

Characterizing entanglement sources

Pavel Lougovski¹ and S. J. van Enk^{1,2}

¹*Department of Physics and Oregon Center for Optics, University of Oregon, Eugene, Oregon 97403, USA*

²*Institute for Quantum Information, California Institute of Technology, Pasadena, California 91125, USA*

(Received 3 August 2009; published 18 November 2009)

We discuss how to characterize entanglement sources with finite sets of measurements. The measurements do not have to be tomographically complete and may consist of POVMs rather than von Neumann measurements. Our method yields a probability that the source generates an entangled state as well as estimates of any desired calculable entanglement measures, including their error bars. We apply two criteria, namely, Akaike's information criterion and the Bayesian information criterion, to compare and assess different models (with different numbers of parameters) describing entanglement-generating devices. We discuss differences between standard entanglement-verification methods and our present method of characterizing an entanglement source.

DOI: [10.1103/PhysRevA.80.052324](https://doi.org/10.1103/PhysRevA.80.052324)

PACS number(s): 03.67.Mn, 02.50.Cw

I. INTRODUCTION

Entanglement is useful but hard to generate and even harder to detect. The most measurement-intensive approach to the problem of experimentally detecting the presence of entanglement is to perform complete quantum-state tomography [1]. Even for just two qubits, this implies a reconstruction of all 15 independent elements of the corresponding density matrix. Subsequently applying the positive partial transpose (PPT) criterion to the reconstructed matrix gives a conclusive answer about entanglement or separability of the state [2,3].

From the practical point of view, it is desirable to have an entanglement detection tool that is more economical than full state tomography but nevertheless is decisive. Already in the original work on PPT [2], it was noticed that one can always construct an observable \mathcal{W} with nonnegative expectation values for all separable states ρ_s and a negative expectation value for at least one entangled state ρ_e . In this way, an experimentally detected violation of the inequality $\langle \mathcal{W} \rangle \geq 0$ is a sufficient condition for entanglement. The observable \mathcal{W} is called an entanglement witness (EW). There always exists an optimal choice of *local* observables such that a given EW can be expressed as a sum of their direct products [4], so that a witness can always be measured locally. The advantage of using EWs for entanglement detection will be appreciated better for multipartite systems with more than two qubits because the number of tomographic measurements would grow exponentially with the number of qubits. On the other hand, a given witness does not detect all entangled states and therefore a variety of different EWs should be tested in order to rule out false negative results.

EWs assume the validity of quantum mechanics and also assume one knows what measurements one is actually performing. A valuable alternative to EW can be sought in using a violation of Bell-Clauser-Horne-Shimony-Holt (CHSH) inequalities [5,6] as a sufficient condition for entanglement (although a Bell-CHSH inequality test can be formulated as a witness, too [7]). Because Bell-CHSH inequalities are derived from classical probability theory without any reference to quantum mechanics, no assumption about what is being measured is necessary. This method is, therefore, safe in the

sense of avoiding many pitfalls arising from unwarranted (hidden) assumptions about one's experiment [8].

Here we propose a different method for characterizing an entanglement source that automatically takes into account finite data as well as imperfect measurements. Our method consists of two parts. The first part, "Bayesian updating," produces an estimate of the relative probabilities that entangled and separable states are consistent with a given finite set of data. This estimate depends on what *a priori* probability distribution (the *prior*) one chooses over all possible states (the more data one has, the less it depends on the prior). That is, there is an *a priori* probability of entanglement and each single measurement updates this probability to an *a posteriori* probability of entanglement. The latter then has to be compared to the former in order to reach the conclusion that one is now either more certain or less certain about having produced an entangled state. In fact, every experiment can only make such probabilistic statements about entanglement, although this is almost never explicitly stated in these terms. Thus our method differs from those in Refs. [9–11] which assume expectation values of EWs are known (corresponding effectively to an *infinite* data set) and try to find the *minimally* entangled state consistent with those expectation values. We use a numerical Bayesian updating method for a probability distribution over density matrices, which is similar to that recently discussed in Ref. [12] in the context of quantum-state tomography. In particular, whereas the reconstruction of a density matrix from experimental data is usually based on the maximum-likelihood estimation (MLE), Ref. [12] discusses its drawbacks and proposes Bayesian updating in its stead as a superior method. Our aim, though, is not to give an estimate of the density matrix, but of entanglement. In fact, any quantity that can be calculated from a density matrix, such as the purity of one's state, can be estimated this way.

The second part of our method introduces two information criteria [13] to judge how different models of a given entanglement generation process can be compared to each other quantitatively. It is probably best to explain this part by giving an example. For simplicity, consider the case of two-qubit states. Suppose an experimentalist has a model for her entanglement-generating source that contains, say, two parameters describing two physically different sources of noise

in the final two-qubit state produced. She may try to fit her data to her two-parameter model, but obviously there are always states in the full 15-dimensional set of all physical states that will fit the data better. There are a number of criteria, standard in the literature on statistical models, that compare quantitatively how different models fit the data and that guard against overfitting (i.e., fitting to the statistical noise). Here we will explore Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) [13]. These information criteria aim to find the most informative model, not the best-fitting model. The idea is that a two-parameter model fitting the data almost as well as the full quantum-mechanical description would provide more physical insight and a more economical (think Occam's razor) and transparent description. At least as importantly, it would predict *future* measurement results better. Each of the two information criteria, AIC and BIC, produces a "score" Ω for each model. One term in Ω is the logarithm of the maximum likelihood possible within each model and the second term subtracts a penalty for each parameter used in the model. The model with the larger value of Ω is then deemed to be the more informative. We propose here to combine information criteria with the Bayesian updating methods for entanglement estimation. Namely, we propose to use the more informative model to generate a "substitute prior." In the case that the data reveal that the more parsimonious model is also the more informative model, the numerical efforts required for our Bayesian updating method are much smaller (for instance, using a prior that depends on only two parameters) and yet should be expected to lead to correct descriptions of the entanglement generated by one's source. We will explicitly verify this expectation in simple cases involving two and three qubits.

This paper is organized as follows. In Sec. II, we give a general formulation of our method of Bayesian updating applicable to any quantum system. We also formulate precisely the two information criteria for model selection. In Secs. III and IV, we discuss numerical examples, which illustrate the Bayesian methods and the information criteria. For concreteness, we consider measurements of Bell-CHSH correlations (although any sort of measurements would do). The examples of Sec. III show that our method detects entangled two-qubit states that escape detection by any of the Bell-CHSH inequalities and even by violations of the stronger version of these bounds, which we call Roy-Uffink-Seevinck bounds [14–16]. Section IV considers three-qubit states. In Sec. V, we discuss the essential difference between our method of characterizing an entanglement source using information criteria and the standard methods of entanglement verification [8,17].

II. QUANTIFYING ENTANGLEMENT VIA BAYESIAN UPDATING

Here we present a numerical Bayesian updating method for one's probability distribution over density matrices. A related Bayesian method was recently advocated in Ref. [12] in the context of quantum-state tomography and quantum-state estimation. We note our aim is not to give an estimate

of the density matrix, but, more modestly, to give estimates of entanglement, purity, and in principle any quantity that can be efficiently calculated from the density matrix. We first discuss the method in general and subsequently we propose a method to choose a prior probability distribution over density matrices.

A. Method

Our method can be formulated as a five-step procedure:

(1) For a system of M qubits, we first choose a finite test set of M -qubit density matrices. We calculate the amount of entanglement (for example, the negativity [18]) for each state in the set.¹ The *a priori* probability that our unknown experimentally generated state, which we denote by $\rho_?$, equals a state ρ in the set is chosen as $p_{\text{prior}}(\rho) = 1/N_s$, where N_s is the number of states in the set.

(2) We assume some sets of positive operator valued measures (POVMs) with elements $\{\Pi_i\}$ are measured. These POVMs can describe any (noisy) set of measurements one performs on the qubits.

(3) For the acquired measurement record $d = \{d_1, \dots, d_i\}$ consisting of the number of times outcome i was obtained,² we calculate the quantum-mechanical probability $p(d|\rho)$ that a given state ρ from the test set generates the measurements outcome d (which follows directly from $\text{Tr } \rho \Pi_i$). Having at hand probabilities $p(d|\rho)$ for all states ρ in the test set, we are now able to calculate the *a priori* probability $p(d) = \sum_{\rho} p(d|\rho) / N_s$ for the measurement record d to occur.

(4) We calculate—using Bayes' rule—the probability $p(\rho|d)$ of having the state ρ given the measurement outcomes d : $p(\rho|d) = p(d|\rho) / [N_s p(d)]$.

(5) We obtain the posterior probability distribution over density matrices in our test set: $p_{\text{posterior}}(\rho) := p(\rho|d)$ for all states ρ .

We can then repeat steps 2–5 for a new set of measurements d if needed.

This procedure gives us, in step 5, a numerical estimate of the *a posteriori* probability that the unknown state $\rho_?$ equals the state ρ from the test set. From $p(\rho|d)$, we can estimate the probability p_e for the state $\rho_?$ to be entangled. We just sum the probabilities $p(\rho|d)$ for all entangled states ρ_{ent} in the set, i.e.,

$$p_e(\rho_?) = \sum_{\rho = \rho_{\text{ent}}} p(\rho|d). \quad (1)$$

Furthermore, we can calculate probability distributions for any function of the density matrix, such as the negativity and purity. We thus infer expectation values such as

¹We remark that in higher dimensions (or for more than two parties), the negativity does not necessarily pinpoint all classes of entangled states. In this case, more than one entanglement monotone should be used to characterize entanglement.

²We thus implicitly assume identical and independent copies of states of M qubits. For discussions of cases where this assumption is correct or incorrect, see [19] or [20], respectively.

$$\bar{N} = \sum_{\rho=\rho_{\text{ent}}} p(\rho|d)N(\rho), \quad (2)$$

where $N(\rho)$ is twice the sum of the absolute values of the negative eigenvalues of the partial transpose of ρ [18] and

$$\bar{P} = \sum_{\rho} p(\rho|d)\text{Tr}(\rho^2), \quad (3)$$

as well as standard deviations $\sigma_N = \sqrt{N^2 - \bar{N}^2}$, etc.

The meanings of our final probability distribution $p(\rho|d)$ and of the above expectation values are as follows. If we were forced to give a *single* density matrix that best describes all data and that includes error bars, we would give the mixed state $\bar{\rho} = \int d\rho p(\rho|d)\rho$, as explained in [12]. The purity and negativity of the state $\bar{\rho}$ are *not* equal to (in fact, smaller than) the estimates \bar{N} and \bar{P} that we use here. The difference is this: if one were to perform more measurements that are tomographically complete, \bar{N} is the expected negativity of the final estimated density matrix. $N(\bar{\rho})$, on the other hand, would be the useful entanglement of a single copy available *without* performing more measurements. For most quantum information processing purposes (such as teleportation), one indeed needs more precise knowledge about the density matrix than just its entanglement [see Ref. [8] for more discussions on this issue].

B. Model testing and information criteria

The only problem standing in the way of a straightforward application of the above Bayesian updating procedure is that a sufficiently dense test set (used in step 1) is in general too hard to handle numerically, since even for two-qubit density matrices the parameter space is 15 dimensional. Although there are certainly ways out of this problem (in particular, sampling directly from the posterior probability distribution can be efficiently done with the Metropolis-Hastings algorithm, see, e.g., [21]), here we stick to the idea of a set of test states by simplifying that set, as follows. [If the data do not allow a simpler model description, so that no small test set can be constructed in a reliable way, there is nothing left to do but bite the bullet and apply Metropolis-Hastings in the full space of all density matrices. It pays to have a good model for one's experiment.]

As an illustrative example (which we will again consider in great detail in the next section), consider an experimentalist trying to produce a maximally entangled Bell state of two qubits, say, $(|00\rangle + |11\rangle)/\sqrt{2}$. She wants to test her entanglement-generating device by measuring some set of Bell correlations. In particular, suppose she measures $2 < K < 15$ independent expectation values (we will be more precise in the next section).

From her previous experience with the same device, she models the generation process by assuming there is both Gaussian phase noise and white noise (mixing with the maximally mixed state $\propto 1$). That is, she assumes her device generates states of the form

$$\rho_{p,\sigma} = p\rho_{\sigma} + (1-p)\frac{1}{4}, \quad (4)$$

with $p \in [0, 1]$ and

$$\rho_{\sigma} = \frac{1}{2} \int_{-\pi}^{\pi} d\phi P(\phi) (|00\rangle + \exp(i\phi)|11\rangle)(\langle 00| + \exp(-i\phi)\langle 11|). \quad (5)$$

Here, $P(\phi)$ is a Gaussian phase distribution of the form

$$P(\phi) = N_{\sigma} \exp(-\phi^2/\sigma^2), \quad (6)$$

with the normalization factor N_{σ} given by

$$N_{\sigma} = \frac{1}{\int_{-\pi}^{\pi} d\phi \exp(-\phi^2/\sigma^2)}. \quad (7)$$

So there are just two parameters the experimentalist has to determine from her measurement results, p and σ .

As a measure to judge how well her data d fit the model (4)–(6), she considers the best likelihood for that model

$$L_{p,\sigma} \equiv \max_{p,\sigma} P(d|\rho_{p,\sigma}). \quad (8)$$

She would like to compare this to the maximum likelihood over *all* physical two-qubit states ρ ,

$$L_a \equiv \max_{\rho} P(d|\rho). \quad (9)$$

There are several ways to compare these two quantities [13]. One criterion is called Akaike information criterion and it defines the quantity³

$$\Omega = \ln(L) - k \quad (10)$$

for each model, where k is the number of parameters in the model and L is the maximum likelihood for the model. The quantity Ω rewards a high value of the best likelihood (indicating a good fit), but penalizes a large number of parameters (to guard against overfitting). Now when fitting the values of $2 < K < 15$ expectation values to a model, the best complete model contains just K , not 15, independent parameters.

Thus the experimentalist would calculate two numbers

$$\begin{aligned} \Omega_{p,\sigma} &= \ln(L_{p,\sigma}) - 2, \\ \Omega_a &= \ln(L_a) - K. \end{aligned} \quad (11)$$

If $\Omega_{p,\sigma} > \Omega_a$, then the Akaike information criterion judges the simple two-parameter model to be more informative than the complete K -parameter model.

There is a Bayesian version of this criterion [13] and it is defined in terms of similar quantities

$$\Omega' = \ln(L) - k \ln(N_m)/2, \quad (12)$$

where L and k have the same meaning as before and N_m is the number of data taken. Again, if $\Omega'_{p,\sigma} > \Omega'_m$, the two-

³For some reason, one often uses -2Ω as the figure of merit and tries to minimize it.

parameter model is considered more informative than the K -parameter description. For $N_m > 8$ (which we will always assume), the BIC puts a larger penalty on the number of parameters than does the AIC. Thus, if the AIC prefers the more parsimonious model, then so does the BIC.

In the case that the simple model turns out to be more informative, according to at least the BIC and possibly also the AIC (this depends on the data), we propose that the experimentalist may well use the simple model to construct a test set of states. For example, she could assume as prior probability distributions for p and σ that p is uniformly distributed on the interval $[0,1]$ and that σ is uniform on, say, the interval $[0, \pi]$ (this is somewhat arbitrary, of course, as every prior is). Then, the test set of states could be sampled by simply choosing N_p uniformly spaced points in the interval $[0,1]$ for p and N_σ uniformly spaced points in the interval $[0, \pi]$ for σ , thus creating a test set of $N_s = N_p N_\sigma$ states.

The above model leads to states that are diagonal in the Bell basis

$$\begin{aligned} |\Phi_1\rangle &= (|00\rangle + |11\rangle)/\sqrt{2}, \\ |\Phi_2\rangle &= (|00\rangle - |11\rangle)/\sqrt{2}, \\ |\Phi_3\rangle &= (|01\rangle + |10\rangle)/\sqrt{2}, \\ |\Phi_4\rangle &= (|01\rangle - |10\rangle)/\sqrt{2}. \end{aligned} \quad (13)$$

In the next section, we will consider not only the above two-parameter model, but also its obvious extension to a three-parameter model by allowing all Bell-diagonal states.

III. EXAMPLES

A. Orthogonal spin measurements

In the following, we consider, as an example, two-qubit states with spin measurements performed on each qubit (considered as a spin- $\frac{1}{2}$ system) in *two* arbitrary spatial directions that are orthogonal and we denote the corresponding spin operators by A_1 and A_2 for the first qubit and B_1 and B_2 for the second qubit. This constitutes a measurement of eight independent expectation values, four single-qubit expectation values, and four correlations.

We note that in this case, we can construct four Bell-CHSH operators from the four measured correlations

$$\begin{aligned} \mathcal{B}_1 &= A_1 \otimes (B_1 + B_2) + A_2 \otimes (B_1 - B_2), \\ \mathcal{B}_2 &= A_1 \otimes (B_1 + B_2) - A_2 \otimes (B_1 - B_2), \\ \mathcal{B}_3 &= A_1 \otimes (B_1 - B_2) + A_2 \otimes (B_1 + B_2), \\ \mathcal{B}_4 &= A_1 \otimes (-B_1 + B_2) + A_2 \otimes (B_1 + B_2). \end{aligned} \quad (14)$$

We then test two-qubit states that may be entangled but that do *not* violate any of the four Bell inequalities that can be constructed from these four operators. In fact, we will not even optimize the choice of spatial directions, given an ini-

tial guess of what state should be produced, for violating a Bell inequality.

Finally, we will add one more correlation to be measured, namely, that involving the third dimension: $A_3 B_3$. That is, whenever A_3 is measured on the first qubit, B_3 is measured on the second qubit. This addition makes the measurements on each qubit separately tomographically complete, it allows better quantitative estimates of entanglement, but it does not lead to additional Bell-CHSH operators. The total number of independent expectation values measured in this case is 11 (four out of 15 are missing). In order to test our methods, we first consider measurement simulations where all observed frequencies (of all measurement outcomes) are simply (and artificially) set equal to the underlying probabilities, given some fixed underlying quantum states. That is, there is a state perfectly fitting the data. In the next section, we will consider simulated data *with* statistical noise. Given that we do not consider statistical noise now, all information in the data are contained in the expectation values of the observables. Thus, the parameter K to be used for evaluating Ω_a of Eq. (11) is $K=11$.

B. Analytical results

Determining the AIC and BIC criteria can be done analytically in most cases that we will consider here. First of all, we can bound the maximum likelihood over all states, given the sort of measurements from the preceding section. There are 20 observed frequencies, as follows: for each of the five correlation measurements $A_i B_j$, where i, j take on the values $(i, j) = (1, 1), (1, 2), (2, 1), (2, 2), (3, 3)$, there are four different outcomes, which we can be denoted by $(+, +), (+, -), (-, +), (-, -)$. If we denote these frequencies by f_{ijk} , for $k=1 \dots 4$, then the (ln of the) maximum likelihood is bounded by

$$\ln(L_a) \leq \sum_{k,(ij)} N_{ij} f_{ijk} \ln(f_{ijk}), \quad (15)$$

where N_{ij} is the number of times the $A_i B_j$ correlation was measured. If we assume all five correlations are measured equally often, then we have

$$\ln(L_a) \leq \frac{N_m}{5} \sum_{k,(ij)} f_{ijk} \ln(f_{ijk}). \quad (16)$$

The bound is achieved when there is a physical state predicting the frequencies exactly as they were observed.

Let us choose directions of our spin measurements as $A_1 = B_1 = X$, $A_2 = B_2 = Y$, and $A_3 = B_3 = Z$. Then, there are two obvious models an experimentalist could choose from: the first is the Bell-diagonal model, containing three parameters, in which states are of the form

$$\rho = \sum_{i=1}^4 p_i |\Phi_i\rangle\langle\Phi_i|. \quad (17)$$

The observed frequencies f_{ijk} for the five correlations cannot be all predicted to arbitrary accuracy by Bell-diagonal states. In fact, most frequencies predicted by this model are independent of the values of $\{p_i\}$ and are equal to $1/4$. The only

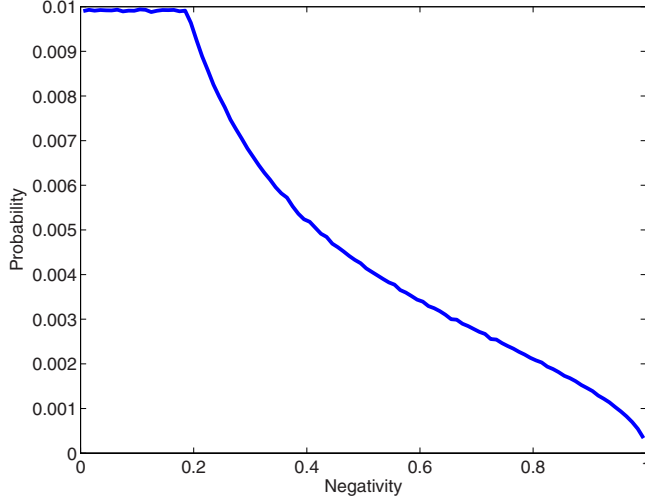


FIG. 1. (Color online) Prior probability distribution of the negativity for the two-parameter states $\rho_{p,\sigma}$ with p and σ drawn uniformly from $[0,1]$ and $[0,\pi]$, respectively. The values of the negativity are binned in 100 bins for entangled states. One additional bin is reserved for separable states (of zero negativity). The latter point is not shown in this graph for visual reasons: the probability of separability is 49.7%. The total number of states drawn from the prior distribution is $N_p \times N_\sigma = 600 \times 600$. With this number of test states, the plotted distribution has converged.

predicted frequencies (which we denote by \tilde{f} so as to distinguish them from the observed frequencies f) that actually depend on the values of $\{p_i\}$ are

$$\begin{aligned}
 \tilde{f}_{111} &= \tilde{f}_{114} = p_1/2 + p_3/2, \\
 \tilde{f}_{221} &= \tilde{f}_{224} = p_2/2 + p_3/2, \\
 \tilde{f}_{331} &= \tilde{f}_{334} = p_1/2 + p_2/2, \\
 \tilde{f}_{112} &= \tilde{f}_{113} = p_2/2 + p_4/2, \\
 \tilde{f}_{222} &= \tilde{f}_{223} = p_1/2 + p_4/2, \\
 \tilde{f}_{332} &= \tilde{f}_{333} = p_3/2 + p_4/2.
 \end{aligned} \tag{18}$$

The best-fitting Bell-diagonal state can only predict the correct correlations between XX , YY , and ZZ measurements. For example, there is a Bell-diagonal state predicting the correct value for the sum $\tilde{f}_{111} + \tilde{f}_{114}$, but its prediction for the difference will always be zero. Thus, the Bell-diagonal state fitting the data best will have the following values for $\{p_i\}$:

$$\begin{aligned}
 p_1 &= [f_{111} + f_{114} - f_{221} - f_{224} + f_{331} + f_{332}]/2, \\
 p_2 &= [-f_{111} - f_{114} + f_{221} + f_{224} + f_{331} + f_{332}]/2, \\
 p_3 &= [f_{111} + f_{114} + f_{221} + f_{224} - f_{331} - f_{332}]/2,
 \end{aligned} \tag{19}$$

provided the observed frequencies are such that the $\{p_i\}$ including $p_4 = 1 - p_1 - p_2 - p_3$ are all nonnegative. In that case,

the (ln of the) maximum likelihood over all Bell-diagonal states is

$$\begin{aligned}
 \ln L_{Bd} &= \frac{N_m}{5} \left[\sum_i (f_{ii1} + f_{ii4}) \ln[(f_{ii1} + f_{ii4})/2] \right. \\
 &\quad \left. + \sum_i (f_{ii2} + f_{ii3}) \ln[(f_{ii2} + f_{ii3})/2] + \sum_{k,i \neq j} f_{ijk} \ln(1/4) \right].
 \end{aligned} \tag{20}$$

For the Bell-diagonal model, we can construct a prior distribution over Bell-diagonal states by choosing the numbers $\{p_i\}$ uniformly over the simplex, as explained in [22].

The two-parameter model is similar in its predictions to the Bell-diagonal model. The only difference is that the parameters p_3 and p_4 are equal. Thus, this model can predict only two correlations correctly, namely, ZZ and $XX - YY$. The maximum likelihood for this model, then, is given by

$$\begin{aligned}
 \ln L_{p,\sigma} &= \frac{N_m}{5} (f_{331} + f_{334}) \ln[(f_{331} + f_{334})/2] \\
 &\quad + (f_{332} + f_{333}) \ln[(f_{332} + f_{333})/2] \\
 &\quad + (f_{111} + f_{114} + f_{222} + f_{223}) \\
 &\quad \times \ln[1/4 + (f_{111} + f_{114} - f_{221} - f_{224})/2] \\
 &\quad + (f_{221} + f_{224} + f_{112} + f_{113}) \\
 &\quad \times \ln[1/4 + (f_{221} + f_{224} - f_{111} - f_{114})/2] \\
 &\quad + \sum_{k,i \neq j} f_{ijk} \ln(1/4),
 \end{aligned} \tag{21}$$

provided all inferred frequencies are nonnegative.

The three parameters to be used for selecting the most informative model are, in the case of AIC,

$$\begin{aligned}
 \Omega_a &= \ln L_a - 11, \\
 \Omega_{p,\sigma} &= \ln L_{p,\sigma} - 2, \\
 \Omega_{Bd} &= \ln L_{Bd} - 3,
 \end{aligned} \tag{22}$$

and similar expressions for the BIC. In fact, it is not the absolute values of these quantities that are significant, but the *differences* between these three quantities for the three different models.

C. Numerics

Let us first discuss the two-parameter substitute prior with p and σ drawn uniformly from $[0,1]$ and $[0,\pi]$, respectively. As our favorite entanglement monotone, we use the negativity [18,22]. The prior probability distribution for negativity is displayed in Fig. 1 (the graph for concurrence is the same for this special case of Bell-diagonal states). The plot shows that states exist in the full range of separable to maximally entangled, with the prior probability of entanglement being $P_{\text{ent}} = 50.3\%$.

Using this prior, we consider the measurement of five different Bell correlations. Sample results are displayed and discussed in Figs. 2–4. Figures 2 and 3 show measurement

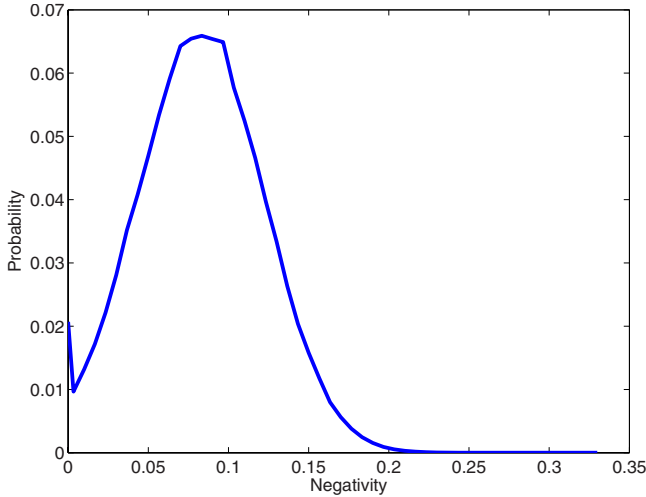


FIG. 2. (Color online) The state underlying the data considered here is of the form $\rho_{p,\sigma}$, with $p=0.4$ and $\sigma=0.4$. Each of the five Bell correlation observables is measured 400 times, so that in total $N_m=2000$ measurements are performed. Plotted is the posterior probability distribution for the negativity. The test set of two-parameter states is the same as that for Fig. 1. The results have been binned together in 50 bins of equal size for entangled states, plus one extra bin for separable states (at zero negativity). The posterior probability for entanglement is 98% in this case (with 2% falling in the first bin at zero negativity). The estimated negativity and its error bar are $\bar{N}=0.082 \pm 0.039$, where the actual negativity of the state $\rho_{0.4,0.4}$ is $N(\rho_{0.4,0.4})=0.0843$.

results generated from an entangled state $\rho_\gamma=\rho_{0.4,0.4}$, as defined in Eq. (4). There is no need to test either AIC or BIC for this case, since the state is chosen from the two-parameter set of states, so the two-parameter model is trivially more informative. The Bayesian posterior probability for entanglement distribution is consistent with the actual en-

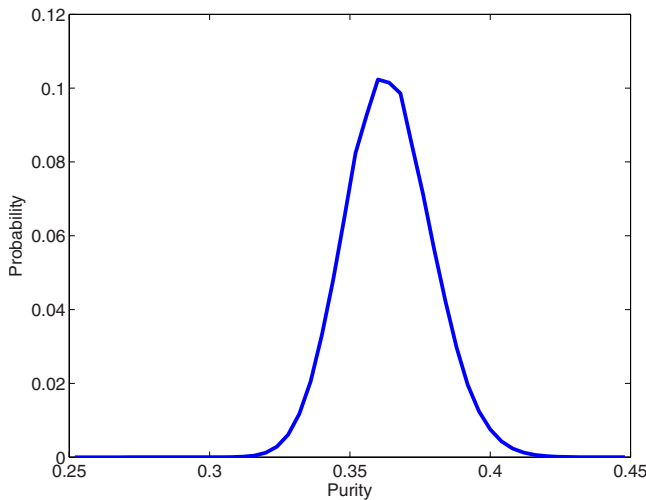


FIG. 3. (Color online) Same as Fig. 2, but for the purity. The purity of the state $\rho_{0.4,0.4}$ equals 0.3638; the estimate (plus error bar) obtained for the purity is $\bar{P}=0.364 \pm 0.016$. The estimate of the purity is, relatively speaking, more accurate than that of entanglement.

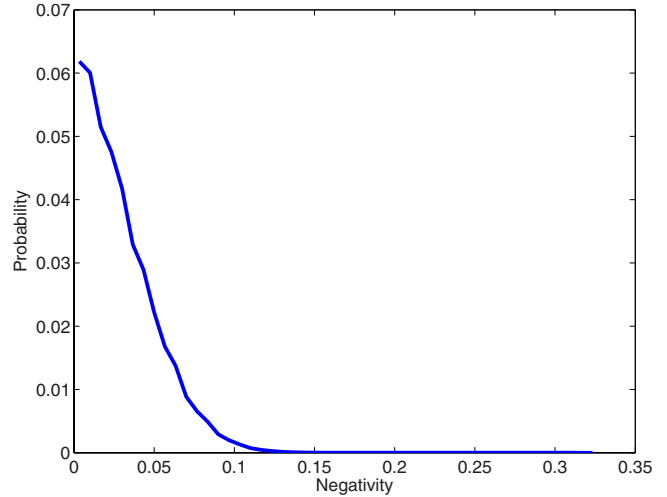


FIG. 4. (Color online) Same as Fig. 2 but for the separable state $\rho_{1/3,1/3}$. The estimate obtained for the negativity is $\bar{N}=0.012$ and its standard deviation is $\sigma_N=0.020$. The probability of entanglement is $P_{\text{ent}}=40.5\%$. The bin at zero negativity contains 59.5% probability and that point is not plotted for visual reasons. No firm conclusion about entanglement is reached in this case. This is to be expected as the state $\rho_{1/3,1/3}$ lies on the border between separable and entangled states.

tanglement properties of ρ_γ , as discussed in the figure caption.

We then also test a state that is just separable, the state $\rho_{1/3,1/3}$. The results are displayed in Fig. 4 and can be summarized as “inconclusive” about the question whether the data inform the experimentalist that the underlying state is entangled or not. This is not surprising given how close the actual state is to the separable or entangled boundary. The plot for purity is not shown, as it is very similar to Fig. 3 (the estimate of the purity is $\bar{P}=0.331 \pm 0.013$, perfectly consistent with the actual purity of 0.3303 of $\rho_{1/3,1/3}$).

Next we consider the following family of states:

$$\rho_k = 0.5|\psi_k\rangle\langle\psi_k| + 0.5\mathbb{1}/4, \tag{23}$$

with $k \leq 1$ and

$$|\psi_k\rangle = (|00\rangle + k|11\rangle)/\sqrt{1+k^2}. \tag{24}$$

For $k=1$, this state is in the two-parameter set, but for $k < 1$ it is not. Obviously, the smaller $k < 1$ is, the less well it is approximated by a state $\rho_{p,\sigma}$. We investigate how well the two-parameter model does by calculating

$$\begin{aligned} \Delta\Omega &\equiv \Omega_{p,\sigma} - \Omega_a, \\ \Delta\Omega' &\equiv \Omega'_{p,\sigma} - \Omega'_a \end{aligned} \tag{25}$$

and tabulating the values for several values of $k < 1$ in Table I. We moreover give the estimated negativities and purities, plus their error bars, as compared to the actual values of those quantities for the states ρ_k . What the table shows is that the two-parameter model ceases to be more informative when $k < 1$ decreases. At that point, the estimate of negativity is still perfectly fine, but the estimate of the purity starts

TABLE I. Comparison, through the Akaike and Bayesian information criteria [using Eq. (25)], of the two-parameter model based on the family of states $\rho_{p,\sigma}$ [Eq. (4)] and the full 15-parameter description of all two-qubit states, with measurement data generated from the family of states ρ_k [Eq. (23)]. Here, the number of measurements is $N_m=5 \times 1000$. The purity of ρ_k is equal to $P=0.4375$ for any value of $0 < k < 1$. For decreasing values of k , $\Delta\Omega$ and $\Delta\Omega'$ decrease, indicating the two-parameter becomes less and less informative. The estimate of purity becomes, likewise, less and less reliable.

k	N	$\bar{N} \pm \sigma_N$	$\bar{P} \pm \sigma_P$	$\Delta\Omega$	$\Delta\Omega'$
0.9	0.247	0.246 ± 0.024	0.436 ± 0.012	7.2	36
0.8	0.238	0.237 ± 0.024	0.431 ± 0.012	0.9	30
0.7	0.220	0.219 ± 0.024	0.423 ± 0.012	-11	18
0.6	0.191	0.190 ± 0.024	0.410 ± 0.011	-29	0.8
0.5	0.150	0.149 ± 0.025	0.392 ± 0.011	-53	-24

to fail. In the last entry, for $k=0.5$, the two-parameter's model's estimate of purity is statistically incorrect. In this particular case, the AIC works better than the BIC.

In order to consider the three-parameter Bell-diagonal model, we first display the prior distribution for negativity of that model in Fig. 5. Next we discuss a state that is not close to any state in the two-parameter set of states, but that is still reasonably well described by the three-parameter model

$$\rho_1 = 0.53|\psi_1\rangle\langle\psi_1| + 0.47|\phi_1\rangle\langle\phi_1|, \quad (26)$$

with

$$\begin{aligned} |\psi_1\rangle &= (|00\rangle + 0.9|11\rangle)/\sqrt{1.81}, \\ |\phi_1\rangle &= (|01\rangle + 0.9|10\rangle)/\sqrt{1.81}. \end{aligned} \quad (27)$$

We consider $N_m=5 \times 1000$ measurements, with each correlation being measured 1000 times. For the calculations with

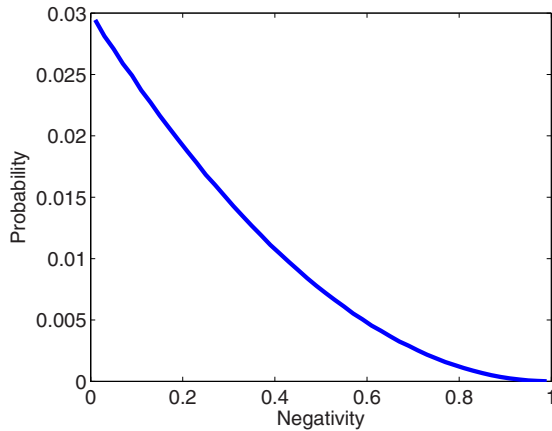


FIG. 5. (Color online) Prior probability distribution of the negativity for the three-parameter set of Bell-diagonal states. The point at zero negativity is left out for visual reason: separable states occupy 50.0% of the total volume. Here, 10^7 states were drawn from the prior distribution over states to obtain numerical convergence.

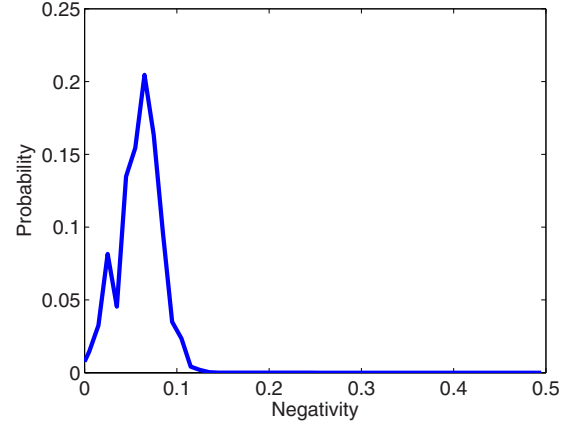


FIG. 6. (Color online) Posterior probability distribution of the negativity, using the three-parameter set of Bell-diagonal states as prior, for data generated from the state ρ_1 of Eq. (26). The size of the test set is 10^7 states, just as in Fig. 5.

the three-parameter model, a test set of size 10^7 was used. In contrast, for the two-parameter model, test sets of size 600×600 were sufficient in all cases (that is, increasing the test set sizes starting from 100×100 showed all results converged by the time sizes of 600×600 were reached). This illustrates that choosing a good physical model with as few parameters as possible pays large dividends. For this state, we calculate the AIC and BIC and compare the two- and three-parameter (Bell-diagonal) models to the full-state model,

$$\Omega_{Bd} - \Omega_a = 2.4,$$

$$\Delta\Omega = \Omega_{p,\sigma} - \Omega_a = -462, \quad (28)$$

for the AIC, and

$$\Omega'_{Bd} - \Omega'_a = 27,$$

$$\Delta\Omega' = \Omega'_{p,\sigma} - \Omega'_a = -433, \quad (29)$$

for the BIC. That is, the three-parameter model is considered more informative than the model containing all physical states. On the other hand, the two-parameter model is *much* less informative. The estimates for negativity and purity are, for the three-parameter model

$$\bar{N}^{Bd} = 0.059 \pm 0.022,$$

$$\bar{P}^{Bd} = 0.4977 \pm 0.0025, \quad (30)$$

where the actual values are

$$N = 0.059,$$

$$P = 0.502. \quad (31)$$

Thus, both purity and negativity are estimated correctly within the three-parameter model and this is what one would expect given the AIC and BIC criteria. The posterior probability distribution for the negativity is plotted in Fig. 6.

For the two-parameter model, in contrast, we get

$$\begin{aligned}\bar{N} &= 0.056 \pm 0.025, \\ \bar{P} &= 0.353 \pm 0.009,\end{aligned}\quad (32)$$

so that again the purity estimated by the two-parameter model is inaccurate, although the estimated negativity is still quite good. Thus, when the AIC and BIC criteria tell one not to trust a certain simple model, it does not imply that *all* estimated quantities from that model are, in fact, incorrect.

Lest one starts to think that the two-parameter model in fact somehow always estimates the negativity correctly, even if the estimated purity is wrong. Here is a counter example to that idea: when the $N_m=5 \times 1000$ data are generated by the mixture

$$\rho_2 = 0.53|\psi_2\rangle\langle\psi_2| + 0.47|\phi_2\rangle\langle\phi_2|, \quad (33)$$

with

$$\begin{aligned}|\psi_2\rangle &= (|00\rangle + 0.5|11\rangle)/\sqrt{1.25}, \\ |\phi_2\rangle &= (|01\rangle + 0.5|10\rangle)/\sqrt{1.25},\end{aligned}\quad (34)$$

whose negativity is $N=0.039$, the two-parameter model that concludes the state is separable with high probability, $P_{\text{ent}}=3.1\%$, and $\bar{N}=3.5 \times 10^{-4} \pm 0.0025$ (and the estimated purity is incorrect as well, $\bar{P}=0.319 \pm 0.008$, instead of the correct value $P=0.502$). Here, $\Delta\Omega=-203$.

IV. BEYOND TWO QUBITS

In this section, we consider *noisy* data generated from a particular state from a one-parameter family. We test whether the information criteria judge that one-parameter model as more informative than a complete model even in the case of large statistical noise.

The advantage of the Bayesian updating method for entanglement quantification purposes can be more fully appreciated in the case of multiqubit systems with the number of qubits greater than 2. Even for a three-qubit system, if one were to use conventional tomography, one would already need to measure 63 independent (real) parameters (compared to only 15 for two qubits). Now, suppose that an experimentalist is trying to quantify entanglement of a three-qubit state but is restricted to measure only a small subset of the observables: $\{O_1, \dots, O_m\}, m \ll 63$. She wants to apply the Bayesian updating method to her data. First, she models the device utilizing her prior experience with it. Let us assume that according to her model the device produces the following states:

$$\rho_p = p\rho_{\text{GHZ}} + (1-p)\frac{\mathbb{1}}{8}, \quad (35)$$

with $p \in [0, 1]$. Here, ρ_{GHZ} is a three-qubit Greenberger-Horne-Zielinger (GHZ) state, i.e.,

$$\rho_{\text{GHZ}} = \frac{1}{2}(|0,0,0\rangle + |1,1,1\rangle)(\langle 0,0,0| + \langle 1,1,1|). \quad (36)$$

Imagine that she can perform the spin measurements on each qubit in two arbitrary but spatially orthogonal directions. Following our previous notations, we denote the corresponding spin operators A_1 and A_2 for the first, B_1 and B_2 for the second, and C_1 and C_2 for the third qubit. This provides her with eight independent three-party correlations. Similarly to the two-qubit case of the preceding section, we assume she can in addition access the observable $A_3 \otimes B_3 \otimes C_3$ along the third spatial dimension, increasing the total number of measured three-qubit correlations to 9. This makes the measurements on each qubit separately tomographically complete.

Each of the nine three-party correlation measurements $A_i B_j C_k$ $[(i, j, k) = (1, 1, 1), (1, 2, 1), (2, 1, 1), (2, 2, 1), (1, 1, 2), (1, 2, 2), (2, 1, 2), (2, 2, 2), (3, 3, 3)]$ is associated with eight possible outcomes, denoted as $(+, +, +), (+, +, -), (+, -, +), (+, -, -), (-, +, +), (-, +, -), (-, -, +),$ and $(-, -, -)$. Each outcome has a relative frequency of occurrence f_{ijkm} ($m=1 \dots 8$) that can be determined analytically using the Born rule, $f_{ijkm} = \langle m | \rho | m \rangle$. For instance, for the one-parameter model in Eq. (35), the frequencies are

$$\begin{aligned}\tilde{f}_{1111} &= \tilde{f}_{1114} = \tilde{f}_{1116} = \tilde{f}_{1118} = (1+p)/8, \\ \tilde{f}_{1112} &= \tilde{f}_{1113} = \tilde{f}_{1115} = \tilde{f}_{1117} = (1-p)/8, \\ \tilde{f}_{2211} &= \tilde{f}_{2214} = \tilde{f}_{2216} = \tilde{f}_{2217} = (1-p)/8, \\ \tilde{f}_{2212} &= \tilde{f}_{2213} = \tilde{f}_{2215} = \tilde{f}_{2218} = (1+p)/8, \\ \tilde{f}_{1222} &= \tilde{f}_{1223} = \tilde{f}_{1225} = \tilde{f}_{1228} = (1+p)/8, \\ \tilde{f}_{1221} &= \tilde{f}_{1224} = \tilde{f}_{1226} = \tilde{f}_{1227} = (1-p)/8, \\ \tilde{f}_{2121} &= \tilde{f}_{2124} = \tilde{f}_{2126} = \tilde{f}_{2127} = (1-p)/8, \\ \tilde{f}_{2122} &= \tilde{f}_{2123} = \tilde{f}_{2125} = \tilde{f}_{2128} = (1+p)/8, \\ \tilde{f}_{3331} &= \tilde{f}_{3338} = (1+3p)/8, \\ \tilde{f}_{3332} &= \dots = \tilde{f}_{3337} = (1-p)/8.\end{aligned}\quad (37)$$

The remaining 32 frequencies are constant and equal $\frac{1}{8}$.

The experimentalist thus measures $63=9 \times 7$ independent frequencies. To distinguish observed from calculated frequencies, which are model dependent, we denote the former by f_{ijkm} . The measured frequencies are used to calculate the likelihood function $P(f_{ijkm} | \rho_p)$ for a given source model ρ_p . Depending on the model, the number of variables in the likelihood function can vary from 1 to 63. In reality, the observed frequencies are always corrupted by noise and, therefore, choosing a source model in which the likelihood function depends on 63 independent parameters will lead to the model with the largest possible likelihood (and this number of 63 *happens* to equal the number of parameters needed to describe a general three-qubit mixed state; note also that

we do consider statistical noise in the data, so expectation values of correlation functions are not sufficient to characterize the data).

In order to determine whether the simple one-parameter model in Eq. (35) is more informative than, say, the 63-parameter model, that predicts the measured frequencies to be independent, the experimentalist calculates the quantities

$$\Omega_p = \ln L_p - 1,$$

$$\Omega_{63} = \ln L_{63} - 63,$$

$$\Omega'_p = \ln L_p - \ln(N_m)/2,$$

$$\Omega'_{63} = \ln L_{63} - 63 \ln(N_m)/2, \quad (38)$$

defined by the corresponding AIC and BIC criteria. Here, L_p and L_{63} are the maximum-likelihood values for the one- and 63-parameter models. Given the measured data f_{ijkm} ,

$$L_p = \max_{\rho_p} P(f_{ijkm} | \rho_p). \quad (39)$$

She can also find L_{63} without providing the source model explicitly,⁴

$$L_{63} = \max_{f_{ijkm}} P(f_{ijkm} | \tilde{f}_{ijkm}). \quad (40)$$

This maximum can be upper bounded by setting calculated frequencies equal to observed frequencies.

If it turns out that $\Omega_p > \Omega_{63}$ and $\Omega'_p > \Omega'_{63}$, then the experimentalist may use the one-parameter model in Eq. (35) to generate the substitute prior for the Bayesian updating. Otherwise, she has to consider increasing the number of model parameters and improving the model.

Numerics

To test the three-qubit example numerically, we first have generated 1000 sets of noisy measurement data. Gaussian noise was added to all 72 frequencies f_{ijkm}^0 calculated assuming that the unknown state is $\rho_{3/4}$. The noise was chosen to be

$$f_{ijkm} = |f_{ijkm}^0 + \epsilon \sqrt{f_{ijkm}^0(1-f_{ijkm}^0)/N_f}|, \quad (41)$$

where ϵ is a normally distributed random variable and N_f is the number of measurements of each observable. We chose N_f to be rather small, $N_f=50$, so as to have a large amount of statistical noise.

Using the simulated data, we have verified how often both $\Omega_p - \Omega_{63}$ and $\Omega'_p - \Omega'_{63}$ are positive, i.e., when AIC and BIC simultaneously conclude that the one-parameter model is more informative. The results are presented in Figs. 7 and 8. The scatter plots clearly indicate that the one-parameter model is indeed more favorable for *every* generated data set

⁴The explicit source model, which predicts 63 independent frequencies, is only needed when one is interested in finding the MLE state.

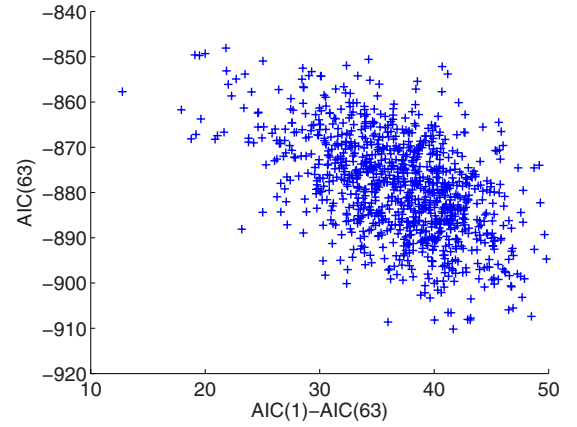


FIG. 7. (Color online) Scatter plot of the difference $\Omega_p - \Omega_{63}$ vs the value Ω_{63} for 1000 noisy data sets generated from the state $\rho_{3/4}$ defined in Eq. (35). The noise is added in accordance with Eq. (41) with $N_f=50$. This plot thus concerns the AIC.

(in spite of the large amount of noise). Following this conclusion, we use it to generate a prior distribution over one-parameter states, with the parameter p uniformly distributed over $[0,1]$.

We use this one-parameter prior to estimate the amount of three-partite entanglement and purity from the simulated data. There is no perfect measure of entanglement for three qubits; the measure of entanglement used is the one advocated in [24],

$$N_{ABC} = (N_{A-BC} N_{B-AC} N_{C-AB})^{1/3}, \quad (42)$$

in terms of the negativities of the three possible bipartite splits of a three-qubit system (the qubits are labeled A , B , and C). The major difference from the two-qubit examples is that here we implement the Bayesian updating with noisy data. For that, we generated 300 new sets of noisy data. We chose 50 000 test states for the prior and the same prior was implemented for each of the 300 sets, and using the Bayesian posterior probabilities, we calculated the mean three-qubit negativity and purity for each. A scatter plot (Fig. 9) of mean negativity vs. mean purity shows the very strong correlation between the two estimates (with a more pure state also hav-

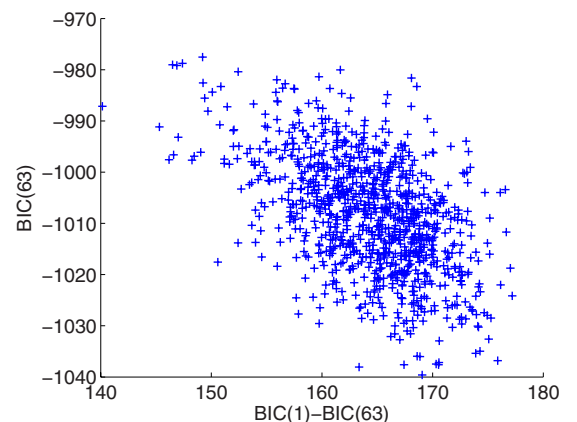


FIG. 8. (Color online) Same as in Fig. 7, but for the BIC.

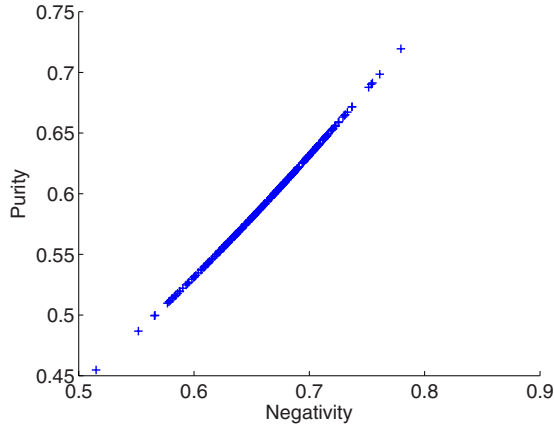


FIG. 9. (Color online) For this plot, 300 different noisy data sets were produced, assuming that the “true” state underlying the data is of the form ρ_p with $p = \frac{3}{4}$. Each of the nine three-party correlations is measured 50 times, so that in total $N_m = 50 \times 9 = 450$ measurements are performed. For each data set, the mean value of the three-qubit negativity is calculated and plotted for the corresponding estimate of purity.

ing more entanglement). Moreover, it shows the large fluctuations in both estimates.

Averaging the mean values over all 300 data sets gives the following negativity and purity estimates:

$$\begin{aligned}\bar{N}_{\text{av}} &= 0.66 \pm 0.04, \\ \bar{P}_{\text{av}} &= 0.59 \pm 0.04,\end{aligned}\quad (43)$$

which are consistent with the actual values for $\rho_{3/4}$: $N=0.6875$, $P=0.6172$. The probability distributions of three-

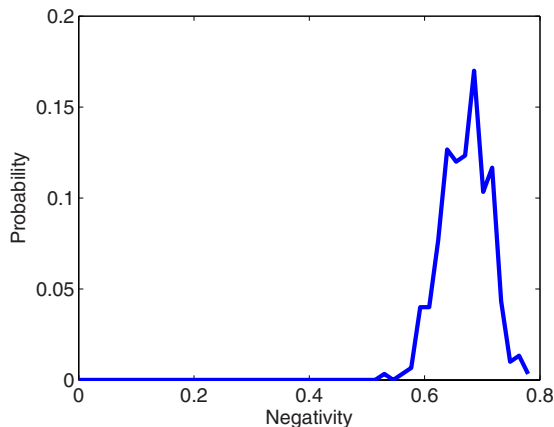


FIG. 10. (Color online) For the same 300 sets of noisy data as in Fig. 9, the probability distribution for the mean negativity over all sets is plotted. The results have been binned together in 50 bins of equal size. The estimated negativity and its error bar are $\bar{N}_{\text{av}} = 0.66 \pm 0.04$, where the actual negativity of the state $\rho_{3/4}$ is $N = 0.6875$. Note this is a different sort of plot than Figs. 2, 4, and 6: there, *one* set of data leads to a probability distribution over negativities, leading to *one* average estimate \bar{N} for the negativity. In this figure, 300 noisy data sets are used to produce 300 such averages and those are in turn averaged to produce \bar{N}_{av} .

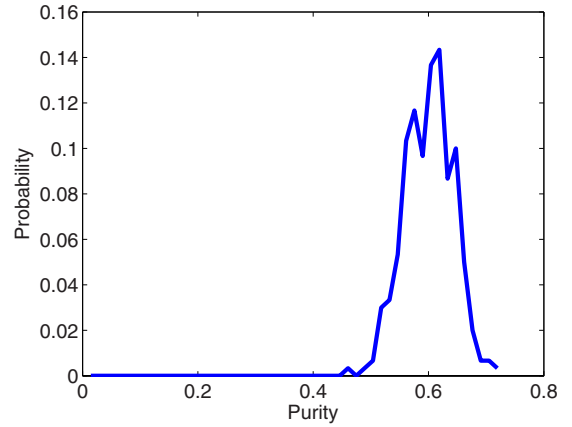


FIG. 11. (Color online) The same as in Fig. 10 but for the purity. Here, the average value found is $\bar{P}_{\text{av}} = 0.59 \pm 0.04$, where the actual value of the purity is $P = 0.6172$.

qubit entanglement and purity over all data sets depicted in Figs. 10 and 11 can also be utilized to confirm that the Bayesian updating method, using just a one-parameter test set of states, produces a statistically accurate estimate of the relevant parameters even in the presence of large measurement noise.

V. DISCUSSION AND CONCLUSIONS

We have demonstrated a method to characterize entanglement sources from finite sets of data using a Bayesian updating procedure for the probability distribution over density matrices. One thus obtains also *posterior* probability distributions for any quantities that can be efficiently calculated from an arbitrary density matrix. For instance, one obtains a probability that one’s state is entangled, as well as expectation values of any computable entanglement monotone, including estimates of their statistical errors. These values should be compared to their *a priori* values to judge whether one’s measurement results lead one to be more certain about entanglement or less.

For two qubits, it is in principle sufficient for the purpose of detecting entanglement to measure spin on each qubit in just two orthogonal directions. On the other hand, empirically, we found that for accurately *quantifying* two-qubit entanglement, adding one more correlation measurement is very beneficial. Thus we concentrated on discussing measurements of five spin-spin correlation functions. Similarly, in our simulations of measurements on three qubits, we used nine correlation measurements, consisting of spin measurements in two directions for each qubit, plus one additional correlation.

It is hard to say in general what sort of measurements, short of fully tomographic measurements, will be sufficient for estimating what sort of quantities. An easy check, though, is to count by how many parameters a given quantity is determined. For instance, purity is determined by the eigenvalues of the density matrix. Thus, for two qubits, one needs only three parameters. Thus, reliably estimating the purity of one’s output states ought to be easier than estimating en-

tanglement. Our simulations confirm this suspicion, producing relatively smaller error bars for estimates of purity than for entanglement.

It is important to note that in the above, we used the phrase “characterizing entanglement sources” rather than “verifying entanglement” because the latter method, in its standard interpretation, has a different meaning: in entanglement verification, one tries to find a proof of entanglement convincing a skeptic outsider. But our Bayesian method rather describes one’s own belief. In particular, the difference is that one’s prior belief of the entanglement-generating source is certainly to be included in a Bayesian description, but in entanglement-verification methods such beliefs are not allowed. Nevertheless, Bayesian methods can be used for the stricter purpose of entanglement verification as well, as is discussed in [23].

In order to characterize one’s entanglement source, then, it is allowed to *tentatively* use a model describing one’s source based on, e.g., previous experiments and experiences with the same (or similar) device. This model can be compared to the complete model that parametrizes the output states by using the full quantum-mechanical description of an arbitrary state of correct Hilbert-space dimension. The latter model, though, while being complete, may have more parameters than wished for or needed. We proposed to use two information criteria, both of which use the data to judge the relative merits of such models. One is the AIC and the other is the BIC [13]. We then showed how the AIC and BIC can be used to construct a test set of states, i.e., an

a priori probability distribution over quantum states generated by one’s source: a Bayesian method, of course, only produces probabilities of entanglement by first constructing a prior.

If a simple model described one’s source very well, then one’s test set can be based on that model. We applied the AIC and BIC criteria to several examples, all involving two or three qubits, and showed that indeed, such criteria indicate whether a given model’s predictions about purity and entanglement of the output of the source (including a probability that one’s output state is entangled, as well as an estimate of the amount of entanglement) can be expected to be reliable or not. We demonstrated this by showing that certain estimates produced from a simple model are wrong if the information criteria deem the model to be less informative than the full description of all quantum states, whereas those estimates are right on the mark when the criteria deem the simple model to be more informative.

Thus, although our proposed method for entanglement source characterization violates one of the strict criteria of entanglement verification by making certain assumptions about the state generated, the actual data are used to verify those assumptions.

ACKNOWLEDGMENTS

S.J.v.E. thanks Robin Blume-Kohout for many useful and inspiring discussions. This research is supported by the Disruptive Technologies Office (DTO) of the DNI.

-
- [1] *Quantum State Estimation*, edited by M. Paris and J. Rehacek, Lecture Notes in Physics Vol. 649 (Springer, Berlin, 2004).
- [2] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).
- [3] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).
- [4] O. Gühne *et al.*, Phys. Rev. A **66**, 062305 (2002).
- [5] J. S. Bell, *Speakable and Unspeakeable in Quantum Mechanics* (Cambridge University Press, Cambridge, England, 1987).
- [6] J. Clauser, M. Horne, A. Shimony, and R. Holt, Phys. Rev. Lett. **23**, 880 (1969); J. F. Clauser and M. Horne, Phys. Rev. D **10**, 526 (1974); J. F. Clauser and A. Shimony, Rep. Prog. Phys. **41**, 1881 (1978).
- [7] P. Hyllus *et al.*, Phys. Rev. A **72**, 012321 (2005).
- [8] S. J. van Enk, N. Lütkenhaus, and H. J. Kimble, Phys. Rev. A **75**, 052318 (2007).
- [9] K. M. R. Audenaert and M. B. Plenio, New J. Phys. **8**, 266 (2006).
- [10] J. Eisert, F. G. S. L. Brandão, and K. M. R. Audenaert, New J. Phys. **9**, 46 (2007).
- [11] O. Gühne, M. Reimpell, and R. F. Werner, Phys. Rev. Lett. **98**, 110502 (2007).
- [12] R. Blume-Kohout, e-print arXiv:quant-ph/0611080.
- [13] G. Claeskens and N. L. Hjort, *Model Selection and Model Averaging*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, 2008); K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer-Verlag, Berlin, 2002).
- [14] S. M. Roy, Phys. Rev. Lett. **94**, 010402 (2005).
- [15] J. Uffink and M. Seevinck, Phys. Lett. A **372**, 1205 (2008).
- [16] P. Lougovski and S. J. van Enk, Phys. Rev. A **80**, 034302 (2009).
- [17] O. Gühne and G. Toth, Phys. Rep. **474**, 1 (2009).
- [18] G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002).
- [19] R. Renner, Nat. Phys. **3**, 645 (2007).
- [20] S. J. van Enk, Phys. Rev. Lett. **102**, 190503 (2009).
- [21] B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004); C. P. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer, New York, 2004).
- [22] K. Życzkowski *et al.*, Phys. Rev. A **58**, 883 (1998).
- [23] R. Blume-Kohout, J. O. S. Yin, and S. J. van Enk (unpublished).
- [24] C. Sabin and G. Garcia-Alcaine, Eur. Phys. J. D **48**, 435 (2008).