experimental setup: the experimental setup consists of an OPA for the generation of the squeezed vacuum states, a phase shifter to induce the random-phase noise, and a homodyne detector to measure the prepared state. The mode cleaner was used to increase the fringe contrast at the balanced homodyne detector.

resulting in a classical gain of about 11. The length of the OPA cavity as well as the phase of the second-harmonic pump beam were controlled using radio-frequency modulation/demodulation techniques. The mode cleaner was operated in high finesse mode $\mathcal{F} = 10\,500$ resulting in a linewidth of 55 kHz. A nonclassical noise power reduction of slightly more than 5.0 dB was directly observed with a homodyne detector in combination with a spectrum analyzer at a Fourier sideband frequency of 6.4 MHz.

The phase noise was induced by reflecting the squeezed field from a piezoelectric transducer (PZT) mounted high-reflection mirror that was quasirandomly moved. The voltages applied to the PZTs were produced as follows. An independent random number generator produced data strings with a Gaussian distribution. The strings were digitally filtered to limit the frequency band to 2–2.5 kHz. The output interface was a common PC sound card with SNR of -110 dB. The sound volume was set to meet the desired standard deviation of channel phase noise.

Homodyne detection confirmed that the squeezing degraded in the same way when phase noise was increased. The detector difference currents were electronically mixed with a 6.4 MHz local oscillator. The demodulated signals were then filtered with a 400 kHz bandwidth low-pass filter, and sampled with 10^6 samples per second and 14 bit resolution using a National Instruments analog-digital sampling card.

B. Results of the MCMC

Figure 2 depicts the resulting chains after 40 000 iterations of the MCMC algorithm. The abscissa represents the number of iterations of the MCMC whereas the ordinate represents the parameter values. After an initial "burn-in" period of approximately 1000 iterations, in which the chain heads
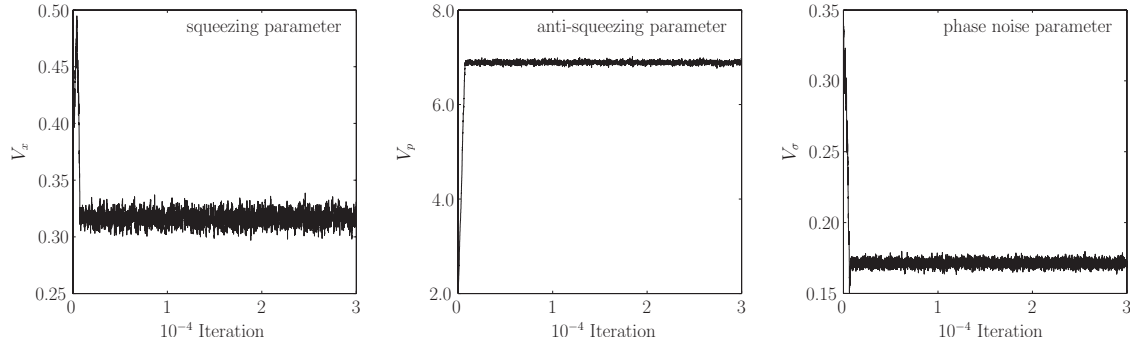
FIG. 2. Markov chains: this figure depicts the evolution of the Markov chains where the abscissa represents the iteration number and the ordinate represents the value of the chain. After an initial burn-in period, in which the chains head to their steady-state positions, the chains eventually converge to a region of parameter space and begin to sample from the posterior distribution. The proposal distributions were taken to be Gaussian with the standard deviations $\Sigma_x=0.0042$, $\Sigma_p=0.022$, and $\Sigma_\phi=0.0037$ corresponding to the squeezing parameter, the anti-squeezing parameter, and the phase noise parameter, respectively. The starting values were randomly chosen with the only constraint that they be non-negative and obey $V_x V_p \geq 1$. The chains settled to their equilibrium positions with means of $\mu_x=0.316$, $\mu_p=6.888$, and $\mu_\phi =0.171$ corresponding to the squeezing, antisqueezing and phase noise parameter, respectively.

toward equilibrium, the chain converges and begins to sample from the posterior distribution (which in this Gaussian case also includes the region of maximum-likelihood). The development of criteria for the determination of chain convergence is a general problem which has been the subject of much research [24,25]. The general idea is to run multiple chains per estimation parameter, and monitor their evolution both within each change and across each change. Convergence is inferred if all chains behave consistently. With respect to the case at hand, convergence of the chain can be inferred by comparing the locations to which the marginalized parameter chains have settled with the independent measurement of the V_x and V_p parameter values performed with a spectrum analyzer. In the absence of such an independent measurement, the criteria in [24,25] can be used to infer convergence.

The width of the marginalized chains, i.e., their standard deviations, quantify the degree of uncertainty on the value of each parameter. By forming histograms of the chain as a function of each of the parameters, we obtain their marginalized posterior probability distributions as shown in Fig. 3. From these posteriors we obtain the following uncertainties on the model parameters: $\sigma_x=0.0056$ for the squeezing parameter, $\sigma_p=0.0289$ for the antisqueezing parameter, and fi-

nally $\sigma_\phi=0.0020$ for the phase noise parameter. The proposal distributions, from which the posterior distribution samples have been drawn, were chosen to be Gaussians with the following standard deviations $\Sigma_x=0.0042$, $\Sigma_p=0.022$, and $\Sigma_\phi =0.0037$ corresponding to the squeezing parameter, anti-squeezing parameter, and phase noise parameter, respectively. These values were obtained through manual tuning of the MCMC algorithm. This is done by adjusting the individual standard deviations, i.e., Σ_x, Σ_p, and Σ_ϕ, until the proportion of accepted jumps reaches approximately 44% [26].

As an independent test of the posterior standard deviations, we also calculated the Fisher information matrix [27,28] given by

$$\mathcal{F}_{ij} = \left\langle \frac{\partial \Lambda}{\partial \lambda^i} \frac{\partial \Lambda}{\partial \lambda^j} \right\rangle. \tag{9}$$

The inverse Fisher matrix represents the covariance of the posterior probability distribution for the true parameters $\vec{\lambda}$ as inferred from a single experiment assuming Gaussian noise and constant priors over the parameter range of interest. The calculated standard deviations from the Fisher matrix are presented in Fig. 3 as well as in Table I where very good
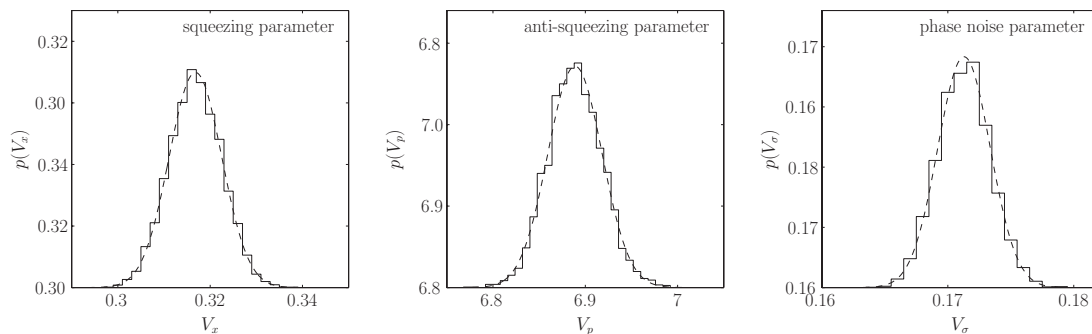

FIG. 3. Marginalized posterior distributions: each probability density was calculated using every tenth point from the marginalized chains corresponding to 38 000 data points. The standard deviation of each density are $\sigma_x=0.0056$, $\sigma_p=0.0289$, and $\sigma_\phi=0.0020$ corresponding to the squeezing parameter, the antisqueezing parameter, and the phase noise parameter, respectively. The dotted line represents the result of the Fisher analysis.

TABLE I. Standard deviations of posterior distributions: this table compares the standard deviations of the parameter posterior distributions obtained from the Markov chain Monte Carlo method and from the Fisher information matrix. The standard deviations represent the error on the estimation of the parameter values. The Fisher matrix returns the actual standard deviations only in the case of Gaussian noise.

	Parameter estimates		Uncertainties	
Parameter	MCMC	Maximum likelihood	MCMC	Fisher
V_x	0.316	0.317	0.0056	0.0055
V_p	6.889	6.880	0.0289	0.0294
V_ϕ	0.171	0.171	0.0020	0.0020

agreement is readily seen. It should be stressed that the posterior distribution standard deviations obtained from the Fisher matrix are only valid when assuming Gaussian noise whereas the posterior distribution standard deviations obtained from the Markov chain is valid regardless of the form of the posterior distribution. The results of the MCMC as well as of the Fisher analysis are compared in Table I.

Figure 4 depicts the evolution of the log-likelihood function [Eq. (8)] for each iteration of the MCMC. It is seen that as the chains evolve through the burn-in stage the log-likelihood quickly increases. After approximately 1000 iterations it has reached equilibrium whereupon parameter space jumps to higher likelihood values are balanced by jumps to lower likelihood values.

Figure 5 represents the spectrum of the squeezed state before the phase diffusion. The measured state is seen to have a squeezing strength of −4.98 dB or a variance of V_x =0.31 and an antisqueezing strength of 8.39 dB or a variance of V_p=6.91. These values have not been corrected for dark noise and were measured using a video bandwidth (VBW) of 10.0 Hz, a resolution bandwidth (RBW) of 100 kHz, and a sweep time (SWT) of 1.5 s. These values lie within the width of the respective posterior distributions obtained from the
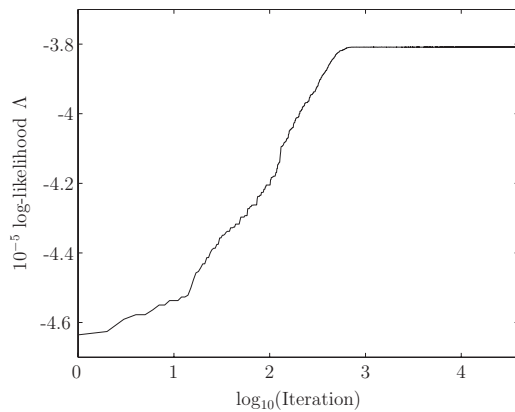

FIG. 4. Log likelihood: evolution of the log likelihood as a function of iteration. The abscissa represents the iteration number and the ordinate the log-likelihood value. After approximately 1000 iterations the log-likelihood appears to have attained its maximum value at which point the chains have reached equilibrium and are sampling from the posterior distribution.
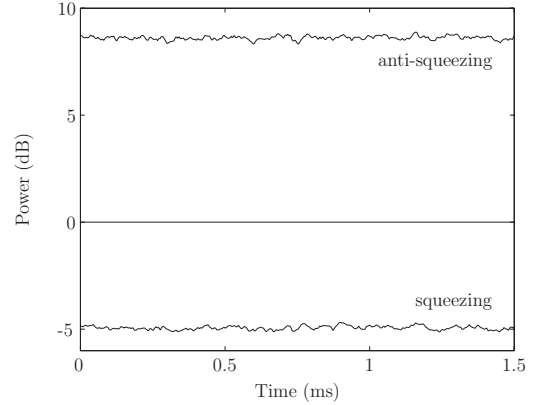

FIG. 5. Zero span measurement of squeezing and antisqueezing: the recorded amount of squeezing and antisqueezing as measured by a balanced homodyne detector and spectrum analyzer at a Fourier frequency of 6.5 MHz. It is seen that approximately −4.98 dB of squeezing and 8.39 dB of antisqueezing were directly measured without dark noise correction. These correspond to variances of V_x=0.31 and V_p=6.91, respectively. The units of the vacuum have been set to 1 corresponding to 0 dB.

MCMC analysis. Furthermore, only *two* quadrature measurements, each containing just 100 000 data samples, were required to obtain these results. This represents a significant savings in terms of experimental effort to reconstruct a non-Gaussian state.

We also compare our results with that of a pure maximum-likelihood approach. This was achieved by altering step 3(d) of the Metropolis-Hastings sampler (see Appendix) such that the chain remains at its current position whenever the likelihood ratio r is less than one. This forces the chain to only move to regions of higher likelihood and hence quickly converge on the true maximum-likelihood location. The parameter values obtained at the maximum-likelihood were: $\hat{V}_x$=0.317 for the squeezing parameter, $\hat{V}_p$=6.880 for the antisqueezing parameter, and $\hat{V}_\phi$=0.171 for the phase noise parameter. These are, however, the only results obtained from the maximum-likelihood estimation alone. Since the MCMC chain contains a complete statistical description of the parameters, the statistical error on the reconstruction of both the quantum state itself as well as derived quantities from it, e.g., purity, can be exactly determined. This will be illustrated in the next section.

C. Reconstruction of the quantum state

Having completely characterized the parameter space, the final state can be reconstructed. In order to generate the Wigner function shown in Fig. 6, the Markov chain together with Eq. (3) were used to calculate the average Wigner function. The characteristic non-Gaussian statistics of the phase-space Wigner distribution is clearly manifested in the reconstructed state. It should be stressed that this averaging already takes into account the standard deviation of each value of the Wigner function since the parameter values are taken from the posterior distribution.
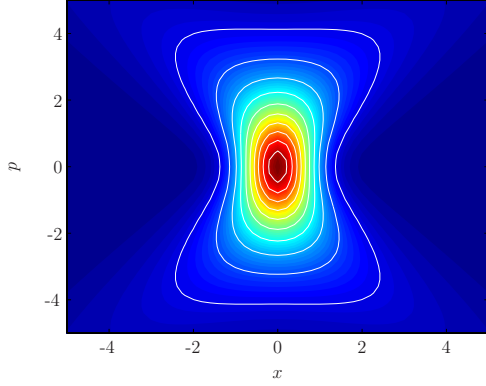
FIG. 6. (Color online) Reconstructed Wigner function: the Wigner function of the measured state as reconstructed using Eq. (3) and the Markov chain. This reconstruction contains all the relevant statistical information.

In addition to generating a phase-space plot of the reconstructed quantum state, the Markov chain can be used in further calculations of such properties as the purity of the reconstructed state. Figure 7 depicts such a result. Using the analytical definition of the purity

$$\mu = 4\pi \int \int W^2(x,p)dxdp, \tag{10}$$

and the resulting chain from the MCMC, the purity can be calculated, automatically taking into consideration the statistical error and correlation on the parameters determined by the MCMC. The result is a probability density whose standard deviation quantifies the degree of uncertainty on the estimation of the purity. For the state in question, we obtain a purity of $\mu = 0.5649 \pm 0.0028$. It is important to note that this information is delivered directly from the MCMC itself; no additional assumptions as to the distribution of the errors and their correlation properties need to be made. Additionally, any one-dimensional quantity can be calculated in this

manner. For example, if estimating the amount of entanglement of a non-Gaussian state, the logarithmic negativity [29] can be calculated over the span of the resulting chain. The result will be a probability density quantifying the uncertainty on its value.

VI. CONCLUSION

We have applied a Bayesian data analysis scheme known as MCMC to the tomographic reconstruction of quantum states. Taking phase-diffused squeezed states as an example, we have provided the details as to the derivation of the likelihood function as well as to the numerical implementation of the MCMC. The results include a set of probability density distributions which exactly quantify the degree of uncertainty on the estimation of the parameters. These results were compared to both a pure maximum likelihood and a Fisher information approach. Furthermore, using the Markov chain in the calculation of the state's purity enabled the construction of a probability density distribution on the value of the purity, thereby quantifying the degree of uncertainty on its calculation.

We note that MCMC scheme is completely general and can be applied to higher dimensional problems, such as the reconstruction of the density matrix, and will be the topic of future publications.
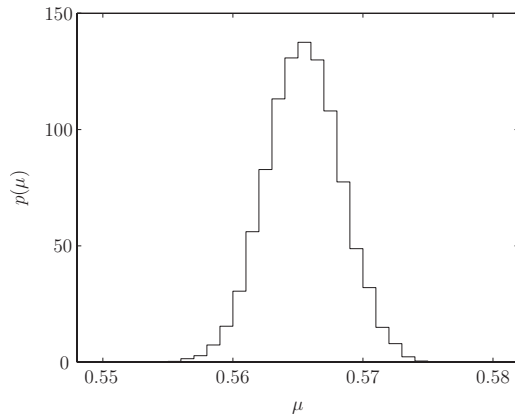
FIG. 7. Reconstructed purity posterior distribution: The purity was calculated using the resulting Markov chain of 38 000 parameter values and Eq. (10). This results in a probability density function which exactly quantifies the uncertainty on the estimation of the purity. The purity is calculated to be $\mu = 0.5649 \pm 0.0028$. No such result is possible with a pure maximum-likelihood approach.
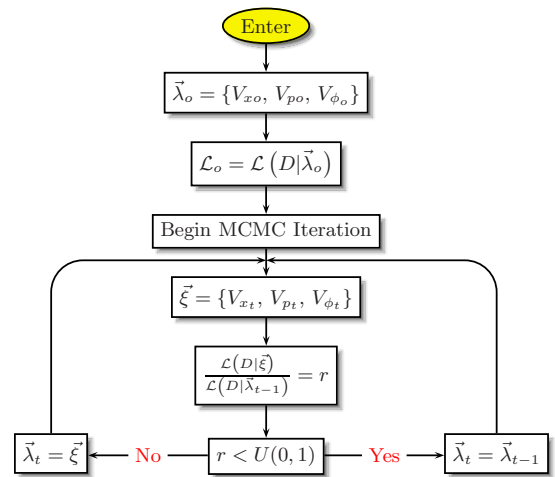

FIG. 8. (Color online) MCMC program flow chart: a schematic explanation of the process of selecting MCMC samples from the posterior distribution. The lower loop iterates until the chain has converged to its equilibrium position and begins to sample from this posterior distribution.

APPENDIX: METROPOLIS-HASTINGS SAMPLER

Our implementation of the MCMC is based on the Metropolis-Hastings sampler. It is defined by performing the following steps (see also Fig. 8):

(1) Generate initial values for parameters, $\vec{\lambda}_o$.

(2) Compute the quantity $\mathcal{L}_0 = \mathcal{L}(D|\vec{\lambda}_0)$.

(3) Iterate the following over the index t until the chain has converged:

(a) Generate a trial parameter vector $\vec{\xi}$ according to a proposal distribution $q(\vec{\lambda}_{t-1})$.

(b) Compute the quantity $\mathcal{L}_t = \mathcal{L}(D|\vec{\xi})$. If $V_x V_p < 1$ or some $V_j < 0$ then set $\mathcal{L}_t = 0$.

(c) Compute the ratio $r = \mathcal{L}_t / \mathcal{L}_{t-1}$.

(d) Sample from a uniform distribution, $U(0,1)$.

$$\text{if } \begin{cases} r > U, & \text{set } \vec{\lambda}_t = \vec{\xi}, \ \mathcal{L}_t = \mathcal{L} & \text{accepted.} \\ r < U, & \text{set } \vec{\lambda}_t = \vec{\lambda}_{t-1}, \ \mathcal{L}_t = \mathcal{L}_{t-1} & \text{rejected.} \end{cases}$$

The proposal distribution $q(\vec{\lambda})$ used in stage 3(a) is used to select trial parameter values within the MCMC. Theoretically q can be any distribution, however in practice it is sensible for the proposal distribution to suggest jumps that are local to the current location but large enough to allow an efficient exploration of the parameter space [30]. Stage 3(b) ensures that the sampling is restricted to subspace of physically admissible values of parameters $\vec{\lambda}$. The simple difference between this MCMC method compared to that of a pure maximum-likelihood method is that for repeated stages within the Metropolis-Hastings sampler the chain has finite probability of jumping *both* to higher or lower values of likelihood.

[1] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).

[2] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997).

[3] B. P. Lanyon, T. J. Weinhold, N. K. Langford, M. Barbieri, D. F. V. James, A. Gilchrist, and A. G. White, Phys. Rev. Lett. **99**, 250505 (2007).

[4] S. Chelkowski, H. Vahlbruch, B. Hage, A. Franzen, N. Lastzka, K. Danzmann, and R. Schnabel, Phys. Rev. A **71**, 013806 (2005).

[5] K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).

[6] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).

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

[8] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, Phys. Rev. A **61**, 010304(R) (1999).

[9] J. Fiurášek, Phys. Rev. A **64**, 024102 (2001).

[10] A. I. Lvovsky, J. Opt. B: Quantum Semiclassical Opt. **6**, S556 (2004).

[11] Z. Hradil, D. Mogilevtsev, and J. Řeháček, Phys. Rev. Lett. **96**, 230401 (2006).

[12] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky, Phys. Rev. A **75**, 042108 (2007).

[13] G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, Phys. Rev. A **62**, 023815 (2000).

[14] A. Ourjoumtsev, R. Tualle-Brouri, and P. Grangier, Phys. Rev. Lett. **96**, 213601 (2006).

[15] D. G. Welsch, W. Vogel, and T. Opatrný, in *Progress in Optics*, edited by E. Wolf (Elsevier, Amsterdam, 1999), Vol. 39.

[16] U. Leonhardt, M. Munroe, T. Kiss, T. Richter, and M. G. Raymer, Opt. Commun. **127**, 144 (1996).

[17] G. M. D'Ariano, M. F. Sacchi, and P. Kumar, Phys. Rev. A **61**, 013806 (1999).

[18] J. Řeháček, D. Mogilevtsev, and Z. Hradil, New J. Phys. **10**, 043022 (2008).

[19] J. Fiurášek, P. Marek, R. Filip, and R. Schnabel, Phys. Rev. A **75**, 050302(R) (2007).

[20] A. Franzen, B. Hage, J. DiGuglielmo, J. Fiurášek, and R. Schnabel, Phys. Rev. Lett. **97**, 150505 (2006).

[21] B. Hage, A. Franzen, J. DiGuglielmo, P. Marek, J. Fiurášek, and R. Schnabel, New J. Phys. **9**, 227 (2007).

[22] Since the parameters are variances, the actual priors should be determined by using Jeffrey's Principle of Invariance. However, as stated, the implementation of flat priors on $\vec{\lambda}$ in this case has negligible effect.

[23] S. Chelkowski, H. Vahlbruch, K. Danzmann, and R. Schnabel, Phys. Rev. A **75**, 043814 (2007).

[24] S. P. Brooks and A. Gelman, J. Comput. Graph. Stat. **7**, 434 (1998).

[25] S. El Adlouni, A.-C. Favre, and B. Bobée, Comput. Stat. Data Anal. **50**, 2685 (2006).

[26] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. (Chapman and Hall/CRC, London, 2004).

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

[28] M. Vallisneri, Phys. Rev. D **77**, 042001 (2008).

[29] G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002).

[30] There are many techniques specific to the efficient exploration of parameter space using MCMCs such as simulated annealing, parallel tempering, and the Fisher information. For the purposes of this analysis none of these techniques were necessary and the proposal distribution used was an uncorrelated multivariate Gaussian with hand-tuned variances.