PRACTICAL ISSUES IN THEORETICAL DESCRIPTIONS OF EXPERIMENTAL QUANTUM STATE AND ENTANGLEMENT ESTIMATION

by

JUN YIN

A DISSERTATION

Presented to the Department of Physics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2011

DISSERTATION APPROVAL PAGE

Student: Jun Yin

Title: Practical Issues in Theoretical Descriptions of Experimental Quantum State and Entanglement Estimation

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics by:

Michael G. Raymer	Chair
Steven J. van Enk	Advisor
Stephen Hsu	Member
Jens U. Nöchel	Member
Jeffrey A. Cina	Outside Member
and	
Kimberly Espy	Vice President for Research and Graduate Studies/ Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2011

DISSERTATION ABSTRACT

Jun Yin

Doctor of Philosophy

Department of Physics

June 2011

Title: Practical Issues in Theoretical Descriptions of Experimental Quantum State and Entanglement Estimation

Approved: _____
Steven J. van Enk

We study entanglement estimation and verification in realistic situations, taking into account experimental imperfections and statistical fluctuations due to finite data. We consider both photonic and spin-1/2 systems. We study how entanglement created with mixed photon wave packets is degraded. We apply statistical analysis to and propose criteria for reliable entanglement verification and estimation. Finally we devote some effort to making quantum state estimation efficient by applying information criteria.

CURRICULUM VITAE

NAME OF AUTHOR: Jun Yin

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Fudan University, Shanghai

DEGREES AWARDED:

Doctor of Philosophy, Physics, 2011, University of Oregon, Eugene
Bachelor of Science, Physics, 2005, Fudan University, Shanghai

PUBLICATIONS:

J. O. S. Yin and S. J. van Enk, *Information criteria for efficient quantum state estimation*, accepted by Phys. Rev. A.

J. O. S. Yin and S. J. van Enk, *Criteria for reliable entanglement quantification with finite data*, Phys. Rev. A **83**, 022326 (2011).

R. Blume-Kohout, J. O. S. Yin and S. J. van Enk, *Entanglement verification with finite data*, Phys. Rev. Lett. **105**, 170501 (2010).

J. O. S. Yin and S. J. van Enk, *Entanglement and purity of one- and two-photon states*, Phys. Rev. A **77**, 062333 (2008).

ACKNOWLEDGEMENTS

Thanks go to the ones to whom I dedicate this work.

This work is dedicated to the ones to whom all my thanks go.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION 1

II. QUANTUM STATE ESTIMATION 5

III. QUANTUM ENTANGLEMENT 9

3.1 Introduction 9

3.2 Separable and Entangled States 11

3.3 Entanglement Monotones 12

3.4 Entanglement Verification 17

IV. ENTANGLEMENT AND PURITY OF SINGLE- AND TWO-PHOTON STATES 21

4.1 Introduction 21

4.2 Single- and Two-Photon Wave Packets: Preliminaries 24

4.3 Entanglement of Single-Photon States 29

4.4 Entanglement of Two-Photon States 38

4.5 Conclusions 49

Chapter Page

V. ENTANGLEMENT VERIFICATION WITH FINITE DATA 52

5.1 Likelihood Ratios 53
5.2 The Distribution of λ 54
5.3 Implementation 59
5.4 Examples 59
5.5 Witness Data 60
5.6 Tomographically Complete Data 61
5.7 Conclusions 62

VI. CRITERIA FOR RELIABLE ENTANGLEMENT QUANTIFICATION WITH FINITE DATA 64

6.1 Introduction 64
6.2 Comparisons Between Maximum Likelihood Estimation and Bayesian Methods 66
6.3 Preliminaries 69
6.4 How Many Measurements? 79
6.5 Conclusions 85

VII. INFORMATION CRITERIA FOR EFFICIENT QUANTUM STATE ESTIMATION 87

7.1 Introduction 87
7.2 The Akaike Information Criterion - A Schematic Derivation . . 90
7.3 Results 94

Chapter Page

7.4 Conclusions 111

VIII. CONCLUSIONS 113

APPENDICES

A. ENTANGLEMENT OF A COLOR-MIXED POLARIZATION-ENTANGLED STATE 116

B. ASYMPTOTIC BEHAVIOR OF THE EXPECTATION VALUE OF THE POSTERIOR DISTRIBUTION 119

C. ASYMPTOTIC BEHAVIOR OF THE MAXIMUM LIKELIHOOD ESTIMATION 124

REFERENCES CITED 125

LIST OF FIGURES

Figure Page

3.1 Entanglement witness 20

4.1 A single photon impinges on a 50/50 beam splitter 32

4.2 Two *identical* photons impinge on a 50/50 beam splitter 33

4.3 Two *distinguishable* photons impinge on a 50/50 beam splitter 33

4.4 Two distinguishable photons with orthogonal polarizations impinge on a 50/50 beam splitter 34

4.5 Entropy of entanglement of pure two-photon states 43

4.6 Logarithmic negativity of pure two-photon states 43

4.7 Logarithmic negativity of mixed two-photon states 45

4.8 Logarithmic negativities of various mixed output states 46

4.9 Logarithmic negativities for the output of a mixed two-photon state accompanied by vacuum 49

4.10 Logarithmic negativities for the output of mixed two-photon states . . 50

5.1 General schema of a likelihood ratio test 56

5.2 Loglikelihood ratios behave dramatically differently for different states 58

5.3 Distribution of λ for a SIC-POVM experiment 60

6.1 Two different prior distributions & GH prior distribution 72

6.2 The posterior distributions resulting from a pure and a mixed GH prior 73

6.3 A Z-prior based Metropolis-Hastings walk 82

6.4 Estimated probability distribution of MLE& bootstrap method against the posterior distributions with Z and GH priors 83

6.5 The difference between the estimations of $\langle N_{1,2} \rangle$ using the Z and GH priors 84

Figure Page

7.1 How AIC ranks the one- and two-parameter models vs. the full-parameter model (FPM) 97

7.2 Comparing single- and double-excitation Dicke states 99

7.3 How the AIC ranks our one-parameter model vs. the FPM for an entanglement witness measurement 102

7.4 Does the witness $W_{J_{xy}}$ detect entanglement if there is a phase error? . . 103

7.5 How one quantifies entanglement from a witness measurement 104

7.6 What fraction of the model Eq. (Equation 7.23.) describes physical states? 108

7.7 How the AIC ranks our two-parameter model vs. the FPM, for a witness measurement? 109

7.8 Quantifying entanglement from a witness measurement 110

7.9 Comparing one- and two-parameter models directly 111

CHAPTER I

INTRODUCTION

A quantum computer is revolutionarily different from a classical computer by exploiting some of the unique attributes of quantum mechanics, entanglement in particular. That is why a quantum computer can be used to simulate the evolution of quantum systems without experiencing a considerable slowdown like its classical counterpart[1,2]. Several quantum algorithms are known now that show a significant improvement in the use of time and other resources for solving particular tasks[3–7]. Among them the most famous one, which brought the whole field of quantum information processing (QIP) to the fore, is Shor's algorithm for factoring large integers[7].

A second branch of QIP is quantum cryptography, which promises to provide absolute security and privacy for communication, something which cannot be obtained within classical information theory[8].

Thus QIP is an intriguing modern subject of research due to its large potential, and by now a great variety of systems have been proposed as candidates for implementing quantum computing and/or quantum communication[9–15]. All those systems have their own technical obstacles, but they also face one common problem in realization, that of quantum decoherence. Decoherence turns quantum superpositions into incoherent mixtures, and these incoherent mixtures are classical in the sense that their information processing capabilities are no better than can be achieved classically. In short, decoherence destroys entanglement.

Decoherence tends to increase with the size of the system. Currently the largest system shown to demonstrate the potential of QIP (by containing entanglement between all systems) is a 14-partite ion-trap system[16].

Much effort has been devoted to characterizing entanglement as a resource in QIP, e.g., in quantum computation[9–11], quantum cryptography[17], quantum teleportation[18,19], etc. It is believed that entanglement is necessary for efficient quantum computing, requiring much less resources than classical algorithms or any quantum algorithm that does not utilize entanglement[20,21]. In this case, no rigorous proof for this statement exists (there cannot be such a proof until one proves the holy grail of complexity theory, that P$\neq$NP). On the other hand, in the case of quantum cryptography there is a proof that entanglement is necessary (although not quite sufficient) for absolute security[22].

It is thus of interest to verify the existence of entanglement in a given QIP experiment, to demonstrate the potential advantage over classical information processing[23]. A detailed introduction of entanglement and entanglement verification methods will be given in Chapter III.

Surprisingly, realistic entanglement quantification and verification have hardly been discussed in the literature. Theoretical treatments tend to assume several idealizations. We will take into account a few aspects left out of the discussion so far, such as the noise in measurements and statistical fluctuations due to finite data in entanglement verification.

The other important topic in this dissertation is quantum state estimation, which is the crucial step in predicting the future measurement outcomes from the past measurement outcomes. The problem can be formidable as the size of the system grows and the number of variables that need to be determined grows exponentially.

There are existing quantum state estimation schemes[24,25] that effectively reduced the exponential scaling while some prior knowledge is required. Following the same spirit we propose a method that is well developed in statistics. The method works when only very limited data is possible and allows in principle any degrees of freedom to be tested therefore scaling is no longer a problem. The method can be extremely convenient for the application of state estimation in quantum information processing.

This dissertation is organized as follows: in Chapter II we go over the concepts and the brief history of development in quantum state estimation. In Chapter III we introduce the concepts of entanglement and separability and elaborate their definitions for different systems. We also talk about several entanglement verification schemes.

In Chapter IV we study how entanglement created with mixed photon wave packets is degraded (in this context, the spectral degrees of freedom had always been implicitly assumed to be pure). We find in particular that the entanglement of a delocalized single-photon state of the electromagnetic field is determined simply by its purity. We also discuss entanglement for two-photon mixed states, as well as the influence of a vacuum component.

Chapters V and VI are focused on the statistical properties in entanglement verification/estimation with finite measurements. Since no experiment is able to produce infinite sequence of data, it is necessary to use the proper statistical language. In Chapter V we propose a reliable method to quantify the weight of evidence for (or against) entanglement, based on a likelihood ratio test. Our method is universal in that it can be applied to any sort of measurements. We demonstrate the method by applying it to two simulated experiments on two qubits. The first measures a so-called entanglement witness, while the second performs a tomographically

complete measurement. In Chapter VI we propose criteria for determining how many measurements are needed to quantify entanglement reliably. We base these criteria on Bayesian analysis of measurement results and apply our methods to four-qubit entanglement, but generalizations to more qubits are straightforward.

Since one popular method for entanglement verification is based on quantum state reconstruction, we also study ways to make quantum state tomography efficient. In Chapter VII we propose a method that allows us to tentatively use in principle any model for the state, using any (small) number of parameters (which can, e.g., be chosen to have a clear physical meaning), and the data are used to verify the model. The proof that this method is valid is based on well-established statistical methods that go under the name of "information criteria." We exploit here, in particular, the Akaike Information Criterion. We illustrate the method by simulating experiments on (noisy) Dicke states of four qubits.

We summarize the results of this dissertation in Chapter VIII.

Chapter IV, VI and VII were published and co-authored with S. J. van Enk. Chapter V was published and co-authored with R. Blume-Kohout and S. J. van Enk.

CHAPTER II

QUANTUM STATE ESTIMATION

In quantum mechanics, states of systems are mathematically described by density matrices. When each element of the density matrix is determined, the future evolution of and the future measurement on the system can be precisely predicted, which is a good feature for any implementation in the real world. State estimation is the process where one infers, generally from past measurement outcomes, as much information as possible about the density matrix. Due to the fact that almost any arbitrary measurement (without prior knowledge) will disturb the system, it is in principle impossible to recover the quantum state of a single system[26]. Therefore an ensemble of independently and identically prepared systems is necessary and repetitive measurements on the systems are required. This standard procedure is called *quantum state tomography*. The first systematic approach to implement quantum tomography was proposed by Fano[27]. In quantum optics, one technique to recover each element of a density matrix is called *optical homodyne tomography* proposed by Vogel and Risken[28]. This technique associates measurement observations with the Wigner function[29–31] through which the density matrix can be reconstructed. The Wigner function is the quasi-probability distribution in the phase space of position and momentum quadratures that connects to the density matrix representation of quantum states. In one dimension, the Wigner function is defined as

$$W(x,p) = \frac{1}{\pi\hbar}\int_{-\infty}^{+\infty} \mathrm{d}x' e^{2ipx'/\hbar} \left\langle x - x' \middle| \rho \middle| x + x' \right\rangle , \tag{2.1}$$

where ρ is the density matrix representation (in the $\{|x\rangle\}$ basis) of the quantum state, and $|x \pm x'\rangle$ is the vector state corresponding to the eigenstates of the position operator $\hat{\mathbf{x}}$ with eigenvalues $x \pm x'$. Given the Wigner function of a quantum state ρ, the expectation value of any operator $\hat{G}$ in that state can be expressed as the overlap between the Wigner transformation of $\hat{G}$ and the Wigner function. That is

$$\left\langle \hat{G} \right\rangle = \mathrm{Tr}\left(\rho \hat{G}\right) = \int_{-\infty}^{+\infty} \mathrm{d}x \int_{-\infty}^{+\infty} \mathrm{d}p W(x,p) G(x,p), \tag{2.2}$$

where $W(x,p)$ is the Wigner function of state ρ as in Eq. (2.1) and $G(x,p)$ is the Wigner transformation of operator $\hat{G}$:

$$G(x,p) = \int_{-\infty}^{+\infty} \mathrm{d}x' e^{2ipx'/\hbar} \left\langle x - x' | \hat{G} | x + x' \right\rangle. \tag{2.3}$$

Thus the Wigner function plays an important role in quantum mechanics, analogous to that of the Louiville density operator in classical statistical mechanics.

The first measurements of the Wigner function were soon realized in a series of experiments by Raymer's group at the University of Oregon[32,33] using optical homodyne tomography. Quantum state tomography has since been demonstrated in a wide variety of systems[34–46].

However, measurement observations contain inevitable statistical fluctuations due to the finiteness of data. Therefore statistical inferences are to be applied for an appropriate reconstruction of quantum states from measurement observations. These methods seek to extract the maximum information about what the quantum state is from the finite data based on different principles in statistics. Consider a generic situation where measurements are denoted by $M_j, \quad j = 1, 2, ...$, where the subscript j denotes the different outcomes. A tomography, or a tomographically complete

measurement, indicates that the reconstruction of state ρ is possible in terms of the frequencies f_j of measurement outcomes M_j. According to Born's rule, the probability of getting measurement outcome M_j in state ρ is

$$p_j = \mathrm{Tr}\left(\rho M_j\right). \tag{2.4}$$

Under any circumstance an *inverted* density matrix exists through the inversion of Eq. (2.4) if the observed frequency f_j is interpreted as the probability p_j. Nevertheless at this stage the inverted density matrix ρ is not necessarily a physical state, especially when the number of measurements is small. Thus further steps are required. The inverted density matrix is in fact the state that maximizes the *likelihood* functional with respect to the observations. The likelihood functional of a state ρ with respect to observation frequencies f_j is defined as

$$Ł(\rho|\{f_j\}) = \prod_j \left(\mathrm{Tr}\rho M_j\right)^{f_j}. \tag{2.5}$$

The maximum likelihood principle[47] assumes that the best estimated state ρ_{MLE} is the *physical* state that maximizes the likelihood functional Eq. (2.5). That is, a single estimate ρ_{MLE} that best describes the current observations f_j. In contrast, the Bayesian method takes on a slightly different perspective where it explicitly and systematically emphasizes the existence of a prior distribution of states ρ and insists that any measurement observations only produce a posterior distribution of states (hence not a single estimate)[48,49]. In other words, the measurement observations work in such a way that they transform the prior distribution $P_{\mathrm{prior}}(\rho)$ to a posterior

distribution $P_{\text{posterior}}(\rho)$ through the likelihood functional:

$$P_{\text{posterior}}(\rho) \ \sim \ P_{\text{prior}}(\rho) \times \text{Ł}(\rho|\{f_j\}). \tag{2.6}$$

Chapter VI applies this method as an intermediate step in entanglement estimation. In principle, the number of variables to be determined in quantum state estimation scales exponentially with e.g. the number of qubits, which is a formidable task when the number of qubits is large. In Chapter VII we discuss how quantum state estimation can be made efficient using information criteria.

CHAPTER III

QUANTUM ENTANGLEMENT

This chapter provides the preliminary knowledge for a smooth accessibility to the contents of Chapters IV-VII, which are dedicated to a specific topic each. In this chapter, we give a brief introduction to quantum entanglement. We will define what separable states and entangled states are and touch upon the complications for multipartite entanglement. We will also talk about the quantification of entanglement and the verification of entanglement.

3.1 Introduction

Quantum entanglement was an interesting topic[50,51] well before its employment in QIP, for it challenges two fundamental principles, namely locality and realism, held for granted in classical physics. Bell[52,53] quantifies this, by showing that any correlations within a classical model that includes both locality and realism at the same time, must obey his famous inequalities. Quantum mechanics can violate these inequalities, but only with the help of entanglement. Thus a violation of these inequalities serves as a sufficiency proof for entanglement. More verification schemes were proposed after entanglement became more appealing in QIP. Among these some schemes involve a reconstruction of the whole density matrix, so that the amount of entanglement can be fairly estimated through, e.g., the positive partial transpose (PPT) criterion (explained below)[54,55]. Other schemes that do not require state reconstruction are usually designed for a specified class of states so that the number of different measurements needed is greatly reduced. The typical examples are so-called entanglement witnesses (explained below)[55]. Yet other schemes claim to measure

entanglement directly in the sense that a certain entanglement measure comes directly from the measurement outcomes[56]. A review on entanglement verification can be found in Ref.[57]. Especially a more thorough study of bipartite entanglement can be found in Ref.[58]. Quantum teleportation using bipartite entanglement was demonstrated using photons with post-selection[59,60], with two-mode squeezed light beams[61], and ions[62,63].

Bipartite entanglement is the simplest form of entanglement. Multipartite entanglement refers to entanglement in a system that contains three or more subsystems and it turns out to be much more complicated than bipartite entanglement. It becomes of crucial importance for the future of QIP on large scales. Interestingly in this case, even pure states can be categorized into different entanglement classes in a way that states belonging to different classes cannot be transformed into each other via local operation and classical communication (LOCC). Such cases have been studied in three-[64] and four-[65] qubit systems. Multipartite entanglement for mixed states are much more complicated and remains a heated topic. One of the most challenging tasks is to identify genuine multipartite entanglement, which means there is absolutely no way of writing a state as a convex sum of bipartite pure states (*bipartite* refers to in general any bi-partition of a multipartite system). Experimentally multipartite entanglement was demonstrated using photons (with post-selection)[66–70], and ions[71–74].

3.2 Separable and Entangled States

Given two systems A and B, the density matrices of all bipartite ($n = 2$) separable states are defined as

$$\rho = \sum_j p_j \rho_A^j \otimes \rho_B^j, \tag{3.1}$$

where $p_j \geq 0$, $\forall\ j$ and $\sum_j p_j = 1$. On the other hand, entangled states are defined as states whose density matrices do **not** bear a form of Eq. (3.1). For example, it can be shown that the density matrix of the singlet state

$$\left|\Psi^-\right\rangle = \frac{1}{\sqrt{2}} \left(|01\rangle - |10\rangle\right) \tag{3.2}$$

cannot be written as a separable state in the form of Eq. (3.1): tracing out one system, leaves the other in a mixed state, which could not happen for a pure product state. For $n > 2$ parties, the definition of separability is richer simply because there are more ways of partitioning the overall system. The fully separable states are of the form

$$\rho = \sum_j \left(p_j \prod_k^{\otimes n} \rho_k^j \right), \tag{3.3}$$

where $p_j \geq 0$, $\forall\ j$ and $\sum_j p_j = 1$. ρ_k^j's are all single party density matrices. At the other extreme we have genuine multipartite entanglement, which only exists in the states that do not have a decomposition of bi-separable form:

$$\rho = \sum_j p_j \rho_{n_j} \otimes \rho_{\overline{n_j}}, \tag{3.4}$$

where $p_j \geq 0$, $\forall\ j$ and $\sum_j p_j = 1$. $n_j|\overline{n_j}$ denote a certain bi-partition of the n parties and ρ_{n_j} and $\rho_{\overline{n_j}}$ are the corresponding density matrices. Whereas the separability problem is solved for 2×2 and 2×3 systems[55], it remains a formidable task to prove the existence of a state decomposition like Eq. (3.3) or even Eq. (3.4) for larger systems. In this work, the multipartite entanglement verification is aimed at identifying the states that are bi-separable, i.e.,

$$\rho = \sum_j p_j \rho_m^j \otimes \rho_{\overline{m}}^j, \tag{3.5}$$

only corresponding to one single bi-partition $m|\overline{m}$. Apparently bi-separability is somewhere in between full separability and genuine multipartite entanglement. For instance, a state like $\rho_{AB} \otimes \rho_C + \rho_{BC} \otimes \rho_A + \rho_{CA} \otimes \rho_B$ is not genuinely multipartite entangled nor bi-separable generically. It is thus identified as a *entangled* state without further specification. With the existing theory concerning multipartite separability, however, we argue that the method in this work is good enough in most situations.

3.3 Entanglement Monotones

Quantification of entanglement is one of the main issues concerning entanglement detection. For a density matrix ρ, the quantity $E(\rho)$, called **entanglement monotone**, needs to satisfy the following conditions[75]:

1. $E(\rho) = 0$ when ρ is a separable state and $E(\rho) > 0$ when ρ is an entangled state. As a convention, $E(|\Psi_-\rangle) = 1$.

2. E does not increase **on average**, under any **local operation and classical communication (LOCC)**. It means that any individual manipulation of the

state that are connected with only classical communication can never induce quantum correlation, if not reduce it.

3. Assume measurement is performed on the state ρ and the outcomes ρ_j are obtained with probability p_j, then $E(\rho) \geq \sum_j p_j E(\rho_j)$. Note that usually $\rho \neq \sum_j p_j \rho_j$.

4. E is convex, i.e., $\sum_j p_j E(\rho_j) \geq E(\sum_j p_j \rho_j)$. The process of $\sum_j p_j \rho_j$ can be regarded as discarding information.

For bipartite systems, a few outstanding candidates for entanglement monotones based on the knowledge of the whole density matrix are entanglement of formation[76], geometric measure of entanglement[77–79], concurrence[80], distance measure[81] and negativity[82]. Among all these entanglement monotones, the negativity is most straightforward to be generalized to higher dimensions and multiple parties, and is always computable given the density matrices.

3.31 Concurrence and Beyond

Given a bipartite density matrix ρ, concurrence is define as

$$\mathcal{C}(\rho) \equiv \max\left(0, \sqrt{\lambda_1} - \sqrt{\lambda_2} - \sqrt{\lambda_3} - \sqrt{\lambda_4}\right), \tag{3.6}$$

where $\lambda_{1,2,3,4}$ are the eigenvalues of

$$\rho(\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)$$

in decreasing order and σ_y is the Pauli y-matrix. It is related with **entanglement of formation** $E_F(\rho)$[76] as

$$E_F(\rho) = H\left(\frac{1+\sqrt{1-\mathcal{C}(\rho)^2}}{2}\right), \tag{3.7}$$

where

$$H(x) = -x\log_2 x - (1-x)\log_2(1-x). \tag{3.8}$$

Entanglement of formation is also an entanglement monotone and is defined as the amount of the entanglement needed to form a state using LOCC. It is, however, not directly computable for mixed states.

Multipartite concurrence were also proposed[83]. Developed from concurrence, 3-tangle[84] as an entanglement monotone of a pure state of three parties A,B and C is defined as

$$\tau_{ABC} = \mathcal{C}^2_{A(BC)} - \mathcal{C}^2_{AB} - \mathcal{C}^2_{AC}. \tag{3.9}$$

$\mathcal{C}_{A(BC)}$ is the concurrence of state ρ_{ABC} when treat as a bipartite state under the partition of A and BC. $\mathcal{C}_{AB}$ is the concurrence of the reduced density matrix $\rho_{AB} = \mathrm{Tr}_C\rho_{ABC}$. However it does not fully characterizes the genuine tripartite entangled states like W-state

$$|W\rangle = \frac{1}{\sqrt{3}}\left(|001\rangle + |010\rangle + |100\rangle\right), \tag{3.10}$$

since $\tau(\rho_W) = 0$. It turns out that three-qubit pure states can be entangled in two inequivalent ways in the sense of SLOCC[85] and are thus categorized by two classes

represented by W-state and GHZ-state

$$|GHZ\rangle = \frac{1}{\sqrt{2}}\left(|000\rangle + |111\rangle\right). \tag{3.11}$$

3.32 Negativity and Bound Entanglement

Given a bipartite density matrix ρ, the negativity is define as

$$\mathcal{N}(\rho) = \left|\left|\rho^{\Gamma}\right|\right|_1 - 1, \tag{3.12}$$

where $||\cdot||_1$ stands for the sum of the absolute values of the eigenvalues. ρ^{Γ} is the partial transpose[54,86] of ρ, where the submatrices of one of the two systems are transposed (since the operation is invariant under the switching of the systems it does not matter which one we pick) , e.g.,

$$\rho = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/2 & -1/2 & 0 \\ 0 & -1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \longrightarrow$$

$$\rho^{\Gamma} = \begin{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1/2 \end{pmatrix}^T & \begin{pmatrix} 0 & 0 \\ -1/2 & 0 \end{pmatrix}^T \\ \begin{pmatrix} 0 & -1/2 \\ 0 & 0 \end{pmatrix}^T & \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \end{pmatrix}^T \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -1/2 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ -1/2 & 0 & 0 & 0 \end{pmatrix}.$$

$||M||_1$ is the sum of the absolute value of all the eigenvalues of matrix M. In this case, $\left|\left|\rho^{\Gamma}\right|\right|_1 = |1/2| + |1/2| + |1/2| + |-1/2| = 1$. Essentially if the positivity is eliminated by the partial transpose, then $\mathcal{N}$ ends up with a positive value. Since all

separable states can be written as Eq. (3.1), the partial transpose

$$\rho^{\Gamma} = \sum_j p_j \rho_j^A \otimes \left(\rho_j^B\right)^T . \tag{3.13}$$

It is still a physical state therefore still positive semi-definite. It turns out that for 2x2 and 2x3 systems, a **positive partial transpose (PPT)** is a necessary and sufficient condition for separability[55]. For all other cases, there exist entangled states with PPT, which are referred to as bound entangled states and their entanglement cannot be distilled[87].

Negativity for multipartite qubit system is a generalized idea of bipartite negativity[88,89]. Given a n-qubit density matrix ρ and a bi-partition of the qubits

$$1, ..., k|k+1, ..., n,$$

the negativity of ρ *with respect to this partition* is defined as

$$\mathcal{N}_{1,...,k|k+1,...,n} = \left|\left|\rho^{\Gamma_{1,...,k|k+1,...,n}}\right|\right|_1 - 1.$$

To make things easier, we group together the negativities with the same number of elements in the subgroups of the bi-partitions. For example, the negativity for a three-qubit state is

$$\mathcal{N} = \left(\mathcal{N}_{1|23}\mathcal{N}_{2|31}\mathcal{N}_{3|12}\right)^{1/3} . \tag{3.14}$$

The negativities for a four-qubit state are

$$\mathcal{N}_{1-3} = \left(\mathcal{N}_{1|234}\mathcal{N}_{2|341}\mathcal{N}_{3|412}\mathcal{N}_{4|123}\right)^{1/4},$$
$$\mathcal{N}_{2-2} = \left(\mathcal{N}_{12|34}\mathcal{N}_{13|24}\mathcal{N}_{14|23}\right)^{1/3}, \tag{3.15}$$

etc. If any single negativity vanishes with respect to a bi-partition, then the two subparts according to that bi-partition are separable thus the state itself must be bi-separable. As mentioned in Section 3.2 a separable state such as $\rho = \rho_{AB} \otimes \rho_C + \rho_{BC} \otimes \rho_A + \rho_{CA} \otimes \rho_B$ is generically not bi-separable thus identified as an "entangled state" by this definition of generalized negativity. However with the existing theory of multipartite entanglement verification, this entanglement monotone works most conveniently with the density matrices and serves the verification purpose ideally in the cases we are interested in (more in Chapter VI), although it does leave out *bound entangled states*, which have positive partial transposes.

3.4 Entanglement Verification

As mentioned above, there are two main types of entanglement verification schemes: one that requires state reconstruction and one that does not. In the first type, everything about the state is determined after tomography and it is thus straightforward to calculate any desired entanglement measure from the density matrix. However, it becomes more and more demanding to perform a reliable tomography as the number of qubits increases. On the other hand, entanglement witnesses provide a highly efficient way of deciding the separability of a state. In the case of linear witnesses only one observable needs to be measured, although the observable has to be carefully designed beforehand to save the toil of reliable

tomography. The expectation value of the witness does not correspond directly to the amount of entanglement, which in most cases are not well-defined. However, the expectation value of the witness does indicate the strength of the entanglement, in the similar way the violation of the Bell's inequalities works for entanglement verification, which can be considered as a special class of entanglement witnesses. It is also ideally possible to measure concurrence directly[90], but it requires measurements carried out on the tensor product of independent and identical copies[91], an assumption too strong to be realistic[92].

3.41 Entanglement Witnesses

Witnesses are always a handy tool for entanglement verification, especially in most cases when experimentalists have a fairly good idea of the states they produce, like a GHZ-state[16,73], a Dicke state[93] or a W-state[94], which are usually the most highly entangled states (HES) for their own kinds. A large enough fidelity with respect to these states indicates entanglement, regardless of the types of noises. Therefore a straightforward linear witness can be expressed as

$$\mathcal{W} = \alpha\mathcal{I} - |\Psi_{\text{HES}}\rangle \langle\Psi_{\text{HES}}| , \tag{3.16}$$

where α is to be determined so that all the negative expectation of $\mathcal{W}$ infers an entangled state in certainty:

$$\text{Tr}(\rho\mathcal{W}) < 0 \Longrightarrow \rho \text{ is entangled.}$$

Note that a linear witness always leaves some entangled states undetected, due to the convexity of the set of separable states. However, for any entangled state there always

exists a linear witness that detects it, as guaranteed by **Hahn-Banach theorem**. An example of linear witness is shown in Chapter V for the likelihood ratio test. To improve the performance of a linear witness, beside optimizing the witness itself[95], one can also construct nonlinear witness[96,97] to accommodate the convexity of the set of separable sets (FIG. 3.1). The fundamental idea is that any states to be detected by a particular witness needs to be close enough to a highly entangled state, which is aimed at by the witness. Other witnesses, not directly measuring the overlap with a highly entangled state, are introduced in the consideration of the local decomposition of measurements. In Chapter VI we will see for the Dicke state[98]

$$\left|D_4^2\right\rangle = \left(|0011\rangle + |0101\rangle + |0110\rangle + |1001\rangle + |1010\rangle + |1100\rangle\right)/\sqrt{6}, \tag{3.17}$$

an effective witness is[99]

$$W_{J_{xy}} = 7/2 + \sqrt{3} - J_x^2 - J_y^2, \tag{3.18}$$

where $J_{x,y} = \sum_j \sigma_{x,y}^{(j)}/2$, with $\sigma_{x,y}^{(j)}$ the Pauli matrices for the j-th subsystem. ρ is Dicke-like entangled if $\langle \rho W_{Jxy} \rangle < 0$.

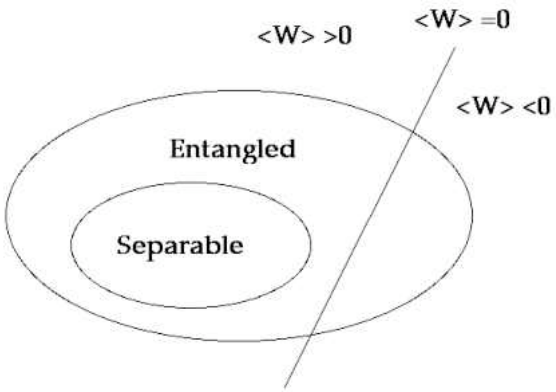

FIGURE 3.1 Entanglement witness

CHAPTER IV

ENTANGLEMENT AND PURITY OF SINGLE- AND TWO-PHOTON STATES

This work was published as *Entanglement and purity of one- and two-photon states*, Phys. Rev. A **77**, 062333 (2008). It was initiated by S. J. van Enk and finished jointly by J. O. S. Yin and S. J. van Enk.

4.1 Introduction

Consider an entangled state containing one or more photons. By how much is the entanglement degraded when the photon wave packets are described by mixed rather than pure states? For example, suppose Alice and Bob are given a two-photon *polarization-entangled* state[100–104]—say the singlet state—but they are not told what the color of the photons is. All they know is the photons are either both blue or both green. Does their ignorance reduce the amount of entanglement they possess? The answer in this case is negative: the entanglement is still one ebit, even though the overall state is mixed. After all, they could, in principle at least, apply a *local* measurement on each photon that measures its color but not its polarization. That way they are guaranteed to end up with a pure maximally entangled state. The fact that polarization and color are independent degrees of freedom is crucial here.

Suppose now Alice and Bob are given a *mode-entangled* state[105–107] containing merely a single delocalized photon[108–113]

$$|\psi\rangle := \frac{|0\rangle_A |1\rangle_B + |1\rangle_A |0\rangle_B}{\sqrt{2}}, \tag{4.1}$$

where A and B denote specific modes in Alice's and Bob's labs and $|0\rangle$ and $|1\rangle$ denote Fock states with zero and one photon, respectively. The notation used implies that modes A and B are well-defined. But suppose that Alice and Bob actually do not know what the color of the single delocalized photon is, green or blue (each with 50% probability). Or suppose they do not know the polarization of the photon, only that it is either left-hand or right-hand circularly polarized. Then they really should ascribe a mixed state to their field modes: an equal mixture of $|\psi\rangle$ and the similar state

$$|\psi'\rangle := \frac{|0\rangle_{A'}|1\rangle_{B'} + |1\rangle_{A'}|0\rangle_{B'}}{\sqrt{2}}, \tag{4.2}$$

where the primed modes refer to modes of different color or different polarization. In Section 4.3 we will find the logarithmic negativity[82,114] of this mixed state to be $E_{\mathcal{N}} = \log_2\left(1+\sqrt{1/2}\right) < 1$, so that in this case Alice's and Bob's state does lose some of its glamor [a pure state of the form (4.1) would contain one ebit of entanglement]. Note that, indeed, Alice and Bob cannot use the same *local* filtering measurement of frequency to filter their state to a pure entangled state: as soon as a photon is detected on, say, Bob's side, the state collapses to either $|0\rangle_A|1\rangle_B$ or $|0\rangle_{A'}|1\rangle_{B'}$, and not to the desired pure entangled state $|\psi\rangle$ or $|\psi'\rangle$. Alternatively, a *nonlocal* filtering measurement of color could upgrade Alice's and Bob's state to a pure entangled state, but that nonlocal operation could and would increase the amount of entanglement. The distinguishing feature of this example compared to the previous one is that color or polarization and mode are dependent degrees of freedom. More precisely, color and polarization are part of what defines a mode.

The purpose of this chapter is to continue investigating questions of this sort: by how much is the entanglement of single- or two-photon states degraded when the

photon wave packets are not pure but mixed? See Figs. 4.1–4.4 for typical examples of questions considered in the present chapter. The motivation for this research is, of course, the simple fact that typically any photon produced in an experiment is represented by a mixed state[104,115–117]. For instance, even if one's source produces a Fourier-limited wave packet, in practice one will not know exactly the timing of the wave packet, or the exact central frequency, or the exact width. For another example, consider a single photon heralded by detection of the other photon of a down-converted pair of photons. Whenever there is some entanglement between the two photons, tracing out one photon necessarily leaves the remaining photon in a mixed state.

In Section 4.2 we start out by collecting some useful results about the description of single- and two-photon wave packets to be used in later sections. In Section 4.3 we focus our attention on entangled states that can be generated by splitting a single photon on a 50/50 beam splitter. Ideally that leads to an output state with one ebit of entanglement, but, as we will show, the entanglement resulting from splitting a *mixed* single-photon input state is less than one ebit but turns out to be a simple function of its purity. In the same Section we will also consider the more realistic case of a nonzero vacuum component of the state of the field and its effect on entanglement. In Section 4.4 we consider typical mixed states of two orthogonally polarized photons arising from type-II down conversion, and in that case too we consider the effects of the presence of a (large) vacuum component. We also compare the mode entanglement one obtains by splitting both photons on a 50/50 beam splitter with the entanglement that may already be present in the input state between two orthogonally polarized modes, and find one may increase the amount of entanglement that way. We conclude

the chapter with a summary that also discusses some possible extensions of the present work.

4.2 Single- and Two-Photon Wave Packets: Preliminaries

4.21 Single Photons

Consider a single photon of a definite polarization propagating in a well-defined direction. Then a pure state can be described in terms of continuous modes[118] as

$$|1_\psi\rangle = \int \mathrm{d}t \widetilde{\psi}(t) a^\dagger(t) \, |\mathrm{v}\rangle \, , \tag{4.3}$$

where $|\mathrm{v}\rangle$ is the vacuum state, $a^\dagger(t)$ an operator that creates a photon at time t and $\widetilde{\psi}(t)$ the temporal mode function of the wave packet. Often it is more useful to Fourier transform this representation into frequency space, and describe the same state by

$$|1_\psi\rangle = \int \mathrm{d}\omega \psi(\omega) a^\dagger(\omega) \, |\mathrm{v}\rangle \, . \tag{4.4}$$

Typically, the function $\psi(\omega)$ will be appreciable only in a small bandwidth σ around a central frequency ω_0, with $\sigma \ll \omega_0$, and the wave packet (4.4) describes a quasi-monochromatic photon. Since the creation operators bear the relation

$$a^\dagger(t) = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}\omega a^\dagger(\omega) e^{-i\omega t}, \tag{4.5}$$

the spectral shape $\psi(\omega)$ is thus completely determined by the temporal mode $\widetilde{\psi}(t)$ in the sense that

$$\psi(\omega) = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}t \widetilde{\psi}(t) e^{-i\omega t}. \tag{4.6}$$

The single-photon states displayed so far are pure. Mixed states of single photons arise, e.g., when they are part of a multipartite system and we trace over the other parties; or if one of the transverse degrees of freedom of the photon is traced out; or if we are simply ignorant about one or more of the properties of the photon. In any case the mixed state of a single photon is described in terms of the density matrix

$$\rho_1 = \int \mathrm{d}\lambda P(\lambda) \left|1_{\psi_\lambda}\right\rangle \left\langle 1_{\psi_\lambda}\right|. \tag{4.7}$$

Here λ stands for any parameter or combination of parameters that can possibly be involved in the mode function, although no concrete form is given yet. Some good examples of what λ could stand for are arrival time ("time jitter"), central frequency ("frequency jitter"), the width of the (Fourier-limited) wave packet, etcetera[117]. When writing the mixed state in terms of parameters with such a clear physical meaning, the states $|1_{\psi_\lambda}\rangle$'s will in general not be orthogonal. However, we can always diagonalize the density matrix, and rewrite it as a discrete sum involving orthogonal states,

$$\rho_1 = \sum_k p_k \left|1_k\right\rangle \left\langle 1_k\right|. \tag{4.8}$$

Here p_k and $|1_k\rangle$ are eigensolutions to

$$\rho_1 |1_k\rangle = p_k |1_k\rangle \tag{4.9}$$

with

$$\langle 1_j |1_k\rangle = \delta_{jk}, \tag{4.10}$$

and $\sum_k p_k = 1$. For our purpose of calculating entanglement of single-photon states, the latter representation is often more useful.

4.22 Two Photons

Now consider states of exactly two photons. Since we have in mind the two-photon component of the state produced by type-II down conversion[115,119,120], we assume the photons have orthogonal polarizations. We will simply indicate the ordinary and extraordinary polarizations by "H" and "V". Pure and mixed states for such photon pairs can then be expressed as

$$|2_\psi\rangle = \int \mathrm{d}\omega \int \mathrm{d}\omega' \psi(\omega, \omega') a_H^\dagger(\omega) a_V^\dagger(\omega') |\mathrm{v}\rangle \,, \tag{4.11}$$

and

$$\rho_2 = \int \mathrm{d}\lambda P(\lambda) |2_{\psi_\lambda}\rangle \langle 2_{\psi_\lambda}| \,, \tag{4.12}$$

respectively. The commutators of the mode operators are

$$[a_k(\omega), a^\dagger_{k'}(\omega')] = \delta_{k,k'}\delta(\omega - \omega'), \tag{4.13}$$

where $\{k, k'\} = \{H, V\}$.

It is well-known that for a pure state of a bipartite system there is always a discrete Schmidt decomposition. In our case this allows us to rewrite

$$|2_\psi\rangle = \sum_k \sqrt{\lambda_k} h^\dagger_k v^\dagger_k |\mathrm{v}\rangle . \tag{4.14}$$

For a state like Eq. (4.11) we can explicitly define the new creation operators as

$$h^\dagger_k = \int \mathrm{d}\omega \varphi_k(\omega) a^\dagger_H(\omega), \tag{4.15a}$$

$$v^\dagger_k = \int d\omega \phi_k(\omega) a^\dagger_V(\omega), \tag{4.15b}$$

where λ_k, φ_k and ϕ_k can be obtained by solving the eigenvalue problems[121,122]

$$\int \mathrm{d}\omega' \widetilde{\rho}_A(\omega, \omega')\varphi_k(\omega') = \lambda_k \varphi_k(\omega), \tag{4.16a}$$

$$\int \mathrm{d}\omega' \widetilde{\rho}_B(\omega, \omega')\phi_k(\omega') = \lambda_k \phi_k(\omega), \tag{4.16b}$$

with the "reduced density matrices" given by

$$\widetilde{\rho}_A(\omega, \omega') = \int d\omega'' \psi(\omega, \omega'')\psi^*(\omega', \omega''), \tag{4.17a}$$

$$\widetilde{\rho}_B(\omega, \omega') = \int d\omega'' \psi(\omega'', \omega)\psi^*(\omega'', \omega'). \tag{4.17b}$$

The mode operators $h_k^\dagger$'s and $v_k^\dagger$'s satisfy the standard commutation relations

$$[v_j, h_k^\dagger] = 0, \quad [v_j, v_k^\dagger] = [h_j, h_k^\dagger] = \delta_{jk}. \tag{4.18}$$

4.23 Mode Entanglement

In the rest of the chapter we will calculate mode entanglement between field modes[105–107] that are spatially separated[123]. For that purpose we expand the density matrix in an appropriate orthonormal basis of Fock states, including states with no photons, single-photon states $|1_k\rangle$ as defined above, etc.

We use two standard measures of entanglement in this chapter: one useful entanglement measure, but only valid for pure bipartite states, is the entropy of entanglement[76]. For example the mode entanglement between modes of orthogonal polarization of the pure state $|2_\psi\rangle$ of Eq. (4.11) is defined in terms of the Schmidt coefficients appearing in Eq. (4.14) as

$$E(|2_\psi\rangle) = -\sum_k \lambda_k \log_2 \lambda_k. \tag{4.19}$$

The other entanglement monotone we will make use of is the logarithmic negativity[82,114], defined in terms of the partial transpose (PT) of a matrix. For example, for a two-photon system PT is easily defined if we expand its density matrix in the basis spanned by $|k\rangle_1 |l\rangle_{21} \langle m|_2 \langle n|$, with $|k\rangle_1$, $|l\rangle_2$, $|m\rangle_1$ and $|n\rangle_2$ referring to orthogonalized single-photon states respectively. Then we have

$$\rho \xrightarrow{PT} \rho^\Gamma, \quad \rho^\Gamma_{klmn} = \rho_{knml},$$

ρ_{klmn} being the matrix element. We then solve for the eigenvalues of ρ^Γ, ending up with a series of real e_k's, since both ρ and ρ^Γ are Hermitian. The logarithmic negativity[82,114] is defined as

$$E_{\mathcal{N}}(\rho) = \log_2 \|\rho^\Gamma\|_1 = \log_2 \sum_k |e_k|. \tag{4.20}$$

It is worthwhile to notice that the absolute sum of all eigenvalues $\sum_k |e_k|$ equals 1 plus twice the absolute value of the sum of all negative eigenvalues of ρ^Γ.

4.3 Entanglement of Single-Photon States

4.31 No Vacuum Component

4.311 General Results

A pure single-photon state split on a 50/50 beam splitter looks like

$$\frac{1}{\sqrt{2}}(|0\rangle\,|1\rangle + |1\rangle\,|0\rangle) \tag{4.21}$$

and possesses exactly one ebit of entanglement[110,111,113]. If the photon entering one input port of the beam splitter is mixed, the calculation of entanglement in the output state is a little more complicated, but the problem can still be solved analytically. We consider single polarization only and start by an input state that is already expanded in its diagonal form

$$\rho_{\text{in}} = \sum_{k=1}^{n} p_k\,|1_k\rangle\,\langle 1_k|\,. \tag{4.22}$$

We put it on a 50/50 beam splitter, which in the Heisenberg picture transforms the mode operators for the input ports $a^\dagger, b^\dagger$ into those for the output $c^\dagger, d^\dagger$ in the following way

$$\begin{pmatrix} a^\dagger \\ b^\dagger \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} c^\dagger \\ d^\dagger \end{pmatrix}. \tag{4.23}$$

In this picture the state stays unchanged, but when expressed in terms of creation operators at the output ports it looks different:

$$\begin{aligned} \rho_{\text{out}} = \rho_{\text{in}} =& \sum_{k=1}^{n} p_k \left|1_k\right\rangle_a \otimes_a \left\langle 1_k\right| \\ =& \frac{1}{2} \sum_{k=1}^{n} p_k (\left|1_k\right\rangle_c \left|0\right\rangle_d + \left|0\right\rangle_c \left|1_k\right\rangle_d) \otimes ({}_c\!\left\langle 1_k\right|_d \left\langle 0\right| + {}_c\!\left\langle 0\right|_d \left\langle 1_k\right|) \\ =& \sum_{k=1}^{n} \rho_k, \end{aligned} \tag{4.24}$$

where each submatrix ρ_k can be written out as

$$\rho_k = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & p_k/2 & p_k/2 & 0 \\ 0 & p_k/2 & p_k/2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.25}$$

Here rows and columns correspond to states $|0\rangle_c|0\rangle_d$, $|0\rangle_c|1_k\rangle_d$, $|1_k\rangle_c|0\rangle_d$, $|1_k\rangle_c|1_k\rangle_d$ and their conjugates respectively. Naively, taking the PT simply gives

$$\rho_k^\Gamma = \begin{pmatrix} 0 & 0 & 0 & p_k/2 \\ 0 & p_k/2 & 0 & 0 \\ 0 & 0 & p_k/2 & 0 \\ p_k/2 & 0 & 0 & 0 \end{pmatrix}. \tag{4.26}$$

In the total density matrix context, the two $p_k/2$ on the diagonal remain independent of other ρ_k's. These give rise to nonnegative eigenvalues of ρ^Γ. Care must be taken, however, with the two off-diagonal elements. Those matrix elements share the vacuum state $|0\rangle_c|0\rangle_d$ with all other submatrices, and therefore beg an eigenvalue solution to the matrix

$$\begin{pmatrix} 0 & p_1/2 & ... & p_n/2 \\ p_1/2 & 0 & ... & 0 \\ ... & ... & & ... \\ p_n/2 & 0 & ... & 0 \end{pmatrix}. \tag{4.27}$$

It can be shown, with a little effort, that the only two nonzero eigenvalues of this matrix are

$$\lambda_{1,2} = \pm\frac{1}{2}\sqrt{\sum_{k=1}^{n} p_k^2}. \tag{4.28}$$

Notice from Eq. (4.8) that

$$\sum_{k=1}^{n} p_k^2 = \mathrm{Tr}\rho_{\mathrm{in}}^2, \tag{4.29}$$

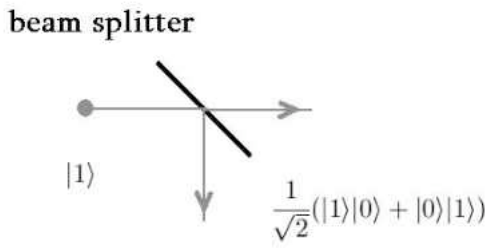

FIGURE 4.1 A single photon impinges on a 50/50 beam splitter and one ebit of entanglement is created between the two output ports. How does the entanglement change when the input photon is in a mixed state?
Answer: When we use the logarithmic negativity[82,114] as our entanglement monotone of choice we find $E_{\mathcal{N}} = \log_2\left(1+\sqrt{\text{purity}}\right)$, in terms of the purity $\text{Tr}\rho^2$ of the input state (see Section 4.3).

and that $\text{Tr}\rho_{\text{in}}^2$ is an invariant quantity under basis transformations, or equivalently, unitary operations. $\lambda_{1,2}$ can thus be evaluated as

$$\lambda_{1,2} = \pm\frac{1}{2}\left(\text{Tr}\rho_{\text{in}}^2\right)^{1/2}. \tag{4.30}$$

Since we learn from the definition that $\|\rho^{\Gamma}\|_1$ equals half of the absolute value of the sum of all negative eigenvalues, we immediately get

$$E_{\mathcal{N}}(\rho_{\text{out}}) = \log_2\left[1+(\text{Tr}\rho_{\text{in}}^2)^{1/2}\right], \tag{4.31}$$

as announced in Fig. 4.1. (Of course, the purity of the output state equals that of the input state, as we assumed the beam splitter to be describable by a unitary operation.)

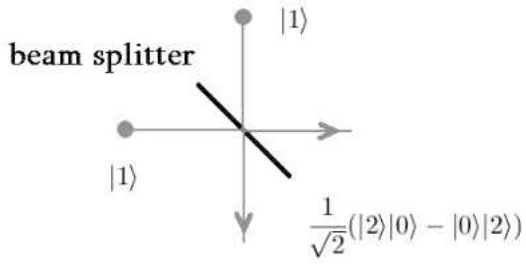

FIGURE 4.2 Two *identical* photons impinge on a 50/50 beam splitter: what is the entanglement at the output?
Answer: 1 ebit, similarly to the answer in Fig.4.1: the output is a delocalized photon *pair*.

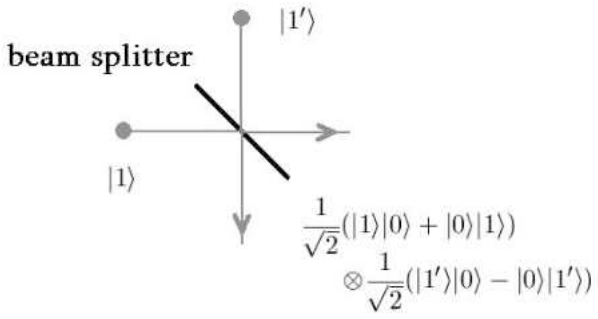

FIGURE 4.3 Two *distinguishable* photons (with, say, orthogonal polarizations or different colors) impinge on a 50/50 beam splitter: what is the entanglement at the output?
Answer: 2 ebits (2 equivalent versions of an entangled delocalized single-photon state) .

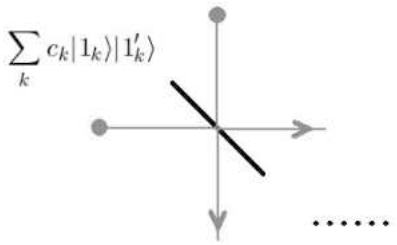

FIGURE 4.4 Two distinguishable photons with orthogonal polarizations impinge on a 50/50 beam splitter. When there is already entanglement between the polarization degrees of freedom in the input state, what is the entanglement at the output?
Answer: $E_{\text{out}} = 2 + E_{\text{in}}/2$ for the entropy of entanglement, generalizing the answer illustrated in Fig. 4.1 (see Section 4.4).
How does this answer change when the input state is mixed?
Answer: it's complicated . . . see Section 4.4 for details.

4.312 Example

With this conclusion, we attempt to calculate the logarithmic negativity for single-photon mixed states for which all wave packet functions $\psi(\omega)$ are Gaussian-shaped

$$\psi(\omega) \sim \exp\left(-\frac{(\omega-\omega_0)^2}{\sigma^2}\right), \tag{4.32}$$

and which is mixed with respect to the arrival time τ of wave-packet peaks. Assuming a Gaussian distribution for arrival times as well, we write

$$\psi(\omega) \to \psi_\tau(\omega) = \psi(\omega)e^{-i\omega\tau}, \tag{4.33}$$

$$P(\tau) \sim \exp\left(-\frac{\tau^2}{\sigma_\tau^2}\right). \tag{4.34}$$

So, explicitly the density matrix is

$$\rho_{\text{in}} = A \int \mathrm{d}\tau \exp\left(-\frac{\tau^2}{\sigma_\tau^2}\right) \int d\omega \int \mathrm{d}\omega' \exp\left[-\frac{(\omega-\omega_o)^2}{\sigma^2} - \frac{(\omega'-\omega_o)^2}{\sigma^2} + i(\omega'-\omega_0)\tau\right] a^\dagger(\omega) \left|\mathrm{v}\right\rangle \left\langle\mathrm{v}\right| a(\omega'), \tag{4.35}$$

with normalization coefficient A. Since the state is profiled in spectral space, we have

$$\mathrm{Tr}\rho_{\text{in}}^2 = \int \mathrm{d}\omega \left\langle\mathrm{v}\right| a(\omega)\rho_{\text{in}}^2 a^\dagger(\omega) \left|\mathrm{v}\right\rangle . \tag{4.36}$$

By using the commutation relation

$$[a(\omega), a^\dagger(\omega')] = \delta(\omega-\omega') \tag{4.37}$$

and basic algebra we can show that

$$\mathrm{Tr}\rho_{\text{in}}^2 = \left(1+4\sigma^2\sigma_\tau^2\right)^{-1/2} . \tag{4.38}$$

The logarithmic negativity after the beam splitter is then according to (4.31)

$$E_{\mathcal{N}}(\rho_{\text{out}}) = \log_2\left[1+(1+4\sigma^2\sigma_\tau^2)^{-1/4}\right], \tag{4.39}$$

which in either limit $\sigma_\tau = 0$ or $\sigma = 0$ reduces to unity. The first limit is just a pure state, which is easily conceived, while the latter fact is a bit harder to reveal, but

think of $\psi(\omega)$ as $\delta(\omega-\omega_o)$ when $\sigma\to 0$. Then according to Eq. (4.35),

$$\begin{aligned}\rho_{\text{in}} =& A\int \mathrm{d}\tau \exp\left(-\frac{\tau^2}{\sigma_\tau^2}\right)\int \mathrm{d}\omega\delta(\omega-\omega_o)e^{-i\omega\tau}\int \mathrm{d}\omega'\delta(\omega'-\omega_o)e^{i\omega'\tau}a^\dagger(\omega)\left|\mathrm{v}\right\rangle\left\langle\mathrm{v}\right| a(\omega')\\ =& A\int \mathrm{d}\tau \exp\left(-\frac{\tau^2}{\sigma_\tau^2}\right) a^\dagger(\omega_o)\left|\mathrm{v}\right\rangle\left\langle\mathrm{v}\right| a(\omega_o), \end{aligned} \tag{4.40}$$

which is equivalently a pure state. Physically, $\delta(\omega-\omega_o)$ corresponds to monochromatic light whose wave function extends homogeneously along the time axis to both infinities, so that the concept of wave-packet arrival time no longer applies.

In conclusion, the entanglement of the output state depends only on the ratio of the "incoherent" width of the mixture in time, σ_τ to the "coherent" width in time of each wave packet $1/\sigma$. For a large incoherent width the entanglement reduces to zero, as expected.

4.32 Adding the Vacuum

Real experiments involving single photons typically are described by a state involving a vacuum component in addition to the single-photon component. The phase between the vacuum and a particular single-photon Fock state may or may not be known or controlled. We may write a state containing the two Fock states as

$$\rho_{\text{vac1}_{\text{in}}} = \int \mathrm{d}\varphi f(\varphi)\left(\sqrt{1-p}\left|\mathrm{v}\right\rangle+\sqrt{p}e^{i\varphi}\left|1\right\rangle\right)\otimes\left(\sqrt{1-p}\left\langle\mathrm{v}\right|+\sqrt{p}e^{-i\varphi}\left\langle 1\right|\right), \tag{4.41}$$

with p the fixed *a priori* probability of detecting a photon. We consider two extreme cases here:

1. $f(\varphi) \sim \delta(\varphi)$ when the state is pure and we simply represent it, instead of ρ, as a pure state

$$|\mathrm{vac1_{in}}\rangle = |\sqrt{1-p}\,|\mathrm{v}\rangle + \sqrt{p}\,|1\rangle\,; \tag{4.42}$$

2. $f(\varphi)$ is a flat distribution so that the cross terms vanish after the integration. The state is thus reduced to

$$\rho_{\mathrm{vac1_{in}}} = (1-p)\,|\mathrm{v}\rangle\,\langle\mathrm{v}| + p\rho_1. \tag{4.43}$$

We treat these two cases one by one. The output state of Eq. (4.42) after a 50/50 beam splitter is

$$|\mathrm{vac1_{out}}\rangle = \sqrt{1-p}\,|00\rangle + \sqrt{\frac{p}{2}}(|10\rangle + |01\rangle). \tag{4.44}$$

The entropy of entanglement and the logarithmic negativity are straightforwardly calculated and the result is

$$E_{\mathcal{N}}(|\mathrm{vac1_{out}}\rangle) = \log_2(1+p). \tag{4.45}$$

The latter expression is particularly simple. Of course, both measures of entanglement vary between 0 and 1 for p varying between 0 and 1. It can be shown that $E_{\mathcal{N}}$ is always larger that E when $p \in [0,1]$.

For the mixture, we are at liberty to assume the diagonal expansion Eq. (4.8) along with Eq. (4.9) and Eq. (4.10). Hence we have

$$\rho_{\mathrm{vac1_{out}}} = (1-p)\left|00\right\rangle\left\langle 00\right| + \frac{p}{2}\sum_k \lambda_k \left(\left|01_k\right\rangle + \left|1_k 0\right\rangle\right)\left(\left\langle 01_k\right| + \left\langle 1_k 0\right|\right). \tag{4.46}$$

The quantities we are interested in are now

$$Pur(\rho_{\mathrm{vac1_{out}}}) = (1-p)^2 + p^2\sum_k \lambda_k^2, \tag{4.47}$$

$$E_{\mathcal{N}}(\rho_{\mathrm{vac1_{out}}}) = \log_2\left(p + Pur^{1/2}\right), \tag{4.48}$$

where Pur denotes the purity of the input state. We see that Eq. (4.48) generalizes the expression Eq. (4.45) which is only a special case when $Pur = 1$. Of course, it also generalizes Eq. (4.31). In conclusion, $E_{\mathcal{N}}$ depends only on p and the purity, which in turn is also affected by the value of p.

4.4 Entanglement of Two-Photon States

4.41 No Vacuum Component

A broadband-pumped down-conversion process produces a state whose two-photon component can be written in the form (4.11), with the mode function in the ideal (pure-state) case described as[115]

$$\psi(\omega_o, \omega_e) = \alpha(\omega_o + \omega_e)\Phi(\omega_o, \omega_e). \tag{4.49}$$

Here $\omega_{o,e}$ are the frequencies of the two photons with ordinary and extraordinary polarization, respectively. $\alpha(\omega_o + \omega_e)$ is the pump spectrum envelope, and the phase-

matching function $\Phi(\omega_o, \omega_e)$, after a great deal of simplification, is

$$\Phi(\bar{\omega}_o + \nu_o, \bar{\omega}_e + \nu_e) = \mathrm{sinc}\left\{\left[\nu_o(k'_o - k'_p) + \nu_e(k'_e - k'_p)\right] L\right\}, \tag{4.50}$$

where $\nu_{o,e}$ are deviations from the perfect match frequencies $\bar{\omega}_{o,e}$, and where L is the length of the nonlinear medium. Moreover, $k'_{p,o,e}$ are the first derivatives of wave vectors with respect to frequency for pump photon and outcoming o- and e- photons respectively at the perfect phase-matching condition

$$\omega_p = \bar{\omega}_o + \bar{\omega}_e. \tag{4.51}$$

We still restrict ourselves to Gaussian wave packets only, and so we assume

$$\alpha(\nu_o + \nu_e) \sim \exp\left[-\frac{(\nu_o + \nu_e)^2}{\sigma^2}\right], \tag{4.52}$$

so that $\psi(\omega_o, \omega_e)$ is real everywhere. In general, due to our ignorance about the precise timing of the pump pulse or about its precise central frequency, the state generated will actually be a mixed state.

We are interested in the entanglement of such a (pure or mixed) state between the two orthogonally polarized modes. Moreover, just like in the preceding section, we wish to calculate the entanglement that results from splitting such a two-photon state on a 50/50 beam splitter. The resulting entanglement after the beam splitter is of a different sort, it's entanglement between the two output modes of the beam splitter, not between orthogonal polarizations. An interesting question is whether that entanglement is larger or smaller than the initial polarization entanglement.

Although we will have to resort to numerical methods to calculate both types of entanglement, we can analytically determine the relation between pure-state entanglement before the beam splitter and that after the beam splitter: Assume that the Schmidt decomposition of a state [those coefficients can be obtained, in some approximation, analytically[124], but we won't need them explicitly] described by Eq. (4.49) is

$$|2_{\text{in}}\rangle = \sum_k \sqrt{\lambda_k} h_k^\dagger v_k^\dagger |\text{v}\rangle , \tag{4.53}$$

(Here we have associated o-photons with horizontal polarization and e-photons with vertical polarization.) The entropy of entanglement and logarithmic negativity are thus expressed in terms of Schmidt coefficients as[82]

$$E(|2_{\text{in}}\rangle) = -\sum_k \lambda_k \log_2 \lambda_k, \tag{4.54}$$

$$E_{\mathcal{N}}(|2_{\text{in}}\rangle) = 2\log_2 \left(\sum_k \sqrt{\lambda_k} \right). \tag{4.55}$$

It can be shown that the state after the beam splitter can be Schmidt-decomposed similarly as

$$|2_{\text{out}}\rangle = \frac{1}{2} \left|\tilde{2}\right\rangle_c |0\rangle_d - \frac{1}{2} |0\rangle_c \left|\tilde{2}\right\rangle_d - \sum_k \frac{\sqrt{\lambda_k}}{2} h_{ck}^\dagger v_{dk}^\dagger |\text{v}\rangle + \sum_k \frac{\sqrt{\lambda_k}}{2} v_{ck}^\dagger h_{dk}^\dagger |\text{v}\rangle , \tag{4.56}$$

where $h_{ck}^\dagger$, $v_{ck}^\dagger$, $h_{dk}^\dagger$ and $v_{dk}^\dagger$ are associated with $c_H^\dagger(\omega)$, $c_V^\dagger(\omega)$, $d_H^\dagger(\omega)$ and $d_V^\dagger(\omega)$ according to Eqs. (4.15). $\left|\tilde{2}\right\rangle_{c,d}$ stands for a specific two photon state, take c e.g.,

$$\left|\tilde{2}\right\rangle_c = \sum_k \sqrt{\lambda_k} h_{ck}^\dagger v_{ck}^\dagger |\text{v}\rangle \tag{4.57}$$

which in frequency space is actually

$$\left|\tilde{2}\right\rangle_c = \int \mathrm{d}\nu_o \int \mathrm{d}\nu_e \alpha(\nu_o+\nu_e)\Phi(\nu_o,\nu_e)c_H^\dagger(\nu_o)c_V^\dagger(\nu_e)\left|\mathrm{v}\right\rangle . \tag{4.58}$$

The notation for the state is just a shorthand notation emphasizing its orthogonality with respect to any single-photon state or vacuum state. Hence we may conclude

$$E(|2_{\mathrm{out}}\rangle) = 2 + \frac{1}{2}E(|2_{\mathrm{in}}\rangle), \tag{4.59}$$

$$E_{\mathcal{N}}(|2_{\mathrm{out}}\rangle) = 2\log_2\left(1+2^{E_{\mathcal{N}}(|2_{\mathrm{in}}\rangle)/2}\right). \tag{4.60}$$

One obvious question is now whether these same relations will still hold for mixed input and output states.

Since analytical solutions to the two-photon problem seem impossible without further approximations, even for the pure-state case (but for an exception see Ref.[124]), we therefore turn to numerical methods where we introduce some standard approximations that were used before in[121,122]. Integrals over continuous frequency are converted into sums over discrete frequencies, and infinity as the integral limit is replaced by an artificial cutoff according to

$$\int_{-\infty}^{\infty}\mathrm{d}\omega\int_{-\infty}^{\infty}\mathrm{d}\omega' \to \Delta\omega\Delta\omega'\sum_{j=1}^{n}\sum_{k=1}^{n'}. \tag{4.61}$$

For convenience we hence choose $\Delta\omega = \Delta\omega'$ and $n = n'$. The choice of $\Delta\omega$ is determined by requiring the integrals to converge numerically. Moreover, we have to choose a scale for the many frequencies that occur in this problem. Quite arbitrarily, we have chosen to rescale all quantities with a dimension of frequency to Ω, defined

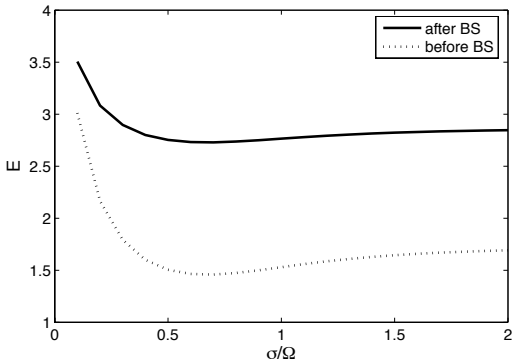

FIGURE 4.5 Entropy of entanglement of pure two-photon states described by Eqs. (4.49)–(4.52) before and after a 50/50 beam splitter as a function of the dimensionless pump width σ/Ω.

through the relation

$$(k'_o - k'_p)L\Omega = 2.25.$$

We also use

$$(k'_e - k'_p)L\Omega = 0.63.$$

These two relations are similar to those used in Ref.[121].

In Figs. 4.5 and 4.6 we plot the numerically evaluated entanglement for pure states, as a function of the width σ of the Gaussian pump pulse in units of Ω. We verified the validity of Eq. (4.59) and Eq. (4.60). Both ν_o and ν_e (deviation from the perfect-match frequencies) are cut off from -2Ω to 2Ω. This is not driven by questions of numerical convergence, but by the freedom one has to consider only those photons in a certain frequency interval. For instance, if one uses narrow-band detectors then only the entanglement between photons of frequency within that bandwidth will be relevant. We simply used cut-off values such that the central peak and two side peaks of the sinc function (4.50) are taken into account.

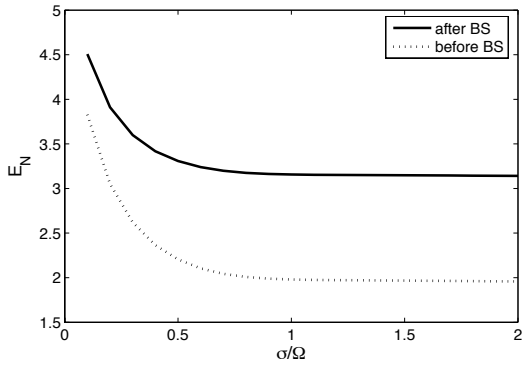

FIGURE 4.6 Logarithmic negativity of pure two-photon states described by Eqs. (4.49)–(4.52) before and after a 50/50 beam splitter as a function of the dimensionless pump width σ/Ω.

Next we consider mixed two-photon states in a way that is similar to what we did for single photon states. We assume Gaussian wave packets with an uncertainty in arrival time. That is, we introduce a time-displaced mode function for two frequencies by

$$\psi(\omega,\omega') \to \psi_\tau(\omega,\omega') = \psi(\omega,\omega')e^{-i(\omega+\omega')\tau}, \tag{4.62}$$

and we choose $P(\lambda)$ in Eq. (4.12) to take exactly the same form as Eq. (4.34) only that λ is replaced by τ. Since the Gaussian shape for τ falls off fairly quickly, we decided to extend the integral in our numerical calculations over the interval $(-2\sigma_\tau, 2\sigma_\tau)$. To be more explicit, after the 50/50 beam splitter the mixed state looks like

$$\rho_{2\text{out}} = \int \mathrm{d}\tau P(\tau) \left|2_{\text{out}}(\tau)\right\rangle \left\langle 2_{\text{out}}(\tau)\right|, \tag{4.63}$$

where $P(\tau) \sim \exp(-\tau^2/\sigma_\tau^2)$ and

$$|2_{\rm out}(\tau)\rangle = \frac{1}{2}\int {\rm d}\nu_o \int {\rm d}\nu_e \alpha_\tau(\nu_o+\nu_e)\Phi(\nu_o,\nu_e)e^{-i(\nu_o+\nu_e)\tau} \times \left(c_H^\dagger(\nu_o)c_V^\dagger(\nu_e) + c_V^\dagger(\nu_e)d_H^\dagger(\nu_o) - c_H^\dagger(\nu_o)d_V^\dagger(\nu_e) - d_H^\dagger(\nu_o)d_V^\dagger(\nu_e)\right)|{\rm v}\rangle \,. \tag{4.64}$$

By realizing that the two-photon states can be expanded in their own space $|2(\tau)\rangle_{c,d}$ instead of single-photon combination space, the size of the density matrix is greatly reduced from $(N+1)^4\times(N+1)^4$ to $(2N+T+1)^2\times(2N+T+1)^2$, N and T being the number of discretization intervals along ω and τ axes respectively. Fig. 4.7 shows our numerical results regarding mixed two-photon states. We display two kinds of curves for three sets of parameters. The solid curves give the numerical results for $E_{\mathcal{N}}(\rho_{\rm out})$, whereas the dotted curves plot $2\log_2\left(1+2^{E_{\mathcal{N}}(\rho_{\rm in})/2}\right)$. In the case of a pure state these two quantities would be the same, according to (4.60), but for mixed states there is a small difference, indicating the relation (4.60) is a fair approximation for mixed states.

From the preceding section we know the precise relation between entanglement generated by a beam splitter and purity for single-photon states. For comparison we plot in Fig. 4.8 the logarithmic negativity as a function of the purity of the mixed state for both single- and two- photon states. The almost linear behavior indicates the close relation between the characteristic purity of the system and the amount of entanglement that can be extracted under ideal conditions. Obviously, entanglement increases with purity.

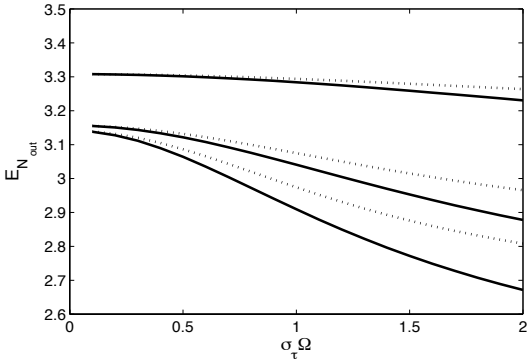

FIGURE 4.7 Logarithmic negativity of mixed two-photon states (4.63) –(4.64) after a 50/50 beam splitter vs. the dimensionless mixed-state width $\sigma_\tau\Omega$. The dotted curves are calculated by plugging the input state's $E_{\mathcal{N}}$ into Eq. (4.60) and the solid curves are direct numerical results. Each set of dotted and solid curves has different wave packet width σ. From top to bottom: $\sigma/\Omega = 0.5, 1, 2$.

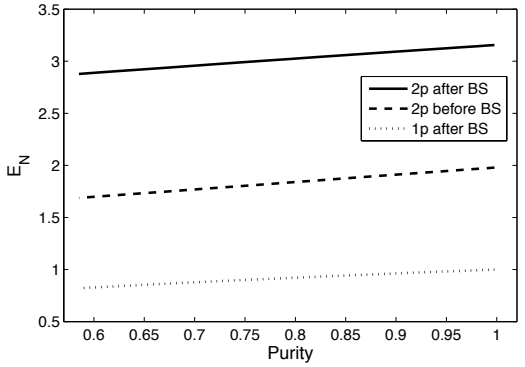

FIGURE 4.8 Logarithmic negativities of various mixed output states of a 50/50 beam splitter are plotted against purity. The single-photon result follows directly from Eq. (4.31), while the two-photon case is calculated from state Eq. (4.63). We have chosen $\sigma = \Omega$ for both two-photon states here and in all remaining figures.

4.42 Adding the Vacuum

Type-II down conversion does not produce a two-photon state. Instead it produces a superposition of a two-photon state and the vacuum. Thus, let us first consider a pure state of the form

$$|\text{vac2}_{\text{in}}\rangle = \sqrt{1-p}\,|\text{v}\rangle + \sqrt{p}\sum_k \sqrt{\lambda_k} h_k^\dagger v_k^\dagger\,|\text{v}\rangle\,, \tag{4.65}$$

where p is the *a priori* probability to detect two photons, which typically will be small. In terms of the Schmidt coefficients we can easily calculate the entanglement present in the state (4.65),

$$E(|\text{vac2}_{\text{in}}\rangle) = -\,(1-p)\log_2(1-p) - p\log_2 p - p\sum_k \lambda_k \log_2 \lambda_k, \tag{4.66}$$

$$E_{\mathcal{N}}(|\text{vac2}_{\text{in}}\rangle) = 2\log_2\left(\sqrt{1-p} + \sqrt{p}\sum_k \sqrt{\lambda_k}\right). \tag{4.67}$$

Just as before we can express the entanglement of a state resulting from splitting this state on a 50/50 beam splitter in terms of the entanglement of the input state. The 50/50 beam splitter accordingly converts the state into

$$\begin{aligned}|\text{vac2}_{\text{out}}\rangle =& \sqrt{1-p}\,|0\rangle_c\,|0\rangle_d + \frac{\sqrt{p}}{2}\left(\left|\tilde{2}\right\rangle_c|0\rangle_d + |0\rangle_c\left|\tilde{2}\right\rangle_d\right)\\ &+ \frac{\sqrt{p}}{2}\sum_k \sqrt{\lambda_k}\left(v_{ck}^\dagger h_{dk}^\dagger + h_{ck}^\dagger v_{dk}^\dagger\right)|\text{v}\rangle\,,\end{aligned} \tag{4.68}$$

where $\left|\tilde{2}\right\rangle$ is given by Eq. (4.57). The entanglement measures are calculated to be

$$\begin{aligned} E(|\mathrm{vac2_{out}}\rangle) = &- \left(\frac{1-p/2+\sqrt{1-p}}{2}\right)\log_2\left(\frac{1-p/2+\sqrt{1-p}}{2}\right) \\ &- \left(\frac{1-p/2-\sqrt{1-p}}{2}\right)\log_2\left(\frac{1-p/2-\sqrt{1-p}}{2}\right) \\ &+ p - \frac{p}{2}\log_2 p - \frac{p}{2}\sum_k \lambda_k \log_2 \lambda_k, \end{aligned} \tag{4.69}$$

$$E_{\mathcal{N}}(|\mathrm{vac2_{out}}\rangle) = 2\log_2\left(1+\sqrt{p}\sum_k \sqrt{\lambda_k}\right). \tag{4.70}$$

We observe that

$$\begin{aligned} E_{\mathrm{out}} - \frac{1}{2}E_{\mathrm{in}} = &- \left(\frac{1-p/2+\sqrt{1-p}}{2}\right)\log_2\left(\frac{1-p/2+\sqrt{1-p}}{2}\right) \\ &- \left(\frac{1-p/2-\sqrt{1-p}}{2}\log_2\frac{1-p/2-\sqrt{1-p}}{2}\right) \\ &+ \frac{1-p}{2}\log_2\left(\frac{1-p}{2}\right) + 1, \end{aligned} \tag{4.71}$$

and

$$2^{E_{\mathcal{N}\mathrm{out}}/2} - 2^{E_{\mathcal{N}\mathrm{in}}/2} = 1 - \sqrt{1-p}. \tag{4.72}$$

These relations are the generalizations of Eq. (4.59) and Eq. (4.60) respectively when vacuum is involved.

For mixed states involving the vacuum no analytical results seem to be possible, so we reverted to numerical calculations. Fig. 4.9 plots some results from those calculations as a function of the probability p.

4.43 Local Filtering

The mixed two-photon state (4.63) is written as a mixture of pure two-photon states (4.64). The latter states are superpositions of two types of states: those with an odd number of photons in each output port of the beam splitter (namely, one), and those with an even number (namely, zero or two). By imagining performing a quantum nondemolition measurement of the parity of the photon number on (one of the) output ports, we are applying a local filter. This filtering cannot increase the entanglement and thus the entanglement after filtering gives a lower bound to the total entanglement present in the mixed state (4.63).

In the pure-state case we can calculate the amount of entanglement resulting from filtering by collapsing the output Eq. (4.68) into either even-number-photon state

$$|\phi_{\text{even}}\rangle = \frac{1}{\sqrt{1-p/2}}\left(\sqrt{1-p}\,|0\rangle\,|_c\,|0\rangle_d + \frac{\sqrt{p}}{2}\,|\tilde{2}\rangle_c\,|0\rangle_d + \frac{\sqrt{p}}{2}\,|0\rangle_c\,|\tilde{2}\rangle_d\right) \tag{4.73}$$

with probability $1-p/2$ or the odd-number-photon state

$$|\phi_{\text{odd}}\rangle = \frac{1}{\sqrt{2}}\sum_k \sqrt{\lambda_k}\left(v_{ck}^\dagger h_{dk}^\dagger + h_{ck}^\dagger v_{dk}^\dagger\right)|\text{v}\rangle \tag{4.74}$$

with probability $p/2$. Averaging the entanglement over the two possible measurement outcomes yields

$$\begin{aligned} E_{\mathcal{N}}^{(\text{even/odd})} =& (1-p/2)E_{\mathcal{N}}(|\phi_{\text{even}}\rangle) + p/2 E_{\mathcal{N}}(|\phi_{\text{odd}}\rangle) \\ =& p + p\log_2\sum_k \sqrt{\lambda_k} - \left(1-\frac{p}{2}\right)\log_2\left(1-\frac{p}{2}\right). \end{aligned} \tag{4.75}$$

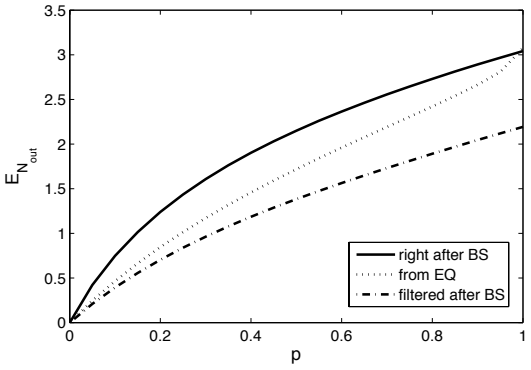

FIGURE 4.9 Logarithmic negativities for the output of a mixed two-photon state accompanied by vacuum at a fixed pump width $\sigma = \Omega$ and a mixture width $\sigma_\tau \Omega = 1$. p is the proportion of the two-photon state, as indicated in Eq. (4.65). The dotted line is calculated from the relation Eq. (4.70), which is true for a pure state only. The dash-dotted line is the entanglement after the beam splitter after the local filtering operation discussed in the text (Section 4.43).

Figs. 4.10 and 4.9 illustrate this filtering effect for *mixed* states as a function of $\sigma_\tau \Omega$ for fixed $\sigma = \Omega$, and as a function of p for fixed σ and σ_τ, respectively. One sees that about a third of the entanglement in the output state arises from coherence between states like $|1,1\rangle$ and $|0,2\rangle + |2,0\rangle$, in symbolic notation.

4.5 Conclusions

We have quantified the entanglement of various mixed states containing exactly one or exactly two photons, as well as states with a nonzero vacuum component, by using the logarithmic negativity. For pure states we found simple relations between the entanglement of single-photon and two-photon states before and after a 50/50 beam splitter. For mixed states such relations are still found to be approximately true.

The simplest result arises for a mixed delocalized single photon. Its entanglement depends only on its purity. The result illustrates that even a perfect deterministic

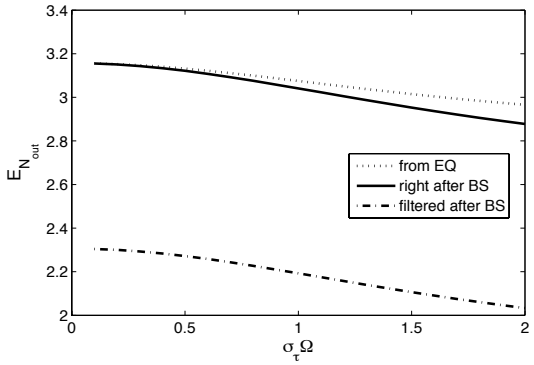

FIGURE 4.10 Logarithmic negativities for the output of mixed two-photon states (4.63)–(4.64). The dotted line is calculated from the relation Eq. (4.60), which is true for a pure state only. The dash-dotted line is the entanglement that would result after local filtering of the output state (see Section 4.43).

single-photon source (never producing more than a single photon) still may not be sufficient for certain quantum-information processing purposes (quantum computing based on dual-rail encoding, for example) if a large degree of entanglement is needed.

We considered fairly realistic cases by explicitly including a vacuum component of photon states, as well as including the spectral and/or temporal shapes of photon wave packets. But an obvious generalization of the present work would be to include full three-dimensional mode structures. Moreover, in the case of two photons impinging on a beam splitter we only treated the case of photons with orthogonal polarizations (having in mind type-II down conversion), but the similar case of identically-polarized photons is interesting as well, and, perhaps surprisingly, more complicated.

CHAPTER V

ENTANGLEMENT VERIFICATION WITH FINITE DATA

This work was published as *Entanglement verification with finite data*, Phys. Rev. Lett. **105**, 170501 (2010). It was initiated and finished jointly by R. Blume-Kohout, J. O. S. Yin and S. J. van Enk.

Entanglement is an essential resource for quantum information processing, and producing and verifying entangled states is considered a benchmark for quantum experiments (for a sample from the most recent experiments on a wide variety of physical systems, see[94,125–137]). Several methods for verifying entanglement have been developed (for overviews, see[57,92]). A bipartite state is entangled if it is not separable, and data $\mathcal{D}$ *demonstrate* entanglement if there is no separable state that could have generated them. As the number of data $N \to \infty$, the data are unambiguous, but for finite N, only probabilistic conclusions can be drawn. In this Letter, we quantify exactly what can be concluded from finite or small data sets, using a simple and efficient *likelihood ratio* test.

We demonstrate the method using two simulated experiments on two-qubit systems [1]. The first measures just one observable, an *entanglement witness*[55,138,139]. The other performs a tomographically complete measurement. In both cases, we use likelihood ratios to draw direct conclusions about entanglement, rather than estimating the quantum state as an intermediate step. A related technique for testing

[1]For larger systems, determining whether a given state is entangled is an NP-hard problem. In multi-partite systems, different classes of entanglement exist, but their classification is still an open problem. Our likelihood ratio method applies to any case where the decision is *binary*: do the data demonstrate entanglement in a particular class or not? So in this paper, the word "separable" can be generalized to "not in the desired entanglement class."

violation of local realism, and based on empirical relative entropy instead of the likelihood ratio, was proposed by van Dam et al[140] and applied by Zhang et al[141].

5.1 Likelihood Ratios

Data $\mathcal{D}$ could have been generated by any one of many independently and identically distributed (i.i.d.) states $\rho^{\otimes N}$. Each state ρ represents a theory about the system, and the relative plausibility of different states is measured by their *likelihood* $\mathcal{L}(\rho)$. A state's likelihood is simply the probability of the observed data *given* that state,

$$\mathcal{L}(\rho) \equiv \Pr(\mathcal{D}|\rho), \tag{5.1}$$

and states with higher likelihood are more plausible. If the *most* likely state is separable, the data clearly do not support entanglement. If it is entangled, then we need to ask how convincing the data are – specifically, whether some separable state is almost as plausible. To judge whether there is (even just one) separable state that fits the data, we compare the likelihoods of (i) the most likely separable state, and (ii) the most likely of all states. Letting $\mathcal{S}$ be the set of separable states, we define

$$\Lambda \equiv \frac{\max_{\rho\in\mathcal{S}} \mathcal{L}(\rho)}{\max_{\text{all } \rho} \mathcal{L}(\rho)}. \tag{5.2}$$

Λ is a *likelihood ratio*, and

$$\lambda = -2\log\Lambda \tag{5.3}$$

represents the weight of evidence in favor of entanglement [2]. To demonstrate entanglement convincingly, an experiment must yield a sufficiently large value for λ.

A likelihood ratio does not assign a probability to "ρ is entangled". Instead, it yields a *confidence level*. We can determine what values of λ typically result from measurements on $\rho^{\otimes N}$, and how their distribution depends on whether ρ is entangled or separable. If we measure $\lambda = \lambda_{\text{exp}}$, and no separable state produces $\lambda \geq \lambda_{\text{exp}}$ with probability higher than ϵ, then we have demonstrated entanglement at the $1-\epsilon$ confidence level. If an experimentalist plans (*before* taking data) to calculate λ and report "ρ is entangled" *only* when the data imply $1-\epsilon$ confidence, then the probability that he erroneously reports entanglement [3] is at most ϵ.

So, ρ may be (i) entangled, (ii) separable, or (iii) on the boundary. Boundary states are still separable, and they are the hardest separable states to rule out. To demonstrate entanglement at the $1-\epsilon$ confidence level, we must show that there is *no* boundary state for which $\Pr(\lambda \geq \lambda_{\text{exp}}) \geq \epsilon$. It is difficult to make rigorous probabilistic statements about λ for small N. But as $N \to \infty$, the following analysis becomes exact, and is generally thought to be reliable for $N \gtrsim 30$[142].

5.2 The Distribution of λ

The set of quantum states ρ is a convex subset of the vector space of trace-1 $d \times d$ Hermitian operators, $\mathbb{R}^{d^2-1}$. An entanglement-verification measurement is

[2]The factor of -2 may seem arbitrary. Statisticians use this convention because λ (as defined) is in many circumstances a χ^2 random variable (see text, below).

[3]Statisticians call this a "Type I error". Erroneously *rejecting* entanglement, even though the experiment is capable of demonstrating entanglement (which is not the same as reporting separability), is a "Type II error". In entanglement verification one tries to avoid Type I errors and is merely mildly unenthusiastic about Type II errors.

represented by a POVM (positive operator-valued measure) $\mathcal{M} = \{E_i \dots E_m\}$, in which each operator E_k represents an event that occurs with probability $p_k = \mathrm{Tr} E_k \rho$ (Born's rule), and each ρ defines a probability distribution $\vec{p} = \{p_1 \dots p_m\}$. Data in which E_k appeared n_k times define empirical frequencies $\vec{f} = \{f_1 \dots f_m\}$, where $f_k \equiv \frac{n_k}{N}$. Both $\vec{p}$ and $\vec{f}$ can be represented as elements of an m-simplex embedded in a vector space $\mathbb{R}^{m-1}$. The probabilities in $\vec{p}$ may be linearly dependent (e.g., if $E_j + E_k = \mathbb{1}$, then $p_j + p_k = 1$ for all ρ), and at most $d^2 - 1$ of them can be independent (because ρ contains only $d^2 - 1$ parameters). We define $\dim(\mathcal{M})$ as the number of independent probabilities.

So Born's rule defines a linear mapping from the operator space containing quantum states into the probability space for measurement $\mathcal{M}$. If $\dim(\mathcal{M}) < d^2 - 1$, then the mapping from states to $\vec{p}$-vectors is many-to-one, and the experiment is completely insensitive to some parameters of ρ. Ignoring these irrelevant parameters makes ρ an (effectively) $\dim(\mathcal{M})$-dimensional parameter. Separable states form a convex subset of all states (see Fig. 5.1). These sets' images in probability space are also nested convex sets (although if $\dim(\mathcal{M}) < d^2 - 1$, then some entangled states will be indistinguishable from separable ones in this experiment).

Suppose that N copies of a state ρ_0 are measured, yielding a likelihood function $\mathcal{L}(\rho)$. $\mathcal{L}(\rho)$ has a unique global maximum $\hat{\rho}_{\mathrm{MLE}}$. As $N \to \infty$, the distribution of $\hat{\rho}_{\mathrm{MLE}}$ approaches a Gaussian around ρ_0 with covariance tensor Δ. $\mathcal{L}(\rho)$ itself is a Gaussian function with the *same* covariance matrix Δ (see note [4]). This defines a characteristic length scale $\delta = |\Delta|_2$ that scales as $\delta = O(1/\sqrt{N})$. We can use Δ to

[4]Technically, this Gaussian ansatz is true only when ρ_0 is full rank – i.e., not on the boundary of the state set. If ρ_0 is rank-deficient, then both the distribution of $\hat{\rho}_{\mathrm{MLE}}$ and $\mathcal{L}(\rho)$ itself are typically truncated by the boundary. However, the analysis remains valid (as $N \to \infty$) except if ρ_0 is simultaneously rank-deficient *and* on the boundary between separable and entangled states.

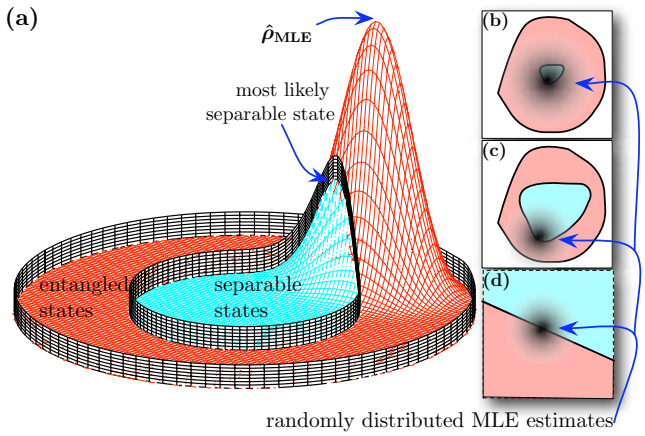

FIGURE 5.1 **General schema of a likelihood ratio test.** The separable states $\mathcal{S}$ (cyan) are a convex subset of all states, surrounded by entangled states (red). Data from an experiment on a state ρ yield a quasiconvex likelihood function [**(a)**] with a unique maximum ($\hat{\rho}_{\mathrm{MLE}}$). $\hat{\rho}_{\mathrm{MLE}}$ is randomly distributed around ρ, at a typical length scale $\delta = O(1/\sqrt{N})$. If $\hat{\rho}_{\mathrm{MLE}}$ is separable then there is no evidence for entanglement, but if it's entangled (as shown), then the relative likelihoods of $\hat{\rho}_{\mathrm{MLE}}$ and the most likely separable state determine the weight of evidence. Data are "convincing" if they are very unlikely to have been produced by a borderline separable state. Typical likelihood ratios for such states depend on the shape of $\mathcal{S}$. In **(b)-(d)** we show three possible cases: in **(b)** $\mathcal{S}$ is smaller than δ and behaves like a point; in **(c)** it is of size δ and its behavior is hard to characterize; in **(d)** it is much bigger than δ and behaves like a half-space.

define a stretched Euclidean metric

$$d(\rho_1, \rho_2) = \sqrt{\mathrm{Tr}\left[(\rho_1 - \rho_2)\Delta^{-1}(\rho_1 - \rho_2)\right]}. \tag{5.4}$$

Using this metric, $\hat{\rho}_{\mathrm{MLE}}$ is univariate Gaussian distributed around ρ_0, and

$$\log \mathcal{L}(\rho) = -\frac{d(\rho, \hat{\rho}_{\mathrm{MLE}})^2}{2}. \tag{5.5}$$

Thus, λ is determined entirely by $d(\hat{\rho}_{\mathrm{MLE}}, \mathcal{S})$, the distance from $\hat{\rho}_{\mathrm{MLE}}$ to the separable set $\mathcal{S}$. If ρ_0 is demonstrably entangled, then λ will grow proportional to N – but if it

is indistinguishable from a separable state, then λ will converge almost certainly to zero (see Figure 5.2).

When ρ_0 is on the boundary, λ neither grows with N nor converges to zero, but continues to fluctuate as $N \to \infty$. Its distribution is controlled by the shape and radius of $\mathcal{S}$, e.g.:

1. If $\mathcal{S}$ is small w/r.t. δ, it behaves like a point (see Figure 5.1b). Then $d(\hat{\rho}_{\text{MLE}}, \mathcal{S}) \approx d(\hat{\rho}_{\text{MLE}}, \rho_0)$, $\lambda = -2\log\left(\mathcal{L}_{max}/\mathcal{L}(\rho_0)\right) = d(\rho, \hat{\rho}_{\text{MLE}})^2$, and so λ is a χ^2 random variable with $\dim(\mathcal{M})$ degrees of freedom (a.k.a. a $\chi^2_{\dim(\mathcal{M})}$ variable).

2. If $\mathcal{S}$ is much larger than δ, then it behaves like a half-space (see Figure 5.1d and note [5]). If $\mathcal{S}$ were a k-dimensional hyperplane, λ would be a $\chi^2_{\dim(\mathcal{M})-k}$ variable. A halfspace behaves like a hyperplane of dimension $(\dim(\mathcal{M}) - 1)$, except with probability $\frac{1}{2}$, $\hat{\rho}_{\text{MLE}}$ is separable. Thus, λ is what we will call a *semi*-χ^2_1 variable: it equals zero with probability $\frac{1}{2}$, and is χ^2_1-distributed otherwise.

As $N \to \infty$, case (2) applies. For small N, however, the real situation is somewhere in between (see Figure 5.1c). $\mathcal{S}$ may be small, and its boundary may be sharply curved, increasing λ. In the absence of a detailed understanding of $\mathcal{S}$'s shape, case (1) provides the best rigorous upper bound on λ. Its cumulative distribution is upper bounded by that of a $\chi^2_{\dim(\mathcal{M})}$ variable – i.e., $\Pr(\lambda > x)$ is no greater than it would be if λ was a $\chi^2_{\dim(\mathcal{M})}$ variable. As $N \to \infty$, the more optimistic semi-χ^2_1 ansatz is valid – but only if we know that N is "large enough".

A χ^2_k variable has expected value k, and higher values are exponentially suppressed. So $\lambda \gg \dim(\mathcal{M})$ is sufficient to demonstrate entanglement at a high

[5] As long as the boundary of $\mathcal{S}$ is differentiable at ρ_0.

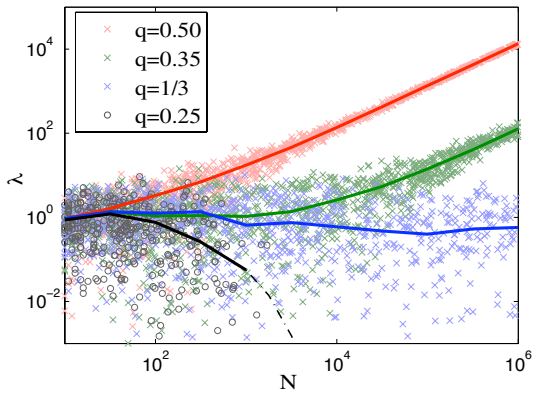

FIGURE 5.2 **Loglikelihood ratios (λ) behave dramatically differently for different states.** 1000 independent simulated tomographically complete experiments were performed, on four different Werner states – separable, barely separable, slightly entangled, and highly entangled. λ is shown for each trial (points), and averaged over all 1000 trials (solid lines). For small N the experiment cannot reliably distinguish them. As N grows, it resolves shorter distances in the state space. For entangled states, typical values of λ increase linearly with N, whereas the separable state almost certainly yields $\lambda = 0$ [not visible in these plots; for $\rho_{q=0.25}$ (black), all trials with more than $N \sim 10^3$ measurements yielded $\lambda = 0$, and the average (dashed line)plunges off the graph]. For barely separable states, λ behaves as a semi-χ^2_k variable with $k = 1$ as $N \to \infty$ (see Fig. 5.3).

confidence level. This implies a tradeoff between an experiment's *power* (ability to identify many entangled states) and its *efficiency* (ability to do so rapidly). Powerful experiments have large dimension – e.g., a tomographically complete measurement can identify *any* entangled state, but has $\dim(\mathcal{M}) = d^2 - 1$. This comes at a price; experiments with large dimension are potentially much more prone to spurious large values of λ, so more data is required to achieve conclusive results [$\lambda \gg \dim(\mathcal{M})$]. Conversely, an entanglement witness (see below) is targeted at a particular state, but it can rapidly and conclusively demonstrate entanglement.

5.3 Implementation

Computing λ involves maximizing $\mathcal{L}(\rho)$ over two convex sets (the set of all states, and the set $\mathcal{S}$ of separable states). $\mathcal{L}(\rho)$ is log-convex, so in principle this is a convex program.

Testing separability is NP-hard, so efficient minimization over $\rho \in \mathcal{S}$ is impossible in general. But for two qubits, the positive partial transpose (PPT) criterion perfectly characterizes entanglement, and λ can be calculated easily (see examples below). For larger systems, $\mathcal{S}$ can be bounded by simpler convex sets, as $\mathcal{S}_- \subset \mathcal{S} \subset \mathcal{S}_+$, (e.g., $\mathcal{S}_+ =$ PPT states, and $\mathcal{S}_- =$ convex combinations of specific product states). Maximizing $\mathcal{L}(\rho)$ over $\mathcal{S}_+$ and $\mathcal{S}_-$ yields bounds on $\max_{\rho\in\mathcal{S}} \mathcal{L}(\rho)$, which may (depending on how wisely the bounding sets were chosen) be tight enough to confirm or deny entanglement.

5.4 Examples

To demonstrate the likelihood ratio test, we simulate two different experiments on two qubits. We imagine an experimentalist trying to produce the singlet state $|\Psi\rangle$, and producing instead a *Werner state*[143],

$$\rho_q = q\Pi_{\text{singlet}} + (1-q)\mathcal{I}/4, \tag{5.6}$$

where $\Pi_{\text{singlet}} = |\Psi\rangle\langle\Psi|$. Werner states are separable when $q \leq 1/3$, and entangled otherwise. The experimentalist's repeated preparations are assumed to be independently and identically distributed (i.i.d.)[144].

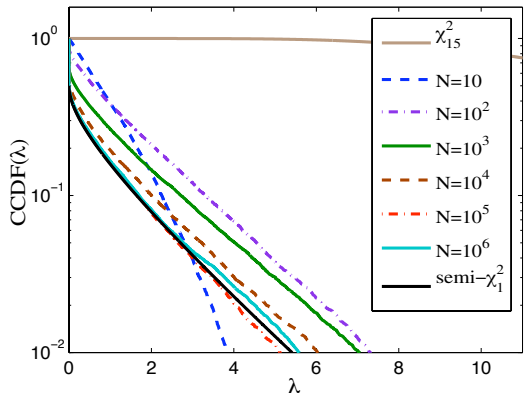

FIGURE 5.3 **Distribution of λ for a SIC-POVM experiment.** We show the empirical *complementary cumulative distribution function* of λ, $CCDF(\lambda_c) = \Pr(\lambda > \lambda_c)$, for the state $\rho_{q=1/3}$ and simulated datasets of size $N = \{10 \ldots 10^6\}$. The CCDF is used to compute confidence levels – e.g., to report entanglement at the 95% confidence level, it is necessary to observe λ such that $CCDF(\lambda) < 0.05$. For this particular state, the chance of a zero λ approaches 50% as N increases. For each N, $CCDF(\lambda_c)$ was based on roughly 10^4 data points from independent trials, each of which generated a value of λ from N tomographically complete measurements on $\rho_{q=1/3}$. We also show CCDFs for a semi-χ^2_1 variable and a $\chi^2_{\dim(\mathcal{M})} = \chi^2_{15}$ variable. The semi-χ^2_1 ansatz is good for large N, but unreliable for small N (yielding too many false positives), while the χ^2_{15} ansatz is very conservative.

5.5 Witness Data

The simplest way to test for entanglement is to repeatedly measure a single *entanglement witness*[55,57,138,139]. An optimal witness for Werner states is $W = \mathcal{I}/2 - \Pi_{\text{singlet}}$. Measuring W yields one of two outcomes – "yes" or "no" – corresponding to POVM (positive-operator valued measure) elements $\{\Pi_{\text{singlet}}, \mathcal{I} - \Pi_{\text{singlet}}\}$. The probability of a "yes" outcome is given by Born's rule as $p = \text{Tr}\rho\Pi_{\text{singlet}}$, so p completely characterizes a state ρ for the purposes of this experiment. The data from N measurements is fully characterized by the frequency of "yes" results, $f = n_{\text{"yes"}}/N$. As $N \to \infty$, $f > \frac{1}{2}$ represents definitive proof that $\langle W \rangle < 0$, and therefore that ρ is entangled. For finite N, $f \leq \frac{1}{2}$ means that a separable state fits as well as any other, so there is no case for entanglement. When $f > \frac{1}{2}$, our likelihood ratio quantifies the

weight of the evidence for entanglement. The likelihood function depends only on p, as

$$\begin{aligned}\mathcal{L}(\rho) = \mathcal{L}(p) = \Pr(\vec{f}|p) &= p^{Nf}(1-p)^{N(1-f)} \\ &= \mathrm{e}^{-N(-f\log p-(1-f)\log(1-p))},\end{aligned} \tag{5.7}$$

making this a single-parameter problem. The maximum likelihood, attained at $p = f$, is $\mathcal{L}_{\max} = \mathrm{e}^{-NH(f)}$, expressed in terms of the data's *empirical entropy*,

$$H(f) = -f\log f - (1-f)\log(1-f). \tag{5.8}$$

If $f > \frac{1}{2}$, the most likely separable state has $p = \frac{1}{2}$, so that $\mathcal{L}_{\text{sep}} = 2^{-N}$, which yields

$$\lambda = -2\log\frac{\mathcal{L}_{\text{sep}}}{\mathcal{L}_{\max}} = 2N\left[\log(2) - H(f)\right]. \tag{5.9}$$

Our numerical explorations (not shown here) confirm that for a barely-separable Werner state, λ behaves as a semi-χ^2_1 variable, even for N as low as 20.

5.6 Tomographically Complete Data

Many entanglement-verification experiments measure a tomographically complete set of observables on a finite-dimensional system (with a heroic example being tomography on 8 ions in an ion trap[74]). Such data identify ρ uniquely as $N \to \infty$, so one can determine with certainty whether ρ is entangled (modulo the computational difficulties in determining whether a specified ρ is separable). Analyzing finite data is more complicated than in the witness example, for the data constrain

a multidimensional parameter space. Ad-hoc techniques are unreliable, and the likelihood ratio test comes into its own.

We consider an apparatus that applies a SIC (symmetric informationally complete)-POVM[145] to each of our two qubits, independently. This measurement (not to be confused with a 4-dimensional SIC-POVM) is tomographically complete, has $4 \times 4 = 16$ outcomes, and yields 15 independent frequencies. Unlike W, it has no special relationship to Werner states, so *any* entangled ρ will yield overwhelmingly convincing data as $N \to \infty$.

We repeatedly simulated $N = 10, \ldots, 10^6$ measurements on a barely-separable Werner state ($\rho_{q=1/3}$), and compared the empirical distribution of λ to those of semi-χ^2_1 and χ^2_{15} random variables (see Figure 5.3). As N gets large, λ becomes indistinguishable from a semi-χ^2_1 variable. For smaller N, this ansatz is too optimistic (and would produce excessive false positives), but the $\chi^2_{d^2-1}$ ansatz is wildly overcautious. We found that for small N, λ behaves like a semi-χ^2_D variable, with D a bit larger than 1 (e.g. $D \approx 1.6$ for $N = 100$).

5.7 Conclusions

Entanglement verification is easy when $N \to \infty$. In practice, N is finite and data are never conclusive. Likelihood ratios provide a simple, reliable test of significance that can be applied to *any* experimental data. Large values of λ are *very* unlikely to be generated by any separable state, but the hardest separable states to rule out are on the boundary. For such states, theory predicts (and our numerics confirm) that λ behaves like a semi-χ^2 random variable. If the underlying state is separable, $\Pr(\lambda > x)$ can be upper bounded using a $\chi^2_{\dim(\mathcal{M})}$ distribution, scaling as e^{-x} for

large x. For entangled states, λ grows linearly with N, and will thus rapidly become distinguishable from any separable state.

CHAPTER VI

CRITERIA FOR RELIABLE ENTANGLEMENT QUANTIFICATION WITH FINITE DATA

This work was published as *Criteria for reliable entanglement quantification with finite data*, Phys. Rev. A **83**, 022326 (2011). It was initiated and finished jointly by J. O. S. Yin and S. J. van Enk.

6.1 Introduction

The study of quantum entanglement never ceases to intrigue researchers[58], and its verification has attracted just as much attention in the quantum information community. Almost all entanglement verification methods[57] are designed for the situation where infinitely many data are (implicitly) assumed to exist. The finite-data regime has not been given much attention until recently[146]. In that paper the main question concerned the binary decision about whether one's quantum systems are entangled or not. In the present chapter we consider the task of quantifying entanglement with finite data. One of the questions we consider here is: how many measurements are needed to quantify entanglement reliably? Obviously, such a question cannot be answered in its full generality, as it will depend on what measurements are performed, on the number of qubits, and, possibly, on how accurate an estimate one wishes to have. Nevertheless, we will develop general criteria for determining a "sufficient" number of measurements based on a Bayesian analysis of measurement data, which can be applied to any sorts of measurements and to any number of qubits. Our criteria do not actually need an accuracy to be specified in advance.

The other goal of this chapter is to develop Bayesian estimation methods for entanglement in nontrivial cases. In particular, we choose to simulate experiments on (mixed, entangled) four-qubit states. Ref.[147] discusses the virtues of Bayesian methods for quantum state estimation, especially as compared to maximum likelihood estimation (MLE), and here we consider that same comparison in the context of entanglement estimation.

An advantage of Bayesian methods is that error bars on entanglement measures are generated automatically. MLE can generate error bars by using a bootstrap method, where ρ_{MLE} is used to numerically generate more data, but this does not work when the number of data is small. In Sec. 6.2 we compare these two methods of generating entanglement estimates and their error bars. The Bayesian methods do require one to choose *prior* probability distributions over states. In Sec. 6.31 we explicitly provide two inherently different standard prior distributions in our systems both of which are numerically feasible and both of which can be applied to any number of qubits. In Sec. 6.32 a convenient entanglement measure is introduced, that can be computed directly from the multipartite density matrices, and which can, likewise, be generalized to any number of qubits. We also briefly discuss the disadvantages of this particular measure (no known multi-partite entanglement measure is without flaws, although a very recent preprint does improve upon the situation[148]. In Sec. 6.33 we derive the relations needed for tomographic state reconstruction that are associated with a special kind of tomographically complete measurements, and in Sec. 6.34 we discuss our implementation of the Metropolis-Hastings algorithm, which allows one to sample from the posterior distribution efficiently. Finally, in Sec. 6.4 we give our main results and attempt to answer the questions laid out in this Introduction.

6.2 Comparisons Between Maximum Likelihood Estimation and Bayesian Methods

In entanglement verification experiments where tomography is adopted, maximum likelihood estimation is widely accepted as the state estimation method of choice. Here the state that best fits the data, ρ_{MLE}, is accepted as the best estimate of the quantum state. While this may sound almost tautological, what MLE fails to give credit to is a large multitude of states that are almost as likely as ρ_{MLE} (see Ref.[147]). Bayesian methods, on the other hand, take these states into account naturally. We will briefly compare MLE and Bayesian methods for entanglement estimation, and in later sections we will get into more details.

Bootstrap methods combined with MLE can be used to generate a distribution of states (somewhat similar to the Bayesian posterior distribution of states). Here one assumes ρ_{MLE} as the real state, from which new sets (of the same size and type as the actual data set) of simulated measurement results are generated. Each such set yields a new MLE state; and thus a distribution of states is generated, which can be used to generate error bars. While this distribution does take into account to some extent the statistical fluctuations, the final estimation of entanglement might still be overly optimistic (see Ref.[147] for a clear account).

The fundamental idea behind Bayesian inference follows from Bayes' theorem. Assume H is a hypothesis and D is the observation data. Bayes' rule tells us that the probability for hypothesis H to be true given observation data D, also known as the posterior probability, is

$$P(H|D) = \frac{P(D|H)}{P(D)} P(H). \tag{6.1}$$

$P(H)$ is the *prior probability*, the probability of H prior to the observations of D. $P(D|H)$ is the *conditional probability* for D to be observed if H is true; it is also called the *likelihood* of the hypothesis given the data; and then it is denoted by $\mathcal{L}(D)$. $P(D)$ is the *marginal probability* for data D, which is usually considered as a normalization factor, namely, as the sum of the conditional probabilities over all mutually exclusive hypotheses

$$P(D) = \sum_j P(D|H_j). \tag{6.2}$$

In our quantum context, the role of hypotheses is played by density matrices ρ.

To be more specific now, we will look at a four-qubit system on which we will perform the simulations presented in the chapter (generalizations are straightforward). We assume some POVM is measured that can be written as a tensor product of local measurements, $\{\Pi_k\}$, (because that tends to be the easiest type of measurement to perform in practice). The outcomes of the POVM measurement can then, likewise, be denoted by $\{\Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n\}$, and f_{jkmn} is the frequency of getting the outcome $\Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n$. The likelihood functional for any state ρ is then by definition

$$\mathcal{L}(\rho) = \prod_{jkmn} \left[\mathrm{Tr}\left(\rho \Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n\right)\right]^{M f_{jkmn}} = \prod_{jkmn} \left(p_{jkmn}\right)^{M f_{jkmn}}, \tag{6.3}$$

where M is the total number of measurements of the POVM, and

$$p_{jkmn} = \mathrm{Tr}\left(\rho \Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n\right) \tag{6.4}$$

is the probability of the outcome $\Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n$, given the state ρ.

The (physical) state that saturates the upper bound of $\mathcal{L}$ is called ρ_{MLE}. Since the estimation is a single state, which can be considered as a distribution with zero width, it is equivalent to taking the limit $M \to \infty$. That is, MLE by reporting a single density matrix essentially assumes that the same data would repeat ad infinitum. Bayesian estimation methods yield the same answer as MLE in that limit, and the influence of the prior is eliminated. In the cases where ρ_{MLE} saturates the upper bound of $\mathcal{L}$ in such a way that $p_{jkmn} \neq \mathrm{Tr}\left(\rho_{\mathrm{MLE}} \Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n\right)$ [1], ρ_{MLE} tends to lie on the boundary of the set of physical states[147]. That is, the state is of non-maximal rank and some eigenvalues are zero. This usually happens when M is "small."

The second step of MLE+bootstrap is to simulate a new dataset by using $\mathrm{Tr}\left(\rho_{\mathrm{MLE}} \Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n\right)$ as probabilities of measurement outcomes. Repeating this procedure many times (using the same ρ_{MLE}) will produce a distribution of measurement outcomes and inferred quantities, and thus error bars on those quantities. As the new data are generated by ρ_{MLE}, entanglement can be easily overestimated if ρ_{MLE} lies on the boundary of the set of physical states, since generically rank-deficient states are more entangled than full-rank states. On the other hand, if ρ_{MLE} is away from the boundary, then the distribution produced this way is expected to resemble the posterior distribution generated by Bayesian methods.

One interesting question is how fast the gap closes up between the two estimates, MLE and Bayesian, as the number of measurements M increases. In fact, this comparison will serve as a (half) criterion for determining how many measurements is "sufficient," provided we choose some "standard" prior distribution to be used in Bayesian entanglement estimation.

[1] In this case there is an Hermitian trace-1 operator σ that does satisfy $p_{jkmn} = \mathrm{Tr}\left(\sigma \Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n\right)$, but σ is not positive definite.

6.3 Preliminaries

Before we can tackle the main questions of this chapter, we need to make several choices, and we need several definitions. These are all collected in this Section.

6.31 Priors and Measures

Let S be the set of all physical states ρ and μ be a measure on the space of S. Particularly in probability theory, $\int_S \mathrm{d}\mu = 1$. If f is any real function of ρ, then the expectation value of f over the space of S is specified by the measure μ:

$$\langle f \rangle = \int_S f(\rho)\mathrm{d}\mu. \tag{6.5}$$

If ρ is parameterized by a set of real parameters $\mathbf{x}$: $\rho = \rho(\mathbf{x})$, then μ becomes the Lebesgue measure over the space of $\mathbf{x}$: $\mathrm{d}\mu = \mathrm{d}\mathbf{x}$, where $\mathrm{d}\mathbf{x}$ is the infinitesimal volume in the corresponding real parameter space. The choice of the parametrization of state ρ , which essentially implies the choice of measure in the space of all states S, induces a prior, $P_\mu(\rho)$. Namely, a uniform random distribution over the parameter space defines a particular prior distribution over the space of the physical states through the relation

$$P_\mu(\rho)d\rho = \rho(\mathbf{x})d\mathbf{x}. \tag{6.6}$$

Thus we claim that we have, in this context, established the connection between the prior and the measure. In numerical implementations where one samples from the

random distribution over $\mathbf{x}$ the integral is replaced by the sum:

$$\int_S f(\rho(\mathbf{x}))\mathrm{d}\mathbf{x} \to \sum_{\mathbf{x}} \Delta\mathbf{x} f((\rho(\mathbf{x})). \tag{6.7}$$

6.311 The GH and Z priors

To study a system consisting of four qubits we choose two inherently different priors: Z and GH, which correspond to two distinct measures of the state space. The measures are chosen for their numerical convenience and for their extendability to arbitrary numbers of qubits. Moreover, they are both dense in the set of all states.

To define the Z measure (or prior) we first write the density matrix for a four-qubit system as

$$\rho = VEV^\dagger, \tag{6.8}$$

where E is a diagonal matrix that carries all the eigenvalues and V is a unitary matrix. The measure of states can be chosen as a product of two particular independent measures introduced in[149,150]

$$\mu(\rho) = \mu(E) \times \mu(V). \tag{6.9}$$

$\mu(E)$ constitutes a 15-dimensional simplex, which is a uniformly distributed manifold defined by a unit sum of 16 nonnegative numbers, and $\mu(V)$ is the Haar measure based on the direct products of four matrices, any single one of which is to be chosen from the set of three Pauli matrices and the identity. We name the prior corresponding to this measure "Z prior".

Alternatively, we can parametrize a four-qubit state as

$$\rho = HH^\dagger / \mathrm{Tr}\left(HH^\dagger\right), \tag{6.10}$$

where H is a random complex 16-by-16 matrix, with both the real and the imaginary part of each entry uniformly distributed on $(-1, 1)$. This is closely related to Cholesky decomposition of the positive semidefinite matrices[151], and similar to the parametrization used in Ref.[152], except that in that paper the unit trace condition is imposed by Lagrangian multipliers while here the condition is satisfied automatically. We name the prior correspond to this measure "GH prior".

Each prior over states induces a prior over any quantity that can be calculated as a function of the state. If we are interested in a quantity $N(\rho)$, then we have a prior $P(N)dN = (P(\rho)dN/d\rho)d\rho$. In particular, in the next subsection we will define two measures of four-qubit entanglement, two "negativities", N_1 and N_2, that both can be calculated (easily) for given states. In FIG.6.1 we show the two induced prior distributions over N_1 (left) and the GH prior distributions for N_1 and N_2 (right). From this point on we will stick to Z and GH priors for the demonstration of further results. Note that these priors are not meant to represent anyone's subjective prior beliefs: rather they are two *standard* priors to be used for our specific purposes of quantifying entanglement and determining how many measurement are needed for that.

It is worth mentioning some observations on simple variations of the above priors. In particular, both priors have the property that the weight of entangled states is larger than that of unentangled states (more precisely, we compare zero negativity states vs. nonzero negativity states, using our definition of multi-partite negativity, see the following section for details). In order to achieve a prior distribution where

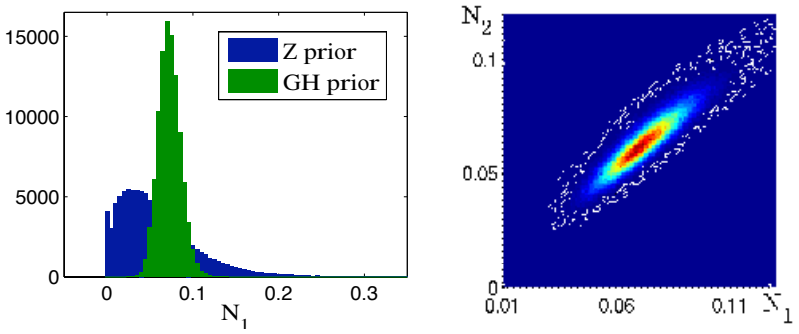

FIGURE 6.1 (Color online) Left: Two different prior distributions of states, named "Z" and "GH" (for details, see main text, Section 6.311 in particular), and the induced prior distributions of the 4-party negativity N_1 (defined in Section 6.32).
Right: GH prior distribution over the two negativities N_1 and N_2, showing the strong correlation between the two measures for randomly drawn states. The "white" noise is due to statistical fluctuations due to the finite sample size.

the ratio of the weight of separable states vs. entangled states is unity, we can mix in an appropriate amount of the identity matrix into the pure Z and pure GH measures. That is, after having picked a random state ρ' from either measure, we take $\rho = \lambda\rho' + (1-\lambda)\mathcal{I}/D$, with $\mathcal{I}/D$ the maximally mixed state in the Hilbert space. λ can be sampled from any distribution that leads to a equal weight between entangled and non-entangled states. In our calculation we chose $\lambda = u^{\beta_{Z,GH}}$, where u is uniformly (Lebesgue) random on $[0,1]$ and $\beta_{Z,GH}$ is an adjustable distortion parameter chosen to ensure a 50% probability of entangled or non-entangled states, $\beta_Z = 0.66$ and $\beta_{GH} = 0.50$.

How different are the pure and mixed priors as far as quantifying entanglement is concerned? FIG.6.2 shows the posterior distributions after just a 1000 measurements for three different states, with pure and mixed GH prior respectively (the plots for the mixed and pure Z distributions are very similar). We find that in every case the "mixed" curve gives results very close to the corresponding "pure" curve, even when the measurements are still far from sufficient for reliable entanglement quantification

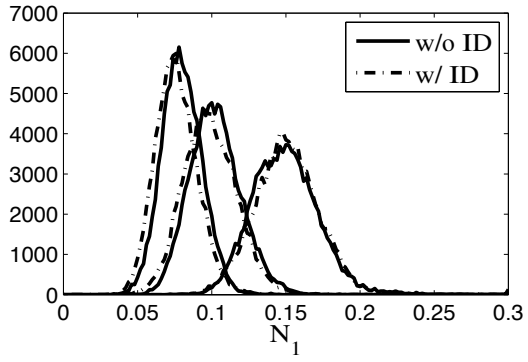

FIGURE 6.2 The posterior distributions resulting from a pure GH prior (w/o ID) and a mixed GH prior (w/ ID, i.e., with the identity mixed in, see text for details). From left to right, the curves describe single trials of just a 1000 measurements on the state ρ_q (Eq. (6.23)) with $q = 0.4, 0.6, 0.8$ from left to right.

(as we will see in Section IV). This indicates that the choice of "pure" Z or GH measures is at least somewhat robust against certain simple modifications.

6.32 Multipartite Entanglement Measures

As mentioned, the system we are particularly interested includes four qubits, which is computationally affordable but sufficiently complicated as a step towards scalable multipartite systems. Despite the intensive studies in the multipartite entanglement[65,85,153–162] over the years, almost all attempts at categorizing multipartite entangled states consider first pure states, and the entanglement measures for pure states can then be extended to mixed states through a convex roof extension, but this involves an arduous minimization over all possible decompositions of the mixed states. To illustrate our ideas without getting too involved in any numerical optimizations, we choose to extend an easily calculable and thereby desirable measure, namely, negativity[82], to the four qubit system. The negativity originated from the idea of the partial transpose[54]. As is well known by now, for 2×2

and 2×3 systems that negative partial transpose (NPT) is a necessary and sufficient condition for entanglement[55]. The negativity has been shown to be closely associated with the fidelity of quantum teleportation[18] and its logarithm bounds the amount of entanglement that can be distilled[163]. The major advantage of the negativity is that it is directly computable for both pure states and mixed states regardless of the size of the system, e.g., the number of qubits, as long as the density matrix is given.

Suppose we have a quantum system consisting of multiple subsystems. We can partition the subsystems into two groups, say X and Y. The negativity of a state ρ with respect to that partition $X - Y$, is defined as

$$\mathcal{N}_{X-Y}(\rho) = \left|\left|\rho^{\Gamma_Y}\right|\right|_1 - 1, \tag{6.11}$$

where Γ_Y stands for partial transpose with respect to subsystem Y and $||\cdot||_1$ for the trace norm of a matrix. For four-qubit systems there are two ways of partitioning into groups of certain sizes: "$2-2$" (partitioning the four qubits into two groups of two qubits) and "$1-3$" (partitioning them into one group of three and one single qubit). Correspondingly we define two negativities as the *geometric means*:

$$\mathcal{N}_{2-2} = \left(\mathcal{N}_{AB-CD}\mathcal{N}_{AC-BD}\mathcal{N}_{AD-BC}\right)^{1/3}, \tag{6.12}$$

$$\mathcal{N}_{1-3} = \left(\mathcal{N}_{A-BCD}\mathcal{N}_{B-CDA}\mathcal{N}_{C-DAB}\mathcal{N}_{D-ABC}\right)^{1/4}, \tag{6.13}$$

where

$$\mathcal{N}_{AB-CD} = \left|\left|\rho^{\Gamma_{CD}}\right|\right|_1 - 1, \tag{6.14}$$

and similar for all others. For simplicity we henceforth denote $\mathcal{N}_{2-2}$ by N_1 and we use N_2 for $\mathcal{N}_{1-3}$.

Despite the fact that these negativities can be computed regardless of the system, they do not necessarily make a distinction between all different types of four-party entanglement. For instance, both measures may be nonzero for states that are *not* genuinely four-party entangled (e.g., a state like $[\rho_{AB} \otimes \rho_{CD} + \rho_A \otimes \rho_{BCD}]/2$, where ρ_{AB}, ρ_{CD}, and ρ_{BCD} are entangled); and it may be zero for certain entangled states, namely those with the property that for at least one partition the entanglement is *bound*.

On the other hand, both N_1 and N_2 are entanglement monotones since each single $\mathcal{N}_{AB-CD}$ or $\mathcal{N}_{A-BCD}$ is an entanglement monotone, as shown in Ref.[82]. Moreover, a vanishing N_1 or N_2, or equivalently a positive partial transpose (PPT) indicates nondistillability with respect to the corresponding partition[164].

Whereas for generic states N_1 and N_2 are correlated to a high degree (FIG.6.1), an illuminating counter-example (showing the independence of the two measures) is the Smolin state[165], given by

$$\begin{aligned}\rho =& \frac{1}{4}\left(\left|\Psi^+\right\rangle_{AB}\left\langle\Psi^+\right| \otimes \left|\Psi^+\right\rangle_{CD}\left\langle\Psi^+\right| + \left|\Psi^-\right\rangle_{AB}\left\langle\Psi^-\right| \otimes \left|\Psi^-\right\rangle_{CD}\left\langle\Psi^-\right|\right. \\ &\left. + \left|\Phi^+\right\rangle_{AB}\left\langle\Phi^+\right| \otimes \left|\Phi^+\right\rangle_{CD}\left\langle\Phi^+\right| + \left|\Phi^-\right\rangle_{AB}\left\langle\Phi^-\right| \otimes \left|\Phi^-\right\rangle_{CD}\left\langle\Phi^-\right|\right), \end{aligned} \tag{6.15}$$

where

$$\left|\Psi^\pm\right\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle \pm |10\rangle\right), \tag{6.16a}$$

$$\left|\Phi^\pm\right\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle \pm |11\rangle\right). \tag{6.16b}$$

For the Smolin state, $N_1 = 0$ (it's separable along any 2-2 cut) and $N_2 = 0.5$ (it's entangled along any 1-3 cut). More specifically, $\mathcal{N}_{AB-CD} = \mathcal{N}_{AC-BD} = \mathcal{N}_{AD-BC} = 0$, $\mathcal{N}_{A-BCD} = \mathcal{N}_{B-CDA} = \mathcal{N}_{C-DAB} = \mathcal{N}_{D-ABC} = 0.5$. The evaluations of the entanglement reflect perfectly what is shown in Ref.[165], that for the Smolin state, entanglement can be distilled between any one of the four qubits and part of the rest of the three qubits, while there is no entanglement between any two groups of two qubits.

6.33 SIC-POVM and the Inverted State

For no particular reason we will assume we measure, on each single qubit, a class of tomographic POVMs that is symmetric informationally-complete (the so-called SIC-POVMs), where any pair of two outcome vectors has exactly the same overlap. A single qubit SIC-POVM is formulated as[145]

$$\Pi_\alpha = \frac{1}{2} |\alpha\rangle \langle\alpha| , \quad \alpha = 1, 2, 3, 4. \tag{6.17}$$

They are linearly independent, tomographically complete and satisfy the normalization condition

$$\sum_{\alpha=1}^{4} \Pi_\alpha = \mathcal{I}, \tag{6.18}$$

and the symmetry condition

$$\mathrm{Tr}(\Pi_\alpha \Pi_\beta) = \begin{cases} \frac{1}{4} & \alpha = \beta \\ \frac{1}{12} & \alpha \neq \beta \end{cases}. \tag{6.19}$$

The four-qubit POVM measurement we refer to is the tensor product of the SIC-POVM on individual qubits so that only local measurements are performed. We label the nonorthogonal compound basis as $M_{jkmn}, j, k, m, n = 1, 2, 3, 4$ and $M_{jkmn} = \Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n$. The linear independence and the completeness of M_{jkmn}'s can be inferred from the same properties of the Π_α's for a single qubit system. This makes it possible to expand arbitrary density matrices in terms of the M_{jkmn}:

$$\rho = \sum_{jkmn} q_{jkmn} \Pi_j \otimes \Pi_k \otimes \Pi_m \otimes \Pi_n. \tag{6.20}$$

Note that the coefficients q_{jkmn} here can be negative without compromising the positivity of ρ. In fact, in order for ρ to be an entangled state, at least one of them must be negative (otherwise, Eq. (6.20) gives a separable form). With the help of Eq. (6.4) we are able to tomographically reconstruct the state by setting the probabilities equal to the measurement frequencies p_{jklm} and then expressing the coefficients q_{jkmn}'s in terms of the probabilities p_{jkmn}'s:

$$\begin{aligned} q_{jkmn} =& 6^4 p_{jkmn} - 6^3 \left(\sum_\alpha p_{\alpha kmn} + \sum_\beta p_{j\beta mn} + \sum_\gamma p_{jk\gamma n} + \sum_\delta p_{jkm\delta} \right) \\ &+ 6^2 \left(\sum_{\alpha\beta} p_{\alpha\beta mn} + \sum_{\alpha\gamma} p_{\alpha k\gamma n} + \sum_{\alpha\delta} p_{\alpha km\delta} + \sum_{\beta\gamma} p_{j\beta\gamma n} + \sum_{\beta\delta} p_{j\beta m\delta} + \sum_{\gamma\delta} p_{jk\gamma\delta} \right) \\ &- 6 \left(\sum_{\beta\gamma\delta} p_{j\beta\gamma\delta} + \sum_{\alpha\gamma\delta} p_{\alpha k\gamma\delta} + \sum_{\alpha\beta\delta} p_{\alpha\beta m\delta} + \sum_{\alpha\beta\gamma} p_{\alpha\beta\gamma n} \right) + 1. \end{aligned} \tag{6.21}$$

In an actual experiment where the readout frequencies f_{jkmn} are considered as p_{jkmn}, the state reconstructed by Eq. (6.21) with $p_{jkmn} = f_{jkmn}$ is called ρ_{tomo}, which is equal to ρ_{MLE} if and only if ρ_{tomo} is physical (see[147]). For the case where it is not

physical, ρ_{MLE} can be approximated by setting the negative eigenvalues of ρ_{tomo} equal to zero, followed by a renormalization of the density matrix.

6.34 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm (or MH), among many Monte Carlo methods, is applied to generate a Markov chain of states to obtain directly the posterior distribution over states (hence a "walk"). The MH walk is known for its fast convergence even when the sampling space is too large for direct random sampling to be efficient. Since the state space of mixed four-qubit states is already 255-dimensional, it makes sense for us to use this method.

The algorithm starts at any random (physical) state and decides to take (or not take) the following random *step* each time towards a new state depending on the relative likelihood of the new and the old state. The process lasts until a converging distribution is reached from the steps taken. The overall outcome is a path in the state space towards the region with the most likely states and wandering about that region. One then counts how often a certain state occurs; that is its weight in the posterior distribution. More precisely, the probability of taking a step is determined by the ratio of the likelihood of the next and the current state. For example, if the likelihood of the next state is 0.7 times the likelihood of the current state, then there is a chance of 70% the next state is accepted. On the other hand, if the next state more likely than the current state, i.e. the ratio of the likelihood is larger than 1, then the acceptance is definite. Since the MH walk spends most of its time on the most likely states, it manages to outperform pure random sampling substantially.

One of the concerns in MH walk is setting the appropriate step size, from one state to the next. It can be defined in a certain chosen measure as

$$d_{\text{step}} = ||\rho_{\text{next}} - \rho_{\text{current}}|| \,. \tag{6.22}$$

A small step size may costs a long time for the algorithm to converge, although still faster than random sampling, while a large step size tends to identify less likely states by getting stuck in a low likelihood region, which then produces a less accurate distribution.

In standard practice the acceptance rate, which is defined as the overall probability of accepting a step, is used as a quantitative reflection of a step size. There is no rigid proof of what an optimal acceptance rate is, as the final distribution converges to a smooth one. In our work we tested a wide range of possible step sizes, balancing the stability and the efficiency of the program, and managed to keep it between 35% and 40%, close to the ideal acceptance rate for Gaussian target distribution[166]. As shown in FIG.6.3, the algorithm quickly navigates to the desired area after about 1,000 steps and stays there "indefinitely" until we terminate the procedure after 10^5 steps.

6.4 How Many Measurements?

In order to examine how many measurements suffice for a reliable report of the amount of entanglement in terms of the negativities, it is enlightening to study states that are unlikely to be mistaken as separable states. We choose a particular class of four-qubit states, namely W states with white noise mixed in, which can be

characterized by

$$\rho(q) = q\,|W\rangle\,\langle W| + (1-q)\mathcal{I}/16, \tag{6.23}$$

where $|W\rangle = \frac{1}{2}\left(|0001\rangle + |0010\rangle + |0100\rangle + |1000\rangle\right)$ and $\mathcal{I}$ is the 16-by-16 identity matrix. $|W\rangle$ is known to possess genuine multipartite entanglement[65], and such genuine entanglement can be detected and distinguished from 3-party and 2-party entanglement, as demonstrated recently in an actual experiment[94]. According to the entanglement monotones given earlier in the chapter, ρ becomes 2-2 separable (i.e., $N_1 = 0$) when $q < 0.1112$ and 1-3-separable (i.e., $N_2 = 0$) when $q < 0.1262$. When a sufficiently large q value is chosen, the state is less likely to be confused as a separable one. Indeed, the similarities shared between our results for the states $\rho(q = 0.8, 0.6, 0.4)$ suggests that the conclusions from these three test states can be validly applied to the class of states with a wide range of q values as long as the state is safely entangled.

In the spirit of Bayesian estimation, the posterior distributions are determined by both the observation data and the prior, with the former becoming more and more important as data accumulates. When the posterior distributions resulting from the two inherently distinct GH and Z priors are laid together, we expect that they will overlap more and more as a function of the number of measurements. Indeed such behavior is demonstrated in FIG. 6.4 and FIG. 6.5, and this behavior forms the basis of our Criterion 1. In particular, FIG. 6.4 shows the evolution of the Bayesian posterior distributions as the number of measurements M increases along 10^4, 10^5 to 10^6. The expectation values $\langle N_{1,2}\rangle$ and the standard errors $\langle 2\delta N_{1,2}\rangle$ are computed and shown in FIG.6.5 on a logarithmic scale for the two priors and different numbers of measurements. Both $\langle 2\delta N_{1,2}\rangle_Z$ and $\langle \delta N_{1,2}\rangle_{GH}$ are fitted with $1/M^{0.5}$ (see Appendix).

On the other hand, in the Figure, $\left|\langle N_{1,2}\rangle_Z - \langle N_{1,2}\rangle_{GH}\right|$ is fitted with $M^{-\alpha}$, where α is approximately 0.81 for N_1 and approximately 0.66 for N_2. The behavior of $\left|\langle N_{1,2}\rangle_Z - \langle N_{1,2}\rangle_{GH}\right|$ is analyzed analytically in the $M \to \infty$ limit in the Appendix, with several important simplifying assumptions made. It shows that for any not-too-pathological prior, the average posterior value of a physical quantity N approaches the true value $N_{\rm r}$ as

$$|\langle N\rangle - N_{\rm r}| \sim 1/\sqrt{M}, \tag{6.24}$$

when M is large. When any two priors are considered with the same observation data, the difference between the average posterior values of N behaves like

$$|\langle N\rangle_Z - \langle N\rangle_{GH}| \sim 1/M, \tag{6.25}$$

which converges faster by a factor of order $\sqrt{M}$. This is because the uncertainty in the data affects each value of $\langle N\rangle$ for each prior in the same linear fashion, and hence this uncertainty is canceled out when the difference is taken. This observation leads directly to our first Criterion.

For entanglement quantification to be reliable we require a number of measurements M such that for M and larger number of measurements, we have

$$\textbf{Criterion 1}: \ |\,\langle N\rangle_Z - \langle N\rangle_{GH}\,| < \langle \delta N\rangle_Z + \langle \delta N\rangle_{GH}\,. \tag{6.26}$$

(Obviously, one can always substitute's one favorite measure of entanglement instead of N to create a new criterion. To repeat, our choice of the negativity is for numerical convenience, as well as the fact our measure can be easily generalized to any number

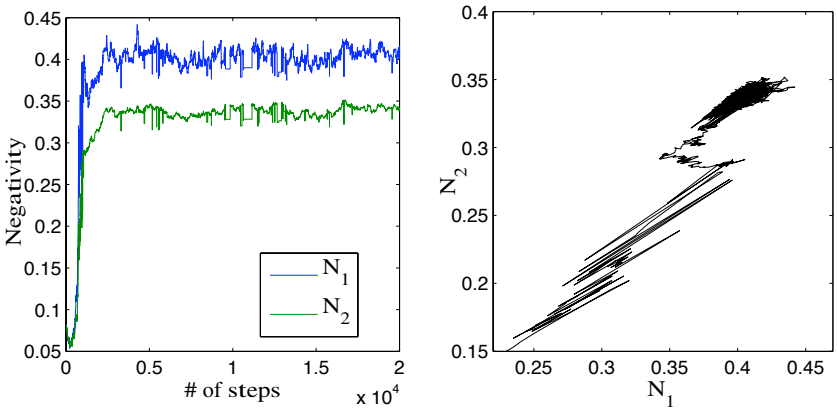

FIGURE 6.3 A Z-prior based Metropolis-Hastings walk towards $\rho(q = 0.8)$ (Eq. (6.23)), for which $N_1 = 0.3875$, $N_2 = 0.3339$. The number of measurements is 10^4.

of qubits.) This means the peaks of the two distributions are closer to each other than their mean standard error. When the two priors are well chosen to be sufficiently distinct, the difference likewise is, presumably, sufficiently large to be spotted. As the measurements accumulate, the distribution will converge towards the true value. And when Eq. (6.26) is satisfied, we claim that the measurements suffice to be trusted and the posterior distribution from either of the priors qualifies as the final result.

According to the Appendix we can write $|\langle N\rangle_Z - \langle N\rangle_{GH}| = A/M$ in the large M limit, where A is a constant. We also write $\langle \delta N\rangle_{Z,GH} = B_{Z,GH}/\sqrt{M}$, where $B_{Z,GH}$ are constants, as indicated by the fitting. Then Eq. (6.26) is satisfied for $\forall M > A^2/(B_Z + B_{GH})^2$. This is observed when the number of measurements is larger than about 10^5 (FIG. 6.5). Therefore 10^5 is the number of the SIC-POVM measurements necessary, according to Criterion 1, for an honest assessment of the amount of entanglement in terms of negativities in a four-qubit system.

Note that both of the two priors used in this chapter are easily generalized to larger number of qubits or other higher dimensional systems. The inherent

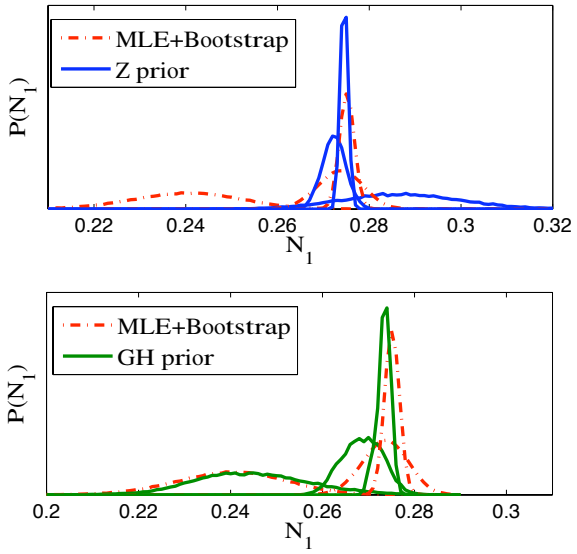

FIGURE 6.4 Estimated probability distribution $P(N_1)$ of MLE & bootstrap method (dot-dashed red) against the posterior distributions with Z and GH priors (solid blue and solid green respectively), after the same series of measurements. The broadest, the medium and the sharpest distribution for each color correspond to $M = 10^4, 10^5, 10^6$. The state being considered is $\rho(q = 0.6)$ (Eq. (6.23)). The red curves are obtained by assuming ρ_{MLE} is the real state, from which the corresponding measurements are simulated and a ρ_{MLE} is found for each set of measurements, thus not requiring any prior. Around $M \approx 10^5$ measurements all three methods more or less agree with each other.

difference between the two, which is observed in terms of the negativities for two qubits and four qubits (FIG.6.1), is expected to persist in similar quantities for larger systems. As a result, the proposed criterion can be extended to multi-qubit systems straightforwardly.

Our next criterion compares estimates of entanglement based on MLE with a Bayesian estimate, using a prior P (either GH or Z). For entanglement quantification to be reliable we require a number of measurements M such that for M and larger

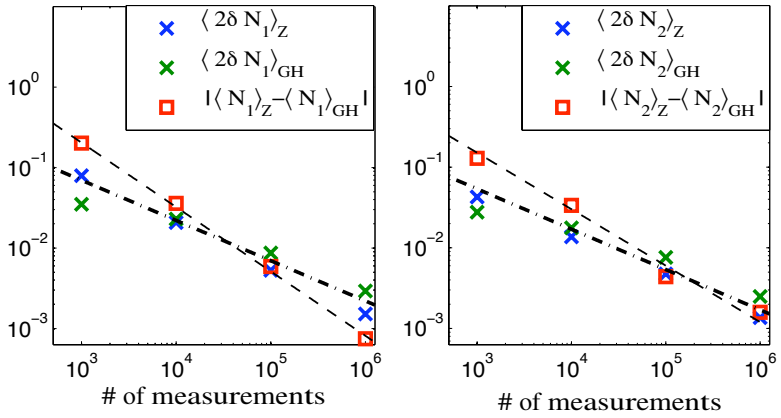

FIGURE 6.5 The difference between the estimations of $\langle N_{1,2}\rangle$ using the Z and GH priors, compared with the standard error $\langle \delta N_{1,2}\rangle$ for $\rho(q = 0.6)$ (Eq. (6.23)). The dash-dotted lines [connecting the sizes of the error bars] are fitted with $c/M^{0.5}$, where c is a number different for the left (N_1) and right (N_2) figures. The dashed lines [connecting the differences between the two estimates] are fitted to guide the eye with $c/M^{0.81}$ (left) and $c/M^{0.66}$ (right) respectively. (Figures for $\rho(q = 0.8)$ and $\rho(q = 0.4)$ look very similar, except for the differences in the slopes of the fitting (dashed) lines.) Our Criterion 1 is formulated in terms of the average of $\langle 2\delta N_1\rangle$ and $\langle 2\delta N_2\rangle$, such that the location where the dashed and the dash-dotted lines cross indicates the number of measurements needed for reliable entanglement estimation.

number of measurements, we have

$$\textbf{Criterion 1.5}: \quad |\langle N\rangle_P - N_{\rm MLE}| < \langle \delta N\rangle_P + \delta N_{\rm MLE}, \tag{6.27}$$

where P stands for either Z or GH. It is, of course, safest (i.e., most conservative) to employ *both* priors, and pick the larger value of M as sufficient. According to the Appendix, $\langle N\rangle_P$ and $N_{\rm MLE}$ approach each other at the rate of $1/M$. The argument that $|N_{\rm MLE} - N_{\rm r}| \sim 1/\sqrt{M}$ can be used to imply that $\delta N_{\rm MLE} \sim 1/\sqrt{M}$, since what $\rho_{\rm r}$ is to $\rho_{\rm MLE}$ is exactly what $\rho_{\rm MLE}$ is to all ρ's that constitute the bootstrap distribution. Therefore, similar to Criterion 1, a number of measurements M can always be found for Criterion 1.5 to be satisfied.

In words, the criterion accepts an estimate of entanglement as reliable if the Bayesian estimate, based on some prior P, and the MLE estimate (using the bootstrap method) agree with each other. It's only half a criterion, as a Bayesian should see no reason to accept the MLE estimate as judge for his estimate; nor should a frequentist accept the Bayesian estimate with some randomly picked prior for that purpose! It is presumably a good criterion for agnostics (and in that case, not independent of the first Criterion, as MLE will agree with both Bayesian estimates only if the latter agree with each other).

As Fig.(6.4) shows, the bootstrap results [for our particular state tested] bear a greater deal of similarities with the GH-based posterior distribution than with the Z-based posterior. Thus the latter determines the critical value of M. For this particular case, one finds once again that $M \approx 10^5$ is necessary for reliable entanglement quantification. Thus, here both Criteria agree with each other.

6.5 Conclusions

We formulated criteria to determine a sufficient number of measurements for reliable entanglement quantification. The main criterion uses two different "standard" prior distributions over states, used in a Bayesian analysis of the measurement data. Namely, if the two posterior distributions resulting from two different priors agree on the amount of entanglement (within error bars) then we can declare that our results have converged and, therefore, that they are reliable. A second criterion, not quite independent of the first, compares the results from maximum likelihood estimation (MLE), without using any prior, to the two Bayesian results. If MLE agrees with the two Bayesian estimates, then, again, we can declare the results sufficiently reliable.

Obviously, in this case the two Bayesian estimates must also agree with each other, and that is why the second criterion is not independent of the first.

We illustrated these criteria by applying them to a particular set of measurements on four qubits [and then both criteria agreed with each other on what constitutes a sufficient number of measurements], but all our results, including the prior distributions, and the measurements considered, and the criteria themselves easily generalize to more (or fewer) qubits.

In order to perform these calculations, we also proposed four-qubit entanglement monotones (based on the negativity) that can be calculated for arbitrary mixed states. Those monotones, too, generalize easily to different number of qubits.

In fact, the extendability of both entanglement measures and priors to arbitrary numbers of qubits is the principal reason to choose these particular criteria (given these ingredients, the criteria then take a standard form for distinguishing two (peaked) probability distributions).

The next question to be answered is how the sufficient number of measurements scales with the number of qubits. How one can analyze this question when the Hilbert space is so large that even the Metropolis-Hastings algorithm fails to work reliably, can be a future subject.

CHAPTER VII

INFORMATION CRITERIA FOR EFFICIENT QUANTUM STATE ESTIMATION

This work was published as *Information criteria for efficient quantum state estimation*, accepted by Phys. Rev. A. It was initiated and finished jointly by J. O. S. Yin and S. J. van Enk.

7.1 Introduction

Quantum state estimation[47,147,167] remains one of the hot topics in the field of quantum information processing. The hope to recover each element in the density matrix from measurement data, however, is impeded by the exponential growth of the number of matrix elements with the number of qubits, and the concomitant exponential growth in time and memory required to compute and store the density matrix. The task can become intimidating when 14 qubits are involved[16], and so efforts have been made to simplify quantum state tomography. One such effort focused on states that have high purity[24] so that the size of the state space shrinks significantly (from $\mathcal{O}(D^2)$ to $\mathcal{O}(D)$ for a system described by a D dimensional Hilbert space). Given that the measurement record is used to verify the assumptions made initially, this method avoids the trap of simplification through imposing *a priori* assumptions merely by *fiat*. Another recent effort[25] in the same spirit considered multi-qubit states that are well represented by matrix product states[168–170] (which require a number of parameters growing only polynomially with the number of qubits). Many states of interest, such as ground states of certain model Hamiltonians in condensed-matter physics, are of that form. Crucially, the particular form of the state can be

verified by the data. Somewhat similarly, a recent paper[171] considered estimating the permutationally invariant part of a density matrix. This, too, requires surprisingly few measurements, and is relevant as many states of interest are permutationally invariant. In fact, the states we will consider in this paper possess that property.

Here we go one step further, and we will allow, tentatively, *any* parametrized form for the density matrix of the quantum system to be tested, possibly containing just a few parameters. In fact, we may have several different tentative ideas of how our quantum state is best parameterized. The questions are then, how the data reveal which of those descriptions work sufficiently well, and which description is the best. This idea corresponds to a well-developed field in statistics: model selection[172–174]. All mathematical descriptions of reality are in fact models (and a quantum state, pure or mixed, is an excellent example of a model), and they can be evaluated by judging their performance relative to that of the true model (assuming it exists). In order to quantify this relative performance, we will make use of the Kullback-Leibler divergence (aka mutual information, aka cross entropy, aka relative entropy)[175], which has the interpretation of the amount of information lost when a specific model is used rather than the true model. Based on the minimization of the Kullback-Leibler divergence over different models, the Akaike Information Criterion (AIC)[176] was developed as a ranking system so that models are evaluated with respect to each other, given measurement data. The only quantities appearing in the criterion are the maximum likelihood obtainable with a given model (i.e., the probability the observed data would occur according to the model, maximized over all model parameters), and the number of independent parameters of the model.

The minimization does not require any knowledge of the true model, only that the testing model is sufficiently close to the true model. The legitimate application

of AIC should, therefore, in principle be limited to "good" models, ones that include the true model (in our case, the exact quantum state that generated the data), at least to a very good approximation. However this does not prevent one from resorting to the AIC for model evaluation when there is no such guarantee. In fact, Takeuchi studied the case where the true model does not belong to the model set and came up with a more general criterion, named the Takeuchi Information Criterion, TIC[177]. However the estimation of the term introduced by Takeuchi to counterbalance the bias of the maximum likelihood estimator used in the AIC, requires estimation of a $K \times K$ matrix (K being the number of independent parameters used by a model) from the data, which, unfortunately, is prone to significant error. This reduces the overall charm and practical use of the TIC. Since in most cases the AIC is still a good approximation to the TIC[173], especially in the case of many data, we stick to the simpler and more robust criterion here.

Information criteria are designed to produce a relative (rather than absolute) ranking of models, so that fixing a reference model is convenient. Throughout this paper we choose the "full-parameter model" (FPM) as reference, that is, a model with just enough independent variables to fully parameterize the measurement on our quantum system. For tomographically complete measurements (discussed in detail in Sec. 7.32) the number of independent variables is given by the number of free parameters in the density matrix ($D^2 - 1$ for a D-dimensional Hilbert space). For tomographically incomplete measurements (see Sec. 7.34), the number of independent variables of FPM is smaller, and equals the number of independent observables. We will, in fact, not even need the explicit form of the FPM (which may be hard to construct for tomographically incomplete measurements), as its maximum possible likelihood can be easily upper-bounded.

We should note an important distinction between maximum likelihood estimation (MLE)[152], a technique often used in quantum tomography, and the method of information criteria and model selection. MLE produces the state that fits the data best. Now the data inevitably contains (statistical) noise, and the MLE state predicts, incorrectly, that same noise to appear in future data. Information criteria, on the other hand, have been designed to find the model that best predicts future data, and tries, in particular, to avoid overfitting to the data, by limiting the number of model parameters. This is how a model with a few parameters can turn out to be the best predictive model, even if, obviously, the MLE state will fit the (past) data better.

We also note that information criteria have been applied mostly in areas of research outside of physics. This is simply due to the happy circumstance that in physics we tend to know what the "true" model underlying our observations is (or should be), whereas this is much less the case in other fields. Within physics, information criteria have been applied to astrophysics[178], where one indeed may not know the "true" model (yet), but also to the problem of entanglement estimation[179]. In the latter case (and in quantum information theory in general) the problem is not that we do not know what the underlying model is, but that that model may contain far too many parameters. Hence the potential usefulness of information criteria. And as we recently discovered, the AIC has even been applied to quantum state estimation, not for the purpose of making it more efficient, but making it more accurate, by avoiding overfitting[180].

7.2 The Akaike Information Criterion - A Schematic Derivation

Suppose we are interested in measuring certain variables, summarized as a vector $\mathbf{x}$, and their probability of occurrence as outcome of our measurement. We

denote $f(\mathbf{x})$ as the probabilistic model that truthfully reflects reality (assuming for convenience that such a model exists) and $g(\mathbf{x}|\vec{\theta})$ as our (approximate) model characterized by one or more parameters, summarized as a vector $\vec{\theta}$. The models satisfy the normalization condition $\int \mathrm{d}\mathbf{x} f(\mathbf{x}) = \int \mathrm{d}\mathbf{x} g(\mathbf{x}|\vec{\theta}) = 1$ for all $\vec{\theta}$. By definition, we say there is no information lost when $f(\mathbf{x})$ is used to describe reality. The amount of information lost when $g(\mathbf{x}|\vec{\theta})$ is used instead of the true model is defined to be the Kullback-Leibler divergence[175] between the model $g(\mathbf{x}|\vec{\theta})$ and the true model $f(\mathbf{x})$:

$$\begin{aligned} I(f, g_{\vec{\theta}}) = & \int \mathrm{d}\mathbf{x} f(\mathbf{x}) \log(f(\mathbf{x})) \\ & - \int \mathrm{d}\mathbf{x} f(\mathbf{x}) \log(g(\mathbf{x}|\vec{\theta})). \end{aligned} \tag{7.1}$$

Eq. (7.1) can be conveniently rewritten as

$$I(f, g_{\vec{\theta}}) = E_{\mathbf{x}} \left[\log(f(\mathbf{x}))\right] - E_{\mathbf{x}} \left[\log(g(\mathbf{x}|\vec{\theta}))\right], \tag{7.2}$$

where $E_{\mathbf{x}}[\cdot]$ denotes an expected value with respect to the true distribution $f(\mathbf{x})$. We see that $\mathbf{x}$ is no longer a variable in the above estimator, as we integrated it out. The only variable that affects $I(f, g_{\vec{\theta}})$ is $\vec{\theta}$. Since the first term in Eq. (7.1) is irrelevant to the purpose of rank-ordering different models g (not to mention we cannot evaluate it when f is not known), we only have to consider the second term. Suppose there exists $\vec{\theta}_0$ such that $g(\mathbf{x}|\vec{\theta}_0) = f(\mathbf{x})$ for every $\mathbf{x}$, that is, the true model is included in the model set. Note that for this to hold, $\vec{\theta}$ does not necessarily contain the same number of parameters as the dimension of the system. To use a simpler notation without the integration over $\mathbf{x}$ we denote the second term in Eq. (7.1) (without the

minus sign) as

$$S(\vec{\theta}_0 : \vec{\theta}) = \int \mathrm{d}\mathbf{x} g(\mathbf{x}|\vec{\theta}_0) \log(g(\mathbf{x}|\vec{\theta})), \tag{7.3}$$

where we have used $g(\mathbf{x}|\vec{\theta}_0)$ to represent the true model $f(\mathbf{x})$. The advantage of this estimator is that it can be approximated without knowing the true distribution $f(\mathbf{x})$. To do that we first consider the situation where $\vec{\theta}$ is close to $\vec{\theta}_0$. This assumption can be justified in the limit of large N, N being the number of measurement records, since the model $\vec{\theta}$ ought to approach $\vec{\theta}_0$ asymptotically (assuming, for simplicity, $\vec{\theta}_0$ is unique). We know that $S(\vec{\theta}_0 : \vec{\theta})$ must have a maximum when $\vec{\theta} = \vec{\theta}_0$, and we may then symbolically expand $S(\vec{\theta}_0 : \vec{\theta})$ in the vicinity of $\vec{\theta}_0$ by

$$\begin{aligned} S(\vec{\theta}_0 : \vec{\theta}) = S(\vec{\theta}_0 : \vec{\theta}_0) - \frac{1}{2}||\vec{\theta} - \vec{\theta}_0||^2_{\vec{\theta}_0} \\ + \mathcal{O}\left(||\vec{\theta} - \vec{\theta}_0||^{3/2}_{\vec{\theta}_0}\right), \end{aligned} \tag{7.4}$$

where

$$||\vec{\theta} - \vec{\theta}_0||^2_{\vec{\theta}_0} = \left(\vec{\theta} - \vec{\theta}_0\right)' \cdot \left.\frac{\partial^2 S(\vec{\theta}_0 : \vec{\theta})}{\partial \vec{\theta}^2}\right|_{\vec{\theta}=\vec{\theta}_0} \cdot \left(\vec{\theta} - \vec{\theta}_0\right) \tag{7.5}$$

denoting a *squared length* derived from a metric defined at $\vec{\theta}_0$. It can be proved that when N is sufficiently large $||\vec{\theta} - \vec{\theta}_0||^2_{\vec{\theta}_0}$ can be approximated by the χ^2_K distribution, with K equal to the number of independent parameters used by the model $\vec{\theta}$. From the properties of the χ^2_K distribution, we know the average value of $||\vec{\theta} - \vec{\theta}_0||^2_{\vec{\theta}_0}$ will approach K.

The next step is to evaluate the estimator $S(\vec{\theta}_0 : \vec{\theta}_0)$, where $\vec{\theta}_0$ is now considered a variable. Suppose we find the maximum likelihood estimate $\vec{\theta}_M$ from the measurement

outcomes such that $S(\vec{\theta}_M : \vec{\theta}_M)$ is the maximum. Now $\vec{\theta}_M$ should also be close to the true model $\vec{\theta}_0$, when N is sufficiently large. Therefore we can similarly expand $S(\vec{\theta}_0 : \vec{\theta}_0)$ in the vicinity of $\vec{\theta}_M$ as

$$S(\vec{\theta}_0 : \vec{\theta}_0) = S(\vec{\theta}_M : \vec{\theta}_M) - \frac{1}{2}||\vec{\theta}_0 - \vec{\theta}_M||^2_{\vec{\theta}_M} + \mathcal{O}\left(||\vec{\theta} - \vec{\theta}_0||^{3/2}_{\vec{\theta}_M}\right), \tag{7.6}$$

$||.||_{\vec{\theta}_M}$ is a length similarly defined as in Eq. (7.4) and has the same statistical attributes as $||\vec{\theta} - \vec{\theta}_0||^2_{\vec{\theta}_0}$ since $\vec{\theta}_0$ is related to $\vec{\theta}_M$ the same way $\vec{\theta}$ is related to $\vec{\theta}_0$ and $\vec{\theta}_M$ is very close to $\vec{\theta}_0$. Its average value, therefore, approaches again K, according to the χ^2_K distribution. Thus we are able to rewrite Eq. (7.3) as

$$S(\vec{\theta}_0 : \vec{\theta}) \approx S(\vec{\theta}_M : \vec{\theta}_M) - K. \tag{7.7}$$

We see that now our target estimator $S(\vec{\theta}_0 : \vec{\theta})$ is evaluated by the MLE solution $\vec{\theta}_M$ only (plus the number of parameters K of the model), with no knowledge of what the true model f is. The assumption that underlies this convenience is constituted by two parts: estimating $S(\vec{\theta}_0 : \vec{\theta})$ with its maximum $\vec{\theta}_0$ and estimating $S(\vec{\theta}_0 : \vec{\theta}_0)$ from the data by its optimum $\vec{\theta}_M$. The deviations from their respective maxima are equal and result simply in the appearance of the constant K.

We now denote $Ł_M = S(\vec{\theta}_M : \vec{\theta}_M)$, which is the maximum likelihood obtainable by our model, with respect to a given set of measurement records. The indicator of the AIC is denoted by Ξ, which is defined by

$$\Xi = -2Ł_M + 2K. \tag{7.8}$$

Apart from the conventional factor 2, and a constant independent of the model $\vec{\theta}$, Ξ is an estimator of the quantity in Eq. (7.1) we originally considered, that is, the Kullback-Leibler divergence between a model that is used to describe the true model and the true model itself. Therefore a given model is considered better than another if it has a *lower* value of Ξ.

Finally, in the case that N is not so large yet that asymptotic relations hold to a very good approximation, one can include a correction factor to the AIC taking the deviation from asymptotic values into account. The corrected AIC gives rise to a slightly different criterion[181]:

$$\Xi_c = -2\text{Ł}_M + 2K + \frac{2K(K+1)}{N-K-1}. \tag{7.9}$$

7.3 Results

7.31 Dicke states

We will apply the AIC to measurements on a popular family of entangled states, the Dicke states of four qubits[74,93,94,98,182]. We simulate two different experiments, one tomographically complete experiment, another measuring an entanglement witness. We include imperfections of a simple type, and we investigate how model selection, according to the AIC, would work. We consider cases where we happen to guess the correct model, as well as cases where our initial guess is, in fact, incorrect.

We consider the four-qubit Dicke states with one or two excitations $\left|D_4^{1,2}\right\rangle$ (with the state $|1\rangle$ representing an excitation):

$$\left|D_4^1\right\rangle = \left(|0001\rangle + |0010\rangle + |0100\rangle + |1000\rangle\right)/2, \tag{7.10a}$$

$$\left|D_4^2\right\rangle = \left(|0011\rangle + |0101\rangle + |0110\rangle + |1001\rangle + |1010\rangle + |1100\rangle\right)\sqrt{6}. \tag{7.10b}$$

For simplicity, let us suppose that white noise is the only random noise in the state generation, and that it corresponds to mixing of the ideal state with the maximally mixed state of the entire space (instead of the subspace with exactly one or two excitations, which could be a reasonable choice, too, depending on the actual implementation of the Dicke states). We thus write the states under discussion as

$$\rho^{1,2}(\alpha) = (1-\alpha)\left|D_4^{1,2}\right\rangle\left\langle D_4^{1,2}\right| + \alpha\mathcal{I}/D, \tag{7.11}$$

where $\mathcal{I}/D$ is the maximally mixed state for dimension $D = 2^4$, and $0 \leq \alpha \leq 1$. We will fix the actual states generating our data to be

$$\rho^{1,2}_{\text{actual}} = \rho^{1,2}(\alpha = 0.2). \tag{7.12}$$

This choice is such that the mixed state is entangled (as measured by our multi-qubit version of the negativity, see below), even though the entanglement witness whose measurement we consider later in Sec. 7.34, just fails to detect it.

For our first model (to be tested by AIC) we wish to pick a one-parameter model (so, $K = 1$) that also includes a wrong guess. A straightforward model choice, denoted

by $M_{1\phi}^{1,2}$, is

$$M_{1\phi}^{1,2} : \rho_{\phi}^{1,2}(q) = (1-q)\left|\Psi_{\text{target}}^{1,2}(\phi)\right\rangle \left\langle\Psi_{\text{target}}^{1,2}(\phi)\right| + q\mathcal{I}/D. \tag{7.13}$$

We refer to the pure states appearing here as the *target* states $\left|\Psi_{\text{target}}^{1,2}(\phi)\right\rangle$, simulating the case where we (possibly incorrectly) think we would be creating a pure state of that form, if only the white noise were absent ($q = 0$). The phase ϕ is included not as a (variable) parameter of the model but as an inadvertently mis-specified property. In this case, it stands for us being wrong about a single relative phase in one of the qubits in state $|1\rangle$. Without loss of generality we assume the first qubit in our representation to carry the wrong phase, and we write

$$\left|\Psi_{\text{target}}^{1}(\phi)\right\rangle = \frac{1}{2}\left(|0001\rangle + |0010\rangle + |0100\rangle + e^{i\phi}|1000\rangle\right), \tag{7.14a}$$

$$\left|\Psi_{\text{target}}^{2}(\phi)\right\rangle = \frac{1}{\sqrt{6}}\left[|0011\rangle + |0101\rangle + |0110\rangle + e^{i\phi}\left(|1001\rangle + |1010\rangle + |1100\rangle\right)\right]. \tag{7.14b}$$

Alternatively, if we *do* consider this a two-parameter model (changing $K = 1$ to $K = 2$), then ϕ is variable, and we would optimize over ϕ. In our case, this optimum value should always be close to $\phi = 0$.

In order to avoid confusion, let us note that we consider model $M_{1\phi}^{1}$ when we simulate measurements on $\rho^1(\alpha = 0.2)$, and we consider model $M_{1\phi}^{2}$ when simulating measurements on $\rho^2(\alpha = 0.2)$.

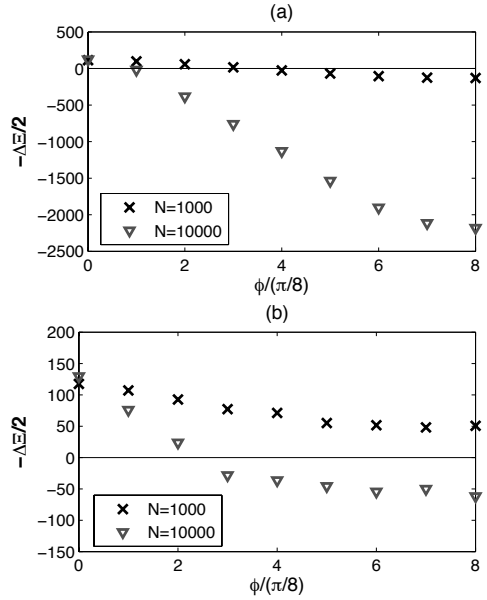

FIGURE 7.1 *How AIC ranks the one- and two-parameter models vs. the full-parameter model (FPM):*
Plot of the difference between Ξ values of our models, the one-parameter model $M^1_{1\phi}$ (a) defined in (7.13) and the two-parameter model $M^1_{2\phi}$ (b) defined in (7.18), and the FPM, *i.e.*, $-\Delta\Xi = \Xi(\text{FPM}) - \Xi(M^1_{1\phi})$ or $-\Delta\Xi = \Xi(\text{FPM}) - \Xi(M^1_{2\phi})$, for various numbers of SIC-POVM measurements, N, with $\left|\Psi^1_{\text{target}}\right\rangle$ as the target state, as functions of the angle ϕ (in units of $\pi/8$). The horizontal line demarcates $\Delta\Xi = 0$: points above (below) that line correspond to cases where the model with fewer (more) parameters is preferred. The figures with $\left|\Psi^2_{\text{target}}\right\rangle$ as the target state look very similar (see FIG. 7.2 for an example of this similarity).

7.32 Tomographically Complete Measurement

We first consider a tomographically complete measurement, in which a so-called SIC-POVM (symmetric informationally complete positive operator values measure[145]) with 4 outcomes is applied to each qubit individually. We first test our one-parameter model, and compare it to the FPM, which contains 255 ($= 4^4 - 1$) parameters, which is the number of parameters needed to fully describe a general

state of 4 qubits. With definition Eq. (7.8) we have

$$\Xi(M_{1\phi}^{1,2}) = -2\text{Ł}_M(M_{1\phi}^{1,2}) + 2, \tag{7.15}$$

since $K = 1$ for $M_{1\phi}^{1,2}$. For the FPM we have

$$\Xi(\text{FPM}) = -2\text{Ł}_M(\text{FPM}) + 2 \times 255, \tag{7.16}$$

where $\text{Ł}_M(\text{FPM})$ is the log of the maximum likelihood obtainable by the FPM. The latter can be bounded from above by noting that the best possible FPM would generate probabilities that exactly match the actual observed frequencies of all measurement outcomes. In the following we will always use that upper bound, rather than the actual maximum likelihood. Even though it is possible to find the maximum likelihood state in principle (and even in practice for small enough Hilbert spaces), we are only concerned with the FPM's ranking according to the AIC, which does not require its density matrix representation. For $M_{1\phi}^{1,2}$ to beat the FPM we require

$$-\Delta\Xi := \Xi(\text{FPM}) - \Xi(M_{1\phi}^{1,2}) > 0. \tag{7.17}$$

This is a sufficient but not necessary requirement, as we use the above-mentioned upper bound to the FPM likelihood.

We plot the difference $\Delta\Xi$ between the two rankings in FIG. 7.1(a) for various values of the number of measurements, and for various values of the phase ϕ. We observe the following: The simple model is, correctly, judged better than the FPM when the phase ϕ is sufficiently small. The more measurements one performs, the

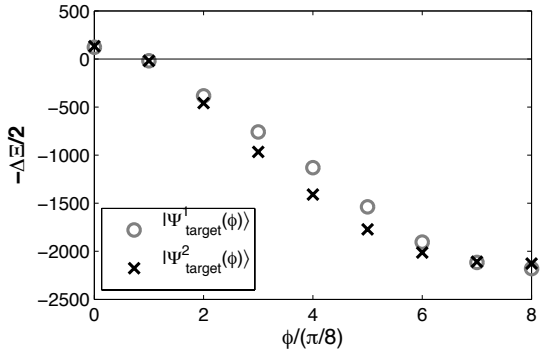

FIGURE 7.2 *Comparing single- and double-excitation Dicke states*: The difference between Ξ's of $M_{1\phi}^{1,2}$ and the FPM as functions of ϕ (in units of $\pi/8$) for N SIC-POVM measurements. There are two different simulations here, one on a noisy Dicke state with 1 excitation using model $M_{1\phi}^{1}$ and another on a noisy Dicke state with two excitations using model $M_{1\phi}^{2}$. Both simulations contain $N = 10000$ measurements. The horizontal line demarcates $\Delta\Xi = 0$.

smaller ϕ has to be for the AIC to still declare the model superior to the FPM (i.e., for the points to stay above the solid line, at $\Delta\Xi = 0$).

Although the correction to the AIC mentioned in Eq. (7.9) is not very small for the FPM for $N = 1000$, applying that correction still does not shift the second and third point below zero: that is, $N = 1000$ measurements is still not sufficiently large for the Ξ_c to recognize that $\phi = \pi/4$ and $\phi = \pi/2$ are incorrect guesses. One can argue about what the cause of this is: it could be that N is just too small for the derivation of the Ξ (or even the Ξ_c) to be correct. Or it could be that the AIC ranking is unreliable because the assumption that the true model is included in the model, is violated. Or it could be that, even with a perfectly valid criterion (perhaps the TIC), the statistical noise present in the data would still be too large.

If we consider the phase ϕ as a second (variable) parameter (thus creating a two-parameter model), then we can give FIG. 7.1 a different interpretation: we would pick $\phi = 0$ as the best choice, and we would increase K by 1. The latter correction is small

on the scale of the plots, and so we find the two-parameter model to be superior to the one-parameter model for any nonzero plotted value of ϕ, and to the FPM. This is a good illustration of the following rather obvious fact: even if one has the impression that a particular property of one's quantum source is (or ought to be) known, it still might pay off to represent that property explicitly as a variable parameter (at the small cost of increasing K by 1), and let the data determine its best value.

7.33 Cross Modeling

Suppose one picked a one-parameter model with a wrong (nonzero) value of ϕ, and the AIC has declared the model to be worse than the FPM. How can one improve the model in a systematic way when one lacks a good idea of which parameters to add to the model (we assume we already incorporated all parameters deemed important *a priori*). Apart from taking more and different measurements, one could use a hint from the existing data. One method making use of the data is to apply "cross modeling," where half the data is used to construct a modification to the model, and the remaining half is used for model validation, again by evaluating Ξ on just that part of the data. So suppose N measurements generate a data sequence $\mathcal{F} = \{f_1, f_2, ..., f_N\}$. One takes, *e.g.*, the first $N/2$ data points, $\{f_1, ..., f_{N/2}\}$, as the training set, and acquires the MLE state ρ_{MLE}, or a numerically feasible approximation thereof, with respect to the training set. We then create a model with two parameters like so:

$$M_{2\phi}^{1,2} : \rho_\phi(\epsilon, q) = (1-\epsilon)\left[(1-q)\rho_{\mathrm{MLE}} + q\left|\Psi_{\mathrm{target}}^{1,2}(\phi)\right\rangle\left\langle\Psi_{\mathrm{target}}^{1,2}(\phi)\right|\right] + \epsilon\mathcal{I}/D. \tag{7.18}$$

For practical reasons ρ_{MLE} does not need to be strictly the MLE state, in particular when the dimension of the full parameter space is large. One would only require it to explain $\{f_1, ..., f_{N/2}\}$ well enough to make sure that part of the data is properly incorporated in the model. Thus, one could, for example, use one of the numerical shortcuts described in[183]. The rest of the data $\{f_{N/2+1}, ..., f_N\}$ is used to evaluate $M_{2\phi}^{1,2}$ against the FPM.

We note the resemblance of this procedure with the method of "cross-validation"[184]. In cross-validation one tries to find out how well a *given* predictive model performs by partitioning the data set into training set and validation set (exactly the same idea as given above). One uses multiple different partitions, and the results are averaged and optimized over those partitions. It can be shown[185] that under certain conditions cross-validation and the AIC are asymptotically equivalent in model selection. This virtually exempts one from having to check multiple partitions of the data set, by applying the AIC to the whole data set.

It is worth emphasizing that what we do here is different in two ways. First, our model is not fixed but modified, based on information obtained from one half of the data. Second, we partition the data set only once, and the reason is, that it would be cheating to calculate the (approximate) MLE state of the full set of data (or, similarly, check many partitions and average), and then consider the resulting MLE state a parameter-free model.

FIG. 7.1(b) shows results for $M_{2\phi}^{1}$ and SIC-POVM measurements. When the number of measurements is $N = 1000$, all $M_{2\phi}^{1}$ models are considered better than the FPM, regardless of the phase error ϕ assumed for the target state. The reason is that around $\phi = \pi/2$ the approximate MLE state obtained from the first half of the

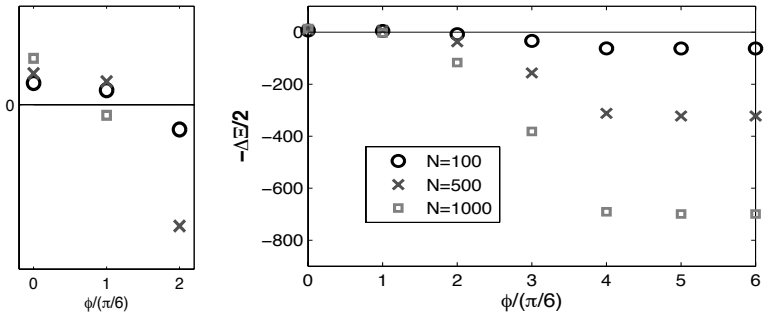

FIGURE 7.3 *How the AIC ranks our one-parameter model vs. the FPM for an entanglement witness measurement*: The difference between Ξ's of $M^2_{1\phi}$ (with $\left|\Psi^2_{\text{target}}(\phi)\right\rangle$) and the FPM, *i.e.*, $-\Delta\Xi = \Xi(\text{FPM}) - \Xi(M^2_{1\phi})$, for different numbers of witness measurements, as functions of ϕ. The horizontal line demarcates $\Delta\Xi = 0$. The left figure zooms in on the leftmost three sets of data points in the right figure.

data is able to "predict" the measurement outcomes (including their large amount of noise!) on the second half better than the 1-parameter model with the wrong phase.

On the other hand, when $N = 10000$ the AIC recognizes only the simple models with small phase errors ($\phi = 0, \pi/8, \pi/4$) as better than the FPM. So, neither the approximate MLE state, nor the 1-parameter model with wrong phase are performing well. This indicates how many measurements are needed to predict a single phase to a given precision.

7.34 Witness Measurement

For states that are close to symmetric Dicke states $\left|D_N^{N/2}\right\rangle$, their entanglement can be verified by using measurements that require only two different local settings, *e.g.*, spins (or polarizations) either *all* in the x-direction or *all* in the y-direction. In particular, when $N = 4$, an efficient witness is $W_{Jxy} = 7/2 + \sqrt{3} - J_x^2 - J_y^2$[99], where $J_{x,y} = \sum_j \sigma_{x,y}^{(j)}/2$, with $\sigma_{x,y}^{(j)}$ the Pauli matrices for the j-th subsystem. This

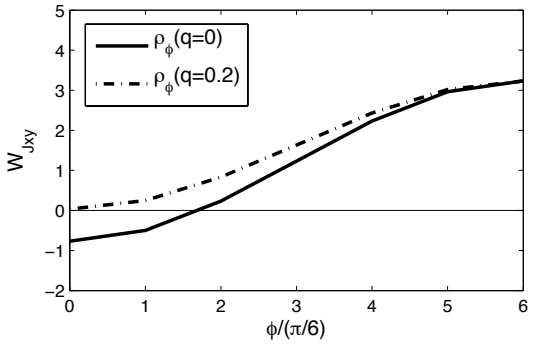

FIGURE 7.4 *Does the witness $W_{J_{xy}}$ detect entanglement if there is a phase error?*: Witness performance $\langle W_{J_{xy}} \rangle$ for different states (defined as in Eq. (7.20)) as a function of ϕ. A negative expectation value detects entanglement.

witness detects (by having a negative expectation value) Dicke states with a white noise background, *i.e.*, $\rho(\alpha) = (1-\alpha)\left|D_4^2\right\rangle \left\langle D_4^2\right| + \alpha \mathcal{I}/D$ whenever $0 \leq \alpha < 0.1920$.

So we suppose we perform $N/2$ measurements on all of the four spins in the x-direction simultaneously, and another $N/2$ similar measurements in the y-direction. Instead of calculating the witness $W_{J_{xy}}$ and ending up with one single value determining entanglement, we make use of the full record of all individual outcomes in order to evaluate (and then maximize) likelihoods. For example, for the measurement of all four spins in the x-direction simultaneously, we can count the number of times they are projected onto the $|x+x+x+x+\rangle$ state, the $|x+x+x+x-\rangle$ state, etc. In both x- or y-directions, the number of independent observables (i.e., the number of independent joint expectation values) is 15, which can be seen as follows: Any density matrix of M qubits can be expressed in terms of the expectation values of 4^M tensor products of the 3 Pauli operators and the identity $\mathcal{I}$, but the expectation value of the product of M identities equals 1 for any density matrix, thus leading to $4^M - 1$ independent parameters encoded in a general density matrix. From having measured just σ_x on all M qubits, we can evaluate all expectation values of all operators that are

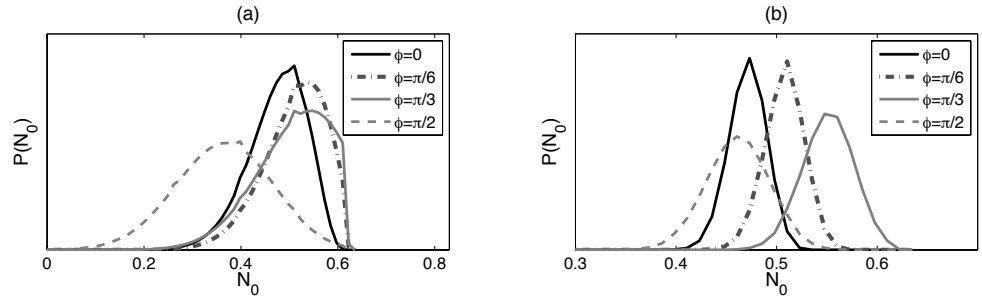

FIGURE 7.5 *How one quantifies entanglement from a witness measurement*: The posterior probability distributions of $\mathcal{N}_0$ for different number of witness measurements ($N = 100$ in (a), $N = 1000$ in (b)), from model $M^2_{1\phi}$ (with target state $\left|\Psi^2_{\text{target}}\right\rangle$), where $\phi = 0, \pi/6, \pi/3$. $\mathcal{N}_0(\rho_{\text{actual}}) = 0.4770$. The posterior distributions are obtained by binning the results into 50 equal-sized bins from the minimum to the maximum value of $\mathcal{N}_0$ for each distribution separately. The prior distribution is assumed to be uniform on [0,1] for both ϵ and q. The distributions of $\mathcal{N}_1$ and $\mathcal{N}_2$ are similar (up to a simple shift).

tensor products of σ_x and the identity. There are 2^M such products, and subtracting the trivial expectation value for $\mathcal{I}^{\otimes M}$ leaves $2^M - 1$ independent expectation values.

This means it only takes $2 \times 15 = 30$ independent parameters to form the FPM, and we have $K = 30$. Similar to the tomographically complete case, we do not need the concrete form of the whole 255-30 dimensional manifold of MLE states, nor do we need to explicitly parameterize the 30-parameter FPM states, as we can simply upper bound the maximum likelihood for this model, $Ł_M(\text{FPM})$, by noting the best one could possibly do is reproduce exactly the observed frequencies of all possible measurement outcomes.

7.35 Estimating Entanglement

Our state $\rho_{\text{actual}} = \rho(\alpha = 0.2)$ is just not detected by the witness W_{Jxy}, but still contains a considerable amount of entanglement. We choose to quantify this entanglement by means of three entanglement monotones (of which only two are

independent), simply constructed from all bipartite negativities. If the four parties are denoted A, B, C and D, the generalized negativities[82,88,89] are defined as

$$\mathcal{N}_1 = \left(\mathcal{N}_{AB-CD}\mathcal{N}_{AC-BD}\mathcal{N}_{AD-BC}\right)^{1/3}, \tag{7.19a}$$

$$\mathcal{N}_2 = \left(\mathcal{N}_{A-BCD}\mathcal{N}_{B-CDA}\mathcal{N}_{C-DAB}\mathcal{N}_{D-ABC}\right)^{1/4}, \tag{7.19b}$$

$$\mathcal{N}_0 = \left(\mathcal{N}_1^3\mathcal{N}_2^4\right)^{1/7}, \tag{7.19c}$$

where $\mathcal{N}_{AB-CD}$ denotes the negativity with respect to partition AB against CD, etc. The main advantage of the generalized negativities is that they are all efficiently computable directly from the density matrix. We have for our state $\mathcal{N}_1 = 0.6293$, $\mathcal{N}_2 = 0.3875$, and $\mathcal{N}_0 = 0.4770$.

Similarly to the tomographically complete case, we first consider the following one-parameter model:

$$M_{1\phi}^2 :\ \rho_\phi(q) = (1-q)\left|\Psi^2_{\text{target}}(\phi)\right\rangle \left\langle\Psi^2_{\text{target}}(\phi)\right| + q\mathcal{I}/D, \tag{7.20}$$

where $\left|\Psi^2_{\text{target}}(\phi)\right\rangle$ is defined in Eg. (7.14b). Ξ's for $M_{1\phi}^2$ and FPM are

$$\Xi(M_{1\phi}^2) = -\,2\text{Ł}_M(M_{1\phi}^2) + 2, \tag{7.21}$$

$$\Xi(\text{FPM}) = -\,2\text{Ł}_M(\text{FPM}) + 2\times 30. \tag{7.22}$$

FIG. 7.3 shows that, as before, the marks above the horizontal solid line correspond to models deemed better than FPM. Compared to the case of full tomography (FIG. 7.1(a)), here the value of $\Xi(M_{1\phi}^2)$ is larger than Ξ(FPM) by a much smaller amount, even when the phase term is correct ($\phi = 0$). The absolute value of the

difference is not relevant, though, and what counts is its sign. The obvious reason for the smaller difference is that the number of independent parameters for the FPM has dropped from 255 to 30. In addition, the FPM in this case does not refer to a specific 30-parameter model. On the contrary, since the number of degrees of freedom of the quantum system is still 255, there is a whole subspace of states, spanning a number of degrees of freedom equal to 225 (=255-30), all satisfying the maximum likelihood condition.

The witness measurement is very sensitive to the phase error, even when the number of measurements is still small. When $N = 1000$, the estimation of ϕ is within an error of $\pi/6$, as the second point plotted is already below the line $\Delta\Xi = 0$. Compared to FIG. 7.1(a), this precision is only reached when $N = 10000$.

An interesting comparison can be made between AIC and the entanglement-detecting nature of witness $W_{J_{xy}}$. FIG. 7.4 shows the performance of $\langle W_{J_{xy}} \rangle$ for the pure state $\left|\Psi^2_{\text{target}}(\phi)\right\rangle$ ($\rho_\phi(q = 0)$, solid curve) and the mixed with 20% of identity mixed in ($\rho_\phi(q = 0.2)$, dot-dashed curve). Even when the state is pure, $\langle W_{J_{xy}} \rangle$ will not be able to witness any entanglement if the phase error is larger than $\pi/3$, just about when AIC declares such a model deficient. Entanglement in the mixed $\rho_\phi(q = 0.2)$ of course is never witnessed. This means $\langle W_{J_{xy}} \rangle$ is only an effective witness in the vicinity of $|D^2_4\rangle$, with limited tolerance of either white noise or phase noise in even just one of the four qubits. (Of course, one would detect the entanglement in the pure state by appropriately rotating the axes in the spin measurement on the first qubit over an angle ϕ.)

To test whether a few-parameter model correctly quantifies entanglement if that model is preferred over the FPM by AIC, we estimate a (posterior, Bayesian) probability distribution over the generalized negativities (defined above). We see that

the first three curves in FIG. 7.5(a) and the first two curves in FIG. 7.5(b), which correspond to the data points above the horizontal line in FIG. 7.3, all give consistent estimates of $\mathcal{N}_0$, compared to the actual value of $\mathcal{N}_0$ for the true state (and the same holds for $\mathcal{N}_{1,2}$ (not shown)). Conversely, the estimate cannot be trusted when AIC deems the simple model inferior to the FPM (of course, it may still happen to be a correct estimate, but one could not be sure). This gives additional evidence for the success of AIC.

7.36 Cross Modeling for a Witness Measurement

We now construct a two-parameter model $M^2_{2\phi}$ similar in spirit to that discussed for tomographically complete measurements: half the data [on which half the time $(\sigma_x)^{\otimes 4}$ is measured, and half the time $(\sigma_y)^{\otimes 4}$] are used to generate a better model, which is then tested on the other half of the data (also containing both types of measurements equally). We write

$$\begin{aligned}\rho(\epsilon, q) =&(1-\epsilon)\left[(1-q)\rho_{\text{observation}}\right.\\ &\left.+q\left|\Psi^2_{\text{target}}(\phi)\right\rangle\left\langle\Psi^2_{\text{target}}(\phi)\right|\right]+\epsilon\mathcal{I}/D. \end{aligned} \tag{7.23}$$

To find a $\rho_{\text{observation}}$—there are many equivalent ones for predicting the outcomes of the witness measurements—we recall that a generic four-qubit state can be expressed as

$$\rho = \sum_{jklm} c_{jklm}\sigma_j \otimes \sigma_k \otimes \sigma_l \otimes \sigma_m, \tag{7.24}$$

where $j, k, l, m = 1, 2, 3, 4$ where $\sigma_{1,2,3}$ denote the Pauli matrices $\sigma_{x,y,z}$ and $\sigma_4 = \mathcal{I}$. The witness measures the coefficients c_{jklm} where j, k, l, m can be combinations of

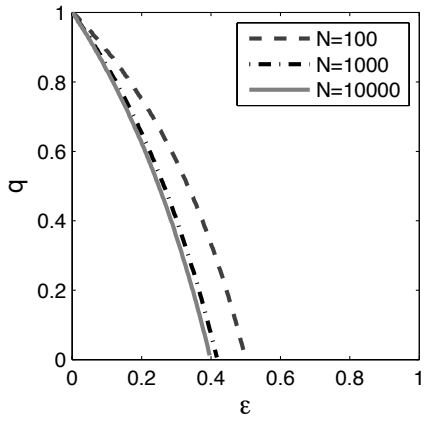

FIGURE 7.6 *What fraction of the model Eq. (7.23) describes physical states?*: The lower left part separated by the curves is where $\rho(\epsilon, q)$ of Eq. (7.23) is unphysical (and so is not actually included in the model), for different number of measurements N.

only 1 and 4 or combinations of only 2 and 4 (*e.g.*, c_{1441} or c_{4222}). We label the c_{jklm}'s that can be recovered from witness measurement as c^w_{jklm} (w as in *witness*). We do not include in c^w_{jklm} the coefficient c_{4444}, which always equals 1/16, so that it does not depend on measurement outcomes. We define

$$\rho_{\text{observation}} = \sum_{jklm} c^w_{jklm} \sigma_j \otimes \sigma_k \otimes \sigma_l \otimes \sigma_m + \mathcal{I}/16. \tag{7.25}$$

Note that $\rho_{\text{observation}}$ can be considered as a trace-one *pseudostate*, since it is not necessarily positive semi-definite. But the most attractive property of $\rho_{\text{observation}}$ is that it preserves the measurement outcomes. It is in fact the unique *pseudostate* that reproduces the exact frequencies of all measurement outcomes *and* that has vanishing expectation values for all other *un*performed collective Pauli measurements. As a component of $\rho(\epsilon, q)$, we allow $\rho_{\text{observation}}$ to be unphysical, but we only keep those $\rho(\epsilon, q)$ that are positive semi-definite. We checked numerically for what values of ϵ and q the states end up being physical, and how this depends on the number

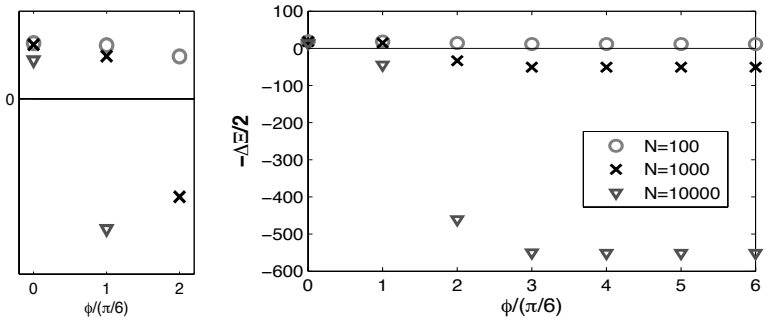

FIGURE 7.7 *How the AIC ranks our two-parameter model vs. the FPM, for a witness measurement?*: The difference between Ξ's of $M^2_{2\phi}$ and the FPM, *i.e.*, $-\Delta\Xi = \Xi(\text{FPM}) - \Xi(M^2_{2\phi})$, for different numbers of witness measurements as a function ϕ. The target state is $\left|\Psi^2_{\text{target}}\right\rangle$. The horizontal line demarcates $\Delta\Xi = 0$. The left figure zooms in on the leftmost three sets of data points in the right figure.

of measurements performed. Physical states are located in the upper right part of the square in FIG. 7.6. That is, only if ϵ and/or q are sufficiently large, so that a sufficiently large amount of $\left|\Psi^2_{\text{target}}\right\rangle$ and/or $\mathcal{I}/16$ has been mixed in, does $\rho(\epsilon, q)$ become physical. Depending on the number of measurements, the area of the upper right part is about 69%-77% of the whole square. The physical/unphysical boundary shifts closer to the origin as the number of measurements increases.

We test the two-parameter model (the physical part of it), and show the results in FIG. 7.7. We find that for $N = 100$ the AIC ranks $\rho_{2\phi}(\epsilon, q)$ better than the FPM, even when the guess about ϕ is very imprecise: 100 witness measurements are, unsurprisingly, not enough for a correct reconstruction of the state. When $N = 1000$, AIC only prefers the models with a value for ϕ within $\pi/6$ of the correct value. And when $N = 10000$, the accepted values of ϕ are even closer to the true value.

The corresponding posterior distributions of negativities $\mathcal{N}_2$ are plotted in FIG. 7.8 for the three better guesses, $\phi = 0, \pi/6, \pi/3$. When $N = 100$ all three give decent predictions of $\mathcal{N}_2$ (and indeed, AIC ranks those models highly). For

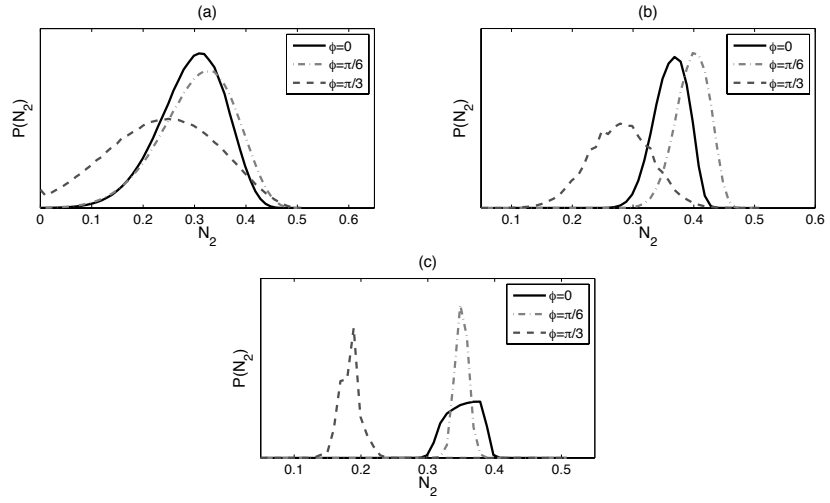

FIGURE 7.8 *Quantifying entanglement from a witness measurement*: The posterior distributions for $\mathcal{N}_2$, for different numbers of measurements ($N = 100$ in (a), $N = 1000$ in (b), $N = 10000$ in (c)), using model $M^2_{2\phi}$ with $\rho_{\text{observation}}$ and target state $\left|\Psi^2_{\text{target}}\right\rangle$, where $\phi = 0, \pi/6, \pi/3$. $\mathcal{N}_2(\rho_{\text{actual}}) = 0.3875$. The posterior distributions are obtained by binning the results into 50 equal-sized bins from the minimum to the maximum value of $\mathcal{N}_2$ for each distribution separately. The same prior is used as in FIG. 7.5. Whenever the AIC declares a model superior to the FPM, the estimated entanglement agrees, within error bars, with the actual value, but may be wrong otherwise.

$N = 1000$ and $N = 10000$, we would only trust the estimates arising from the lower two values of ϕ, or just the correct value of ϕ, respectively. This trust is rewarded in FIG. 7.7(b) and FIG. 7.7(c), as those estimates are indeed correct, within the error bars. In addition, the untrusted estimate for $\phi = \pi/6$ for $N = 10000$ still happens to be correct, too.

7.37 Comparing One- and Two-Parameter Models Directly

Finally, the AIC can compare the one- and two-parameter models $M_{1\phi}$ and $M_{2\phi}$ directly. For that purpose one needs to use the *same* validation set of data, which implies that the two-parameter model needs *additional* data to generate $\rho_{\text{observation}}$.

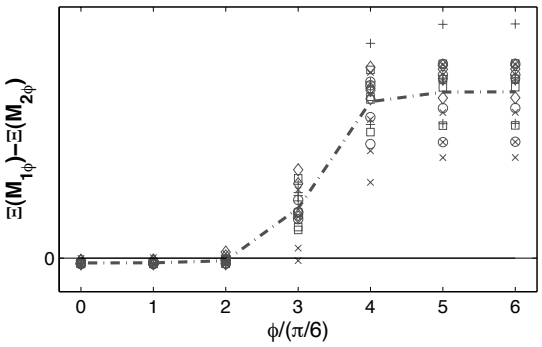

FIGURE 7.9 *Comparing one- and two-parameter models directly*: The difference between Ξ's of $M^2_{2\phi}$ and $M^2_{1\phi}$ for 20 different sets of witness measurements ($N = 50$) as functions of ϕ. The target state is $\left|\Psi^2_{\text{target}}\right\rangle$. The horizontal line demarcates $\Delta\Xi = 0$. The dotted-dashed line is the average of all 20 points at each different ϕ.

Here we display results for just 50 witness measurements, and an additional set of 50 measurements for $M^2_{2\phi}$. FIG. 7.9 shows that even such a small number of additional data is useful if the angle ϕ is wrong, and, similarly, it shows that the same small number suffices to detect a wrong single-qubit phase when it is larger than $\pi/3$.

7.4 Conclusions

We applied information criteria, and the Akaike Information Criterion (AIC) developed in Ref.[176] in particular, to quantum state estimation. We showed it to be a powerful method, provided one has a reasonably good idea of what state one's quantum source actually generates.

For each given model, which may include several parameters describing error and noise, as well as some parameters—call them the ideal-state parameters— describing the state one would like to generate in the ideal (noiseless and error-free) case, the AIC determines a ranking from the observed data. One can construct multiple models, for instance, models where some ideal-state parameters and some noise parameters are

fixed (possibly determined by previous experiments in the same setup), with others still considered variable. Crucially, the AIC also easily ranks the full-parameter model (FPM), which uses in principle all exponentially many parameters in the full density matrix, and which is, therefore, the model one would use in full-blown quantum state tomography. This ranking of the FPM can be accomplished without actually having to find the maximum-likelihood state (or its likelihood)—which quickly would run into insurmountable problems for many-qubit systems—by using a straightforward upper bound.

This way, observed data is used to justify *a posteriori* the use of the few-parameter models—namely, if the AIC ranks that model above the FPM—and thus our method is in the same spirit as several other recent proposals[24,25] to simplify quantum tomography, by tentatively introducing certain assumptions on the quantum state generated, after which data is used to certify those assumptions (and if the certification fails, one at least knows the initial assumptions were incorrect).

We illustrated the method on (noisy and mis-specified) four-qubit members of the family of Dicke states, and demonstrated its effectiveness and efficiency. For instance, we showed that one can detect mis-specified ideal-state parameters and determine noise and error parameters. We also showed by example the successful application of the method to a specific and useful subtask, that of quantifying multi-qubit entanglement.

CHAPTER VIII

CONCLUSIONS

Inspired by the importance of entanglement in quantum information processing, we studied its quantification and verification in different realistic scenarios. For example, as is well-known, a single photon in a pure state impinging on a 50/50 beamsplitter will lead to an output state that has one unit of entanglement. In Chapter III we found analytically that a single photon in a mixed state (a case, surprisingly, not analyzed so far) leads to an output state with less entanglement, determined solely and simply by the purity of the input state. For more complicated cases we performed numerical calculations finding the entanglement of mixed and noisy multi-photon states, including the effects of inevitable vacuum components.

Though not an observable by itself, the presence of entanglement can be verified, and the amount of entanglement can be estimated, by experimental data. While all entanglement quantification and verification methods discussed so far in the literature assumed implicitly an infinite amount of data, we took into account the statistical fluctuations inevitably resulting from finite data. Moreover, whereas almost all theoretical work considered idealized (von Neumann) measurements, our methods easily and automatically take into account realistic noisy and imperfect measurements.

We used several different statistical inference methods to analyze various (finite) experiments. We used both frequentist and Bayesian methods, whichever we deemed more appropriate. For example, an experimental test to verify that there is, in fact, entanglement present, must, by the tough standards in use in this field[92], present evidence that there is no (not even a single) unentangled state fitting the data.

For this particular case, a likelihood ratio (a typical frequentist's tool), is perfectly suited to quantify the evidence in favor of entanglement, as we showed in Chapter IV. One main advantage is that one always calculates the same quantity, no matter what one measured. We applied our proposed likelihood ratio to two standard sorts of entanglement verification measurement–quantum tomography and measurement of an entanglement witness– and showed that our quantity behaves like a semi-χ^2 distribution.

On the other hand, when trying to quantify entanglement from finite data, it is appropriate and useful to calculate a probability distribution over the amount of entanglement created in a given experiment. This requires use of the Bayesian methodology, where a prior distribution (over states) is chosen, the data converts this to a posterior distribution over states, which in turn automatically leads to a posterior probability distribution over entanglement. This Bayesian method we demonstrated in Chapter 5, where we used it to make precise how many measurements one needs to reliably estimate the amount of entanglement (and their error bar) in an experiment: simply said, data is sufficient when the estimates resulting from different priors agree to within their error bars. We applied our method to different states of four qubits that have been of recent experimental interest. This Bayesian method then competes with a standard frequentist method used in, for example, quantum-state tomography, namely the maximum likelihood method. The disadvantage of the latter is that it tends to lead to state estimates that are too pure, especially when relatively little data is available. Our Bayesian method, at the cost of requiring more analysis, corrects for this and leads to more balanced estimates of entanglement and purity.

Finally, we also addressed the issue of analyzing data on systems of many qubits. The Hilbert space grows exponentially with the number of qubits, and the current

bottleneck in multi-qubit experiments on trapped ions (by far the most advanced quantum information technology at the moment, in terms of number of qubits) is determined by the computer power and time needed to analyze the data, not by the experimental capabilities. We proposed and analyzed a known statistical method of model selection, using so-called information criteria, to drastically reduce the amount of computing needed, by tentatively limiting the number of parameters describing one's state. Crucially, that tentative model assumption is then verified by the data. The particular information criterion we used is based on an estimate of the so-called Kullback-Leibler divergence between the true probability distribution and the model.

So, provided one's a priori idea of how one's quantum source works turns out to be good, calculations become much easier to handle, so that even experiments with 14 qubits (the current record number of qubits[16] can be analyzed. In Chapter VI we checked that such calculations do lead to correct estimates of entanglement if the data verify one's model, whereas the data and information criteria will correctly flag incorrect models and resulting incorrect estimates of entanglement as unreliable. Moreover, the information criterion can identify which parameter values of the model were incorrect.

APPENDIX A

ENTANGLEMENT OF A COLOR-MIXED POLARIZATION-ENTANGLED STATE

Here we calculate explicitly the logarithmic negativity for the first example (in Chapter IV) to show that the entanglement of a color-mixed polarization-entangled state is still one ebit, in spite of the mixed nature of the state. The two-photon singlet state, maximally entangled in polarization – horizontal (H) or vertical (V) – but equally and classically mixed in color – green (G) or blue (B) – can be expressed in modes as

$$\rho = \frac{1}{2} |\phi_1\rangle \langle\phi_1| + \frac{1}{2} |\phi_2\rangle \langle\phi_2| , \tag{A.1}$$

where

$$|\phi_1\rangle = \frac{1}{\sqrt{2}} \left(|GH\rangle_A |GV\rangle_B - |GV\rangle_A |GH\rangle_B \right) \tag{A.2}$$

and

$$|\phi_2\rangle = \frac{1}{\sqrt{2}} \left(|BH\rangle_A |BV\rangle_B - |BV\rangle_A |BH\rangle_B \right) . \tag{A.3}$$

Expanding ρ in the basis of $(BH, BV, GH, GV)_A \otimes (BH, BV, GH, GV)_B \oplus (GH, GV)_A \otimes (GH, GV)_B$ gives

$$\rho = \begin{pmatrix} 0 & 0 & 0 & 0 & & & & \\ 0 & \frac{1}{4} & -\frac{1}{4} & 0 & & & 0 & \\ 0 & -\frac{1}{4} & \frac{1}{4} & 0 & & & & \\ 0 & 0 & 0 & 0 & & & & \\ & & & & 0 & 0 & 0 & 0 \\ & & 0 & & 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\ & & & & 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\ & & & & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{A.4}$$

The partial transpose of ρ in the same basis is then

$$\rho^{\Gamma} = \begin{pmatrix} 0 & 0 & 0 & -\frac{1}{4} & & & & \\ 0 & \frac{1}{4} & 0 & 0 & & & 0 & \\ 0 & 0 & \frac{1}{4} & 0 & & & & \\ -\frac{1}{4} & 0 & 0 & 0 & & & & \\ & & & & 0 & 0 & 0 & -\frac{1}{4} \\ & & 0 & & 0 & \frac{1}{4} & 0 & 0 \\ & & & & 0 & 0 & \frac{1}{4} & 0 \\ & & & & -\frac{1}{4} & 0 & 0 & 0 \end{pmatrix}. \tag{A.5}$$

The eight eigenvalues are easily found to be these: six times the eigenvalue $\frac{1}{4}$, and twice $-\frac{1}{4}$. This yields the logarithmic negativity of the state:

$$E_{\mathcal{N}} = \log_2 \left(6 \times \frac{1}{4} + 2 \times \left| -\frac{1}{4} \right| \right) = 1, \tag{A.6}$$

as we announced.

APPENDIX B

ASYMPTOTIC BEHAVIOR OF THE EXPECTATION VALUE OF THE POSTERIOR DISTRIBUTION

Suppose we are interested in a particular quantity $N(\rho)$, where ρ is a physical state. Suppose the number of measurements M is large and the posterior for N, $P(N)$, can be approximated by a normal distribution. Then the estimated value of N is where the maximum of $P(N)$ is. We have

$$\left.\frac{\mathrm{d}P(N)}{\mathrm{d}N}\right|_{\rho_{\max}} = \left.\left(\frac{\mathrm{d}P(N(\rho))}{\mathrm{d}\rho}\Big/\frac{\mathrm{d}N(\rho)}{\mathrm{d}\rho}\right)\right|_{\rho_{\max}} = 0 \quad\Longrightarrow\quad \left.\frac{\mathrm{d}\log P(\rho)}{\mathrm{d}\rho}\right|_{\rho_{\max}} = 0, \tag{B.1}$$

provided that $\mathrm{d}N/\mathrm{d}\rho$ is analytical in the range of ρ.

Recall that

$$P(\rho) = \frac{P_o(\rho)\mathcal{L}(\rho)}{\int d\rho' P_o(\rho')\mathcal{L}(\rho')}, \tag{B.2}$$

where

$$\mathcal{L}(\rho) = \prod_j p_j(\rho)^{F_j}, \tag{B.3}$$

$p_j(\rho)$ is the probability of the j'th result to be observed if the tested state is ρ and F_j is the number of times the j'th result is actually observed. We can approximate F_j's in terms of

$$F_j = Mp_j(\rho_{\mathrm{r}}) + \sqrt{Mp_j(\rho_{\mathrm{r}})[1-p_j(\rho_{\mathrm{r}})]}X_j, \tag{B.4}$$

where $\rho_{\rm r}$ is the real state and X_j is a normally distributed variable with variance 1. It means

$$\overline{X_j} = 0, \ \ \overline{X_j^2} = 1, \ \text{ for every } j. \tag{B.5}$$

The bar average $\overline{X_j}$, instead of the bracket average as in $\langle N \rangle$, indicates that the average is not taken over an ensemble of possible states. Instead, a random X_j value is generated each time a measurement record is collected, as M varies. Whether or not this average is to be taken depends on the specific questions and may be cleared up later. Moreover, X_j and X_k are independent of each other except for one constraint:

$$\sum_j X_j = 1. \tag{B.6}$$

We define

$$Q(\rho) = \prod_j p_j(\rho)^{p_j(\rho_{\rm r})} \tag{B.7}$$

and

$$C_M(\rho) = \prod_j p_j(\rho)^{\sqrt{p_j(\rho_{\rm r})[1-p_j(\rho_{\rm r})]}X_j}. \tag{B.8}$$

Then the likelihood function becomes

$$\mathcal{L}(\rho) = Q^M(\rho) C_M^{\sqrt{M}}(\rho). \tag{B.9}$$

Note that the subscript in C_M suggests the subtle dependence on M through variables X_j's. Hence $C_M(\rho)$ indeed corresponds to a single trial correction. However, note that

the behavior we study are not limited to a single trial. In fact, as in FIG.6.5, the red squares that correspond to $\left|\langle N_{1,2}\rangle_Z - \langle N_{1,2}\rangle_{GH}\right|$ really come from multiple trials that are affected by different noise profiles X_j's. Eventually the average over multiple trials is to be taken and the statistics of X_j's will be applied.

Since the integral in the denominator is just a constant, the zero derivative condition Eq. (B.1) gives

$$\left(M\frac{\mathrm{d}\log Q(\rho)}{\mathrm{d}\rho} + \sqrt{M}\frac{\mathrm{d}\log C_M(\rho)}{\mathrm{d}\rho} + \frac{\mathrm{d}\log P_0(\rho)}{\mathrm{d}\rho}\right)\Bigg|_{\rho_{\max}} = 0. \tag{B.10}$$

Note that at large M, whichever ρ that satisfies Eq. (B.1) or Eq. (B.10) is going to be very close to the actual state ρ_{r}, which is also the maximum of $\log Q(\rho)$, i.e. $\log Q(\rho_{\mathrm{r}}) = \log Q_{\max}$. We expand $\log Q(\rho)$ around ρ_{r} up to $O\left((\rho-\rho_{\mathrm{r}})^2\right)$:

$$\begin{aligned}\log Q(\rho) &\simeq \log Q(\rho_{\mathrm{r}}) - \frac{1}{2}(\rho-\rho_{\mathrm{r}})^T \cdot \frac{\mathrm{d}^2\log Q(\rho)}{\mathrm{d}\rho^2}\Bigg|_{\rho_{\mathrm{r}}} \cdot (\rho-\rho_{\mathrm{r}}) \\ &= L_Q - \frac{1}{2}(\rho-\rho_{\mathrm{r}})^T\widetilde{\alpha}(\rho-\rho_{\mathrm{r}}),\end{aligned} \tag{B.11}$$

where $L_Q = \log Q(\rho_{\mathrm{r}})$ and $\widetilde{\alpha} = -\,\mathrm{d}^2\log Q(\rho)/\mathrm{d}\rho^2\big|_{\rho_{\mathrm{r}}}$. The first derivative term is absent since ρ_{r} is the local maximum and therefore $\widetilde{\alpha} > 0$. This implies

$$\frac{\mathrm{d}\log Q(\rho)}{\mathrm{d}\rho} = -(\rho-\rho_{\mathrm{r}})^T\widetilde{\alpha}. \tag{B.12}$$

Similar expansion is applied for $\log P_0(\rho)$ and $\log C_M(\rho)$ so that

$$\frac{\mathrm{d}\log P_0(\rho)}{\mathrm{d}\rho} = \beta^T + (\rho-\rho_{\mathrm{r}})^T\widetilde{\gamma}, \tag{B.13}$$

where $\beta^T = \mathrm{d}\log P_0(\rho)/\mathrm{d}\rho\big|_{\rho_\mathrm{r}}$ and $\widetilde{\gamma} = \mathrm{d}^2\log P_0(\rho)/\mathrm{d}\rho^2\big|_{\rho_\mathrm{r}}$.

$$\frac{\mathrm{d}\log C_M(\rho)}{\mathrm{d}\rho} = \zeta^T + (\rho - \rho_\mathrm{r})^T\widetilde{\eta}, \tag{B.14}$$

where $\zeta^T = \mathrm{d}\log C_M(\rho)/\mathrm{d}\rho\big|_{\rho_\mathrm{r}}$ and $\widetilde{\eta} = \mathrm{d}^2\log C_M(\rho)/\mathrm{d}\rho^2\big|_{\rho_\mathrm{r}}$.

Therefore the zero derivative condition becomes

$$-M(\rho - \rho_\mathrm{r})^T\widetilde{\alpha} + \sqrt{M}\left(\zeta^T + (\rho - \rho_\mathrm{r})^T\widetilde{\eta}\right) + \beta^T + (\rho - \rho_\mathrm{r})^T\widetilde{\gamma} = 0. \tag{B.15}$$

Solving it for the maximum state:

$$\rho_\mathrm{max} = \rho_\mathrm{r} + \delta\rho,$$

where

$$\delta\rho = \frac{1}{\sqrt{M}}\widetilde{\alpha}^{-1}\zeta + \frac{1}{M}\widetilde{\alpha}^{-1}(\beta + \widetilde{\eta}\widetilde{\alpha}^{-1}\zeta) + O\left(\frac{1}{M^{3/2}}\right). \tag{B.16}$$

Now we suppose that $N(\rho_\mathrm{max})$ is also where the largest probability of $N(\rho)$ is, which is again assumed to be the expectation value, $\langle N\rangle$. Then

$$\begin{aligned} N(\rho_\mathrm{max}) &\simeq N(\rho_\mathrm{r}) + \left.\frac{\mathrm{d}N}{\mathrm{d}\rho}\right|_{\rho_\mathrm{r}} \cdot \delta\rho \\ &= N_\mathrm{r} + \frac{1}{\sqrt{M}}\lambda^T\widetilde{\alpha}^{-1}\zeta + \frac{1}{M}\lambda^T\widetilde{\alpha}^{-1}(\beta + \widetilde{\eta}\widetilde{\alpha}^{-1}\zeta) + O\left(\frac{1}{M^{3/2}}\right), \end{aligned} \tag{B.17}$$

where $N_{\rm r} = N(\rho_{\rm r})$ and $\lambda = {\rm d}N(\rho)/{\rm d}\rho|_{\rho_{\rm r}}$. Since $\widetilde{\alpha}$, β, ζ, $\widetilde{\eta}$ and λ are all fixed and presumably nonzero, the behavior of $\langle N\rangle$ as it approaches its true value $N_{\rm r}$ goes

$$|\langle N\rangle - N_{\rm r}| \sim 1/\sqrt{M}, \tag{B.18}$$

in large M limit.

Note that the first correction in Eq. (B.16), $\widetilde{\alpha}^{-1}\zeta/\sqrt{M}$ is merely influenced by the fluctuation of the data through ζ and the shape of the likelihood function through $\widetilde{\alpha}$, which is determined by the measurement setup. It implies a consistent behavior with no regard of the choice of the prior distribution. Yet the second correction does. We label $\rho_{\rm max}$ and β with subscript Z or GH to differentiate the priors. We have

$$\rho_{Z{\rm max}} - \rho_{GH{\rm max}} = \frac{\widetilde{\alpha}^{-1}}{M}(\beta_Z - \beta_{GH}). \tag{B.19}$$

From the previous analysis we realise that the M-dependence in the difference in the state ρ will carry on to the difference in the negativity N. Eq. (B.19) indicates

$$|\langle N\rangle_Z - \langle N\rangle_{GH}| \sim 1/M. \tag{B.20}$$

APPENDIX C

ASYMPTOTIC BEHAVIOR OF THE MAXIMUM LIKELIHOOD ESTIMATION

Since we are concerned with the large M region, we assume that ρ_{MLE} is not on the boundary, so that it satisfies

$$\left.\frac{\mathrm{d}\log\mathcal{L}(\rho)}{\mathrm{d}\rho}\right|_{\rho_{\mathrm{MLE}}} = 0. \tag{C.1}$$

Using the same expansion in the vicinity of the real state ρ_{r} as in the last section, we obtain

$$\rho_{\mathrm{MLE}} = \rho_{\mathrm{r}} + \frac{1}{\sqrt{M}}\widetilde{\alpha}^{-1}\zeta + \frac{1}{M}\widetilde{\alpha}^{-1}\widetilde{\eta}\widetilde{\alpha}^{-1}\zeta + O\left(\frac{1}{M^{3/2}}\right). \tag{C.2}$$

Compared to Eq. (B.16), the only difference in the higher-order terms is that the term containing β is missing. When the negativity $N(\rho)$ is considered, similar conclusions can be reached:

$$|N(\rho_{\mathrm{MLE}}) - N_{\mathrm{r}}| \sim 1/\sqrt{M}, \tag{C.3}$$

and

$$\left|N(\rho_{\mathrm{MLE}}) - \langle N\rangle_{Z,GH}\right| \sim 1/M. \tag{C.4}$$

REFERENCES CITED

[1] Feynman, R. P. *International Journal of Theoretical Physics* **1982**, *21*, 467.

[2] Lloyd, S. *Science* **1996**, *273*, 1073.

[3] Deutsch, D.; Jozsa, R. *Proceedings: Mathematical and Physical Sciences* **1992**, *439*, 553.

[4] Simon, D. R. *Foundations of computer science, 1994 Proceedings., 35th Annual Symposium* **1994**, 116.

[5] Grover, L. K. *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* **1996**, 212.

[6] Cleve, R.; Ekert, A.; Macchiavello, C.; Mosca, M. *Proc. R. Soc. Lond. A* **1998**, *454*, 339.

[7] Shor, P. W. *Siam J. Comput.* **1999**, *26*, 1484.

[8] Gisin, N.; Ribordy, G.; Tittel, W.; Zbinden, H. *Rev. Mod. Phys.* **2002**, *74*, 145.

[9] Barenco, A.; Deutsch, D.; Ekert, A.; Jozsa, R. *Phys. Rev. Lett.* **1995**, *74*, 4083.

[10] Sleator, T.; Weinfurter, H. *Phys. Rev. Lett.* **1995**, *74*, 4087.

[11] Cirac, J. I.; Zoller, P. *Phys. Rev. Lett.* **1995**, *74*, 4091.

[12] Monroe, C.; Meekhof, D. M.; King, B. E.; Itano, W. M.; Wineland, D. J. *Phys. Rev. Lett.* **1995**, *75*, 4714.

[13] O'Brien, J. L.; Pryde, B. J.; White, A. G.; Ralph, T. C.; Branning, D. *Nature* **2003**, *426*, 264.

[14] Pittman, T. B.; Fitch, M. J.; Jacobs, B. C.; Franson, J. D. *Phys. Rev. A* **2003**, *68*, 032316.

[15] Plantenberg, J. H.; de Groot, P. C.; Harmans, C. J. P. M.; Mooij, J. E. *Nature* **2007**, *447*, 836.

[16] Monz, T.; Schindler, P.; Barreiro, J. T.; Chwalla, M.; Nigg, D.; Coish, W. A.; Harlander, M.; Hänsel, W.; Hennrich, M.; Blatt, R. *Phys. Rev. Lett.* **2011**, *106*, 130506.

[17] Ekert, A. K. *Phys. Rev. Lett.* **1991**, *67*, 661.

[18] Bennett, C. H.; Brassard, G.; Crépeau, C.; Jozsa, R.; Peres, A.; Wootters, W. K. *Phys. Rev. Lett.* **1993**, *70*, 1895.

[19] Schumacher, B. *Phys. Rev. A* **1996**, *54*, 2614.

[20] Clauser, J. F.; Dowling, J. P. *Phys. Rev. A* **1996**, *53*, 4587.

[21] Ahn, J.; Weinacht, T. C.; Bucksbaum, P. H. *Science* **2000**, *287*, 463.

[22] Curty, M.; Lewenstein, M.; Lütkenhaus, N. *Phys. Rev. Lett.* **2004**, *92*, 217903.

[23] Braunstein, S. L.; Caves, C. M.; Jozsa, R.; Linden, N.; Popescu, S.; Schack, R. *Phys. Rev. Lett.* **1999**, *83*, 1054.

[24] Gross, D.; Liu, Y.-K.; Flammia, S.; Becker, S.; Eisert, J. *Phys. Rev. Lett.* **2010**, *105*, 150401.

[25] Cramer, M.; Plenio, M. B.; Flammia, S. T.; Somma, R.; Gross, D.; Bartlett, S. D.; Landon-Cardinal, O.; Poulin, D.; Liu, Y.-K. *Nature Communications* **2010**, *1*, 149.

[26] D'Ariano, G. M.; Yuen, H. P. *Phys. Rev. Lett.* **1996**, *76*, 2832.

[27] Fano, U. *Rev. Mod. Phys.* **1957**, *29*, 74.

[28] Vogel, K.; Risken, H. *Phys. Rev. A* **1989**, *40*, 2847.

[29] Wigner, E. *Phys. Rev.* **1932**, *40*, 749.

[30] Cahill, K. E.; Glauber, R. J. *Phys. Rev.* **1969**, *177*, 1857.

[31] Cahill, K. E.; Glauber, R. J. *Phys. Rev.* **1969**, *177*, 1882.

[32] Smithey, D. T.; Beck, M.; Raymer, M. G.; Faridani, A. *Phys. Rev. Lett.* **1993**, *70*, 1244.

[33] Raymer, M. G.; Beck, M.; McAlister, D. *Phys. Rev. Lett.* **1994**, *72*, 1137.

[34] Ashburn, J. R.; Cline, R. A.; van der Burgt, P. J. M.; Westerveld, W. B.; Risley, J. S. *Phys. Rev. A* **1990**, *41*, 2407.

[35] Dunn, T. J.; Walmsley, I. A.; Mukamel, S. *Phys. Rev. Lett.* **1995**, *74*, 884.

[36] Wallentowitz, S.; Vogel, W. *Phys. Rev. Lett.* **1995**, *75*, 2932.

[37] Munroe, M.; Boggavarapu, D.; Anderson, M. E.; Raymer, M. G. *Phys. Rev. A* **1995**, *52*, R924.

[38] Janicke, U.; Wilkens, M. *J. Mod. Opt.* **1995**, *42*, 2183.

[39] Schiller, S.; Breitenbach, G.; Pereira, S. F.; Müller, T.; Mlynek, J. *Phys. Rev. Lett.* **1996**, *77*, 2933.

[40] Leibfried, D.; Meekhof, D. M.; King, B. E.; Monroe, C.; Itano, W. M.; Wineland, D. J. *Phys. Rev. Lett.* **1997**, *77*, 4281.

[41] Kurtsiefer, C.; Pfau, T.; Mlynek, J. *Nature* **1997**, *386*, 150.

[42] Breitenbach, G.; Schiller, S.; Mlynek, J. *Nature* **1997**, *387*, 471.

[43] Chuang, I. L.; Gershenfeld, N.; Kubinec, M. G.; Leung, D. W. *Proc. R. Soc. Lond. A* **1998**, *454*, 447.

[44] White, A. G.; James, D. F. V.; Eberhard, P. H.; Kwiat, P. G. *Phys. Rev. Lett.* **1999**, *83*, 3103.

[45] Banaszek, K.; Radzewicz, C.; Wódkiewicz, K.; Krasiński, J. S. *Phys. Rev. A* **1999**, *60*, 674.

[46] Lvovsky, A. I.; Hansen, H.; Aichele, T.; Mlynek, O. B. J.; Schiller, S. *Phys. Rev. Lett.* **2001**, *87*, 050402.

[47] Hradil, Z. *Phys. Rev. A* **1997**, *55*, R1561.

[48] Schack, R.; Brun, T. A.; Caves, C. M. *Phys. Rev. A* **2001**, *64*, 014305.

[49] Caves, C. M.; Fuchs, C. A.; Schack, R. *J. Math. Phys.* **2002**, *43*, 4537.

[50] Schrödinger, E. *Die Naturwissenschaften* **1935**, *23*, 807,823,844.

[51] Einstein, A.; Podolsky, B.; Rosen, N. *Phys. Rev.* **1935**, *47*, 777.

[52] Bell, J. S. *Physics* **1964**, *1*, 195.

[53] Clauser, J. F.; Horne, M. A.; Shimony, A.; Holt, R. A. *Phys. Rev. Lett.* **1969**, *23*, 880.

[54] Peres, A. *Phys. Rev. Lett.* **1996**, *77*, 1413.

[55] Horodecki, M.; Horodecki, P.; Horodecki, R. *Phys. Lett. A* **1996**, *223*, 1.

[56] Horodecki, P.; Ekert, A. *Phys. Rev. Lett.* **2002**, *89*, 127902.

[57] Gühne, O.; Tóth, G. *Physics Reports* **2009**, *474*, 1.

[58] Horodecki, R.; Horodecki, P.; Horodecki, M.; Horodecki, K. *Rev. Mod. Phys.* **2009**, *81*, 865.

[59] Bouwmeester, D.; Pan, J.; Mattle, K.; Eibl, M.; Weinfurter, H.; Zeilinger, A. *Nature* **1997**, *390*, 575.

[60] Boschi, D.; Branca, S.; Martini, F. D.; Hardy, L.; Popescu, S. *Phys. Rev. Lett.* **1998**, *80*, 1121.

[61] Furusawa, A.; Sorensen, J. L.; Braunstein, S. L.; Fuchs, C. A.; Kimble, H. J.; Polzik, E. S. *Science* **1998**, *282*, 706.

[62] Riebe, M.; Häffner, H.; Roos, C. F.; Hänsel, W.; Benheim, J.; Lancaster, G. P. T.; Körber, T. W.; Becher, C.; Schmidt-Kaler, F.; James, D. F. V.; Blatt, R. *Nature* **2004**, *429*, 734.

[63] Barrett, M. D.; Chiaverini, J.; Schaetz, T.; Britton, J.; Itano, W. M.; Jost, J. D.; Knill, E.; Langer, C.; Leibfried, D.; Ozeri, R.; Wineland, D. J. *Nature* **2004**, *429*, 737.

[64] Dür, W.; Vidal, G.; Cirac, J. I. *Phys. Rev. A* **2000**, *62*, 062314.

[65] Verstraete, F.; Dehaene, J.; Moor, B. D.; Verschelde, H. *Phys. Rev. A* **2002**, *65*, 052112.

[66] Bouwmeester, D.; Pan, J.; Daniell, M.; Weinfurter, H.; Zeilinger, A. *Phys. Rev. Lett.* **1999**, *82*, 1345.

[67] Pan, J.; Bouwmeester, D.; Daniell, M.; Weinfurter, H.; Zeilinger, A. *Nature* **2000**, *403*, 515.

[68] Pan, J.; Daniell, M.; Gasparoni, S.; Weihs, G.; Zeilinger, A. *Phys. Rev. Lett.* **2001**, *86*, 4435.

[69] Lu, C.-Y.; Zhou, X.-Q.; Gühne, O.; Gao, W.-B.; Zhang, J.; Yuan, Z.-S.; Goebel, A.; Yang, T.; Pan, J.-W. *Nature Phys.* **2007**, *3*, 91.

[70] Gao, W.-B.; Lu, C.-Y.; Yao, X.-C.; Xu, P.; Gühne, O.; Goebel, A.; Chen, Y.-A.; Peng, C.-Z.; Chen, Z.-B.; Pan, J.-W. *Nature Phys.* **2010**, *6*, 331.

[71] Sackett, C. A.; Kielpinski, D.; King, B. E.; Langer, C.; Meyer, V.; Myatt, C. J.; Rowe, M.; Turchette, Q. A.; Itano, W. M.; Wineland, D. J.; Monroe, C. *Nature* **2000**, *404*, 256.

[72] Rauschenbeutel, A.; Nogues, G.; Osnaghi, S.; Bertet, P.; Brune, M.; Raimond, J.-M.; Haroche, S. *Science* **2000**, *288*, 2024.

[73] Leibfried, D.; Knill, E.; Seidelin, S.; Britton, J.; Blakstad, R. B.; Chiaverini, J.; Hume, D. B.; Itano, W. M.; Jost, J. D.; Langer, C.; Ozeri, R.; Reichle, R.; Wineland, D. J. *Nature* **2005**, *438*, 639.

[74] Häffner, H.; Hänsel, W.; Roos, C. F.; Benhelm, J.; al kar, D. C.; Chwalla, M.; Körber, T.; Rapol, U. D.; Reibe, M.; Schmidt, P. O.; Becher, C.; Gühne, O.; ür, W.; Blatt, R. *Nature* **2005**, *438*, 643.

[75] Vidal, G. *J. Mod. Opt.* **2000**, *47*, 355.

[76] Bennett, C. H.; DiVincenzo, D. P.; Smolin, J. A.; Wootters, W. K. *Phys. Rev. A* **1996**, *54*, 3824.

[77] Shimony, A. *Ann. N. Y. Acad. Science* **1995**, *755*, 675.

[78] Barnum, H.; Linden, N. *J. Phys. A: Math. Gen.* **2001**, *34*, 6768.

[79] Wei, T.-C.; Goldbart, P. M. *Phys. Rev. A* **2003**, *68*, 042307.

[80] Wootters, W. K. *Phys. Rev. Lett.* **1998**, *80*, 2245.

[81] Vedral, V.; Plenio, M. B. *Phys. Rev. A* **1998**, *57*, 1619.

[82] Vidal, G.; Werner, R. F. *Phys. Rev. A* **2002**, *65*, 032314.

[83] Mintert, F.; Kuś, M.; Buchleitner, A. *Phys. Rev. Lett* **2005**, *95*, 260502.

[84] Coffman, V.; Kundu, J.; Wootters, W. K. *Phys. Rev. A* **2000**, *61*, 052306.

[85] Bennett, C. H.; Popescu, S.; Rohrlich, D.; Smolin, J. A.; Thapliyal, A. V. *Phys. Rev. A* **2000**, *63*, 012307.

[86] Peres, A. *Quantum Theory: Concepts and methods*; Kluwer Academic Publishers, 1995.

[87] Horodecki, M.; Horodecki, P.; Horodecki, R. *Phys. Rev. Lett.* **1998**, *80*, 5239.

[88] Sabin, S.; Garcia-Alcaine, G. *Eur. Phys. J. D* **2008**, *48*, 435.

[89] Yin, J. O. S.; van Enk, S. J. *Phys. Rev. A* **2011**, *83*, 022326.

[90] Walborn, S. P.; Ribeiro, P. H. S.; Davidovich, L.; Mintert, F.; Buchleitner, A. *Nature* **2006**, *440*, 1022.

[91] Horodecki, P. *Phys. Rev. A* **2003**, *68*, 052101.

[92] van Enk, S. J.; Lütkenhaus, N.; Kimble, H. J. *Phys. Rev. A* **2007**, *75*, 052318.

[93] Kiesel, N.; Schmid, C.; Tóth, G.; Solano, E.; Weinfurter, H. *Phys. Rev. Lett.* **2007**, *98*, 063604.

[94] Papp, S. B.; Choi, K. S.; Deng, H.; Lougovski, P.; van Enk, S. J.; Kimble, H. J. *Science* **2009**, *324*, 764.

[95] Lewenstein, M.; Kraus, B.; Cirac, J. I.; Horodecki, P. *Phys. Rev. A* **2000**, *62*, 052310.

[96] Gühne, O.; Lütkenhaus, N. *Phys. Rev. Lett.* **2006**, *96*, 170502.

[97] Moroder, T.; Gühne, O.; Lütkenhaus, N. *Phys. Rev. A* **2008**, *78*, 032326.

[98] Dicke, R. H. *Phys. Rev.* **1954**, *93*, 99.

[99] Tóth, G. *J. Opt. Soc. Am. B* **2007**, *24*, 275.

[100] Ou, Z. Y.; Mandel, L. *Phys. Rev. Lett.* **1988**, *61*, 50.

[101] Kwiat, P. G.; Mattle, K.; Weinfurter, H.; Zeilinger, A.; Sergienko, A. V.; Shih, Y. *Phys. Rev. Lett.* **1995**, *75*, 4337.

[102] Sergienko, A. V.; Shih, Y. H.; Rubin, M. H. *J. Opt. Soc. Am. B* **1995**, *12*, 859.

[103] Kwiat, P. G.; Waks, E.; White, A. G.; Appelbaum, I.; Eberhard, P. H. *Phys. Rev. A* **1999**, *60*, R773.

[104] U'Ren, A. B.; Silberhorn, C.; Banaszek, K.; Walmsley, I. A.; Erdmann, R.; Grice, W. P.; Raymer, M. G. *Las. Phys.* **2005**, *15*, 146.

[105] Zanardi, P. *Phys. Rev. A* **2002**, *65*, 042101.

[106] van Enk, S. J. *Phys. Rev. A* **2003**, *67*, 022303.

[107] Shi, Y. *Phys. Rev. A* **2003**, *67*, 024301.

[108] Tan, S. M.; Walls, D. F.; Collett, M. J. *Phys. Rev. Lett.* **1991**, *66*, 252.

[109] Hardy, L. *Phys. Rev. Lett.* **1994**, *73*, 2279.

[110] Björk, G.; Jonsson, P.; Sánchez-Soto, L. L. *Phys. Rev. A* **2001**, *64*, 042106.

[111] Hessmo, B.; Usachev, P.; Heydari, H.; Björk, G. *Phys. Rev. Lett.* **2004**, *92*, 180401.

[112] Babichev, S. A.; Appel, J.; Lvovsky, A. I. *Phys. Rev. Lett.* **2004**, *92*, 193601.

[113] van Enk, S. J. *Phys. Rev. A* **2005**, *72*, 064306.

[114] Plenio, M. B. *Phys. Rev. Lett.* **2005**, *95*, 090503.

[115] Grice, W. P.; Walmsley, I. A. *Phys. Rev. A* **1997**, *56*, 1627.

[116] Raymer, M. G.; Noh, J.; Banaszek, K.; Walmsley, I. A. *Phys. Rev. A* **2005**, *72*, 023825.

[117] Legero, T.; Wilk, T.; Kuhn, A.; Rempe, G. *Adv. At. Mol. Opt. Phys.* **2006**, *53*, 253.

[118] Blow, K. J.; r. Loudon,; Phoenix, S. J. D.; Shepherd, T. J. *Phys. Rev. A* **1990**, *42*, 4102.

[119] Hong, C. K.; Mandel, L. *Phys. Rev. A* **1985**, *31*, 2409.

[120] Ou, Z. Y.; Zou, X. Y.; Wang, L. J.; Mandel, L. *Phys. Rev. A* **1990**, *42*, 2957.

[121] Law, C. K.; Walmsley, I. A.; Eberly, J. H. *Phys. Rev. Lett.* **2000**, *84*, 5304.

[122] Parker, S.; Bose, S.; Plenio, M. B. *Phys. Rev. A* **2000**, *61*, 032305.

[123] Spreeuw, R. J. C. *Phys. Rev. A* **2001**, *63*, 062302.

[124] Lvovsky, A. I.; Wasilewski, W.; Banaszek, K. *J. Mod. Opt.* **2005**, *54*, 721.

[125] Jost, J. D.; Home, J. P.; Amini, J. M.; Hanneke, D.; Ozeri, R.; Langer, C.; Bollinger, J. J.; Leibfried, D.; Wineland, D. J. *Nature* **2009**, *459*, 683.

[126] DiCarlo, L.; Chow, J. M.; Gambetta, J. M.; Bishop, L. S.; Johnson, B. R.; Schuster, D. I.; Majer, J.; Blais, A.; Frunzio, L.; Girvin, S. M.; Schoelkopf, R. J. *Nature* **2009**, *460*, 240.

[127] Ansmann, M.; Wang, H.; Bialczak, R. C.; Hofheinz, M.; Lucero, E.; Neeley, M.; O'Connell, A. D.; Sank, D.; Weides, M.; Wenner, J.; Cleland, A. N.; Martinis, J. M. *Nature* **2009**, *461*, 504.

[128] Hofstetter, L.; Csonka, S.; Nygård, J.; Schönenberger, C. *Nature* **2009**, *461*, 960ö.

[129] Ourjoumtsev, A.; Ferreyrol, F.; Tualle-Brouri, R.; Grangier, P. *Nature Phys.* **2009**, *5*, 189.

[130] Fedrizzi, A.; Ursin, R.; Herbst, T.; Nespoli, M.; Prevedel, R.; Scheidl, T.; Tiefenbacher, F.; Jennewein, T.; Zeilinger, A. *Nature Phys.* **2009**, *5*, 389.

[131] Böhi, P.; Riedel, M. F.; Hoffrogge, J.; Reichel, J.; Hänsch, T. W.; Treutlein, P. *Nature Phys.* **2009**, *5*, 592.

[132] Amselem, E.; Bourennane, M. *Nature Phys.* **2009**, *5*, 748.

[133] Matthews, J. C. F.; Politi, A.; Stefanov, A.; O'Brien, J. L. *Nature Photonics* **2009**, *3*, 346.

[134] Janousek, J.; Wagner, K.; Morizur, J.-F.; Treps, N.; Lam, P. K.; Harb, C. C.; Bachor, H. *Nature Photonics* **2009**, *3*, 399.

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

[136] Olmschenk, S.; Matsukevich, D. N.; Maunz, P.; Hayes, D.; Duan, L.-M.; Monroe, C. *Science* **2009**, *323*, 486.

[137] Coelho, A. S.; Barbosa, F. A. S.; Cassemiro, K. N.; Villar, A. S.; Martinelli, M.; Nussenzveig, P. *Science* **2009**, *326*, 823.

[138] Terhal, B. M. *Phys. Lett. A* **2000**, *271*, 319.

[139] Hyllus, P.; Gühne, O.; Bruß, D.; Lewenstein, M. *Phys. Rev. A* **2005**, *72*, 012321.

[140] van Dam, W.; Gill, R. D.; Grunwald, P. D. *IEEE Trans. Inf. Th.* **2005**, *51*, 2812.

[141] Zhang, Y.; Knill, E.; Glancy, S. *Phys. Rev. A* **2010**, *81*, 032117.

[142] Geweke, J. F.; Singleton, K. J. *J. Am. Stat. Assoc.* **1980**, *75*, 133.

[143] Werner, R. F. *Phys. Rev. A* **1989**, *40*, 4277.

[144] Renner, R. *Nature Phys.* **2007**, *3*, 645.

[145] Renes, J. M.; Blume-Kohout, R.; Scott, A. J.; Caves, C. M. *Journal of Mathematical Physics* **2004**, *45*, 2171.

[146] Blume-Kohout, R.; Yin, J. O. S.; van Enk, S. J. *Phys. Rev. Lett.* **2010**, *105*, 170501.

[147] Blume-Kohout, R. *New J. Phys.* **2010**, *12*, 043034.

[148] Jungnitsch, B.; Moroder, T.; Gühne, O. *arXiv:1010.6049v1 [quant-ph]* **2010**.

[149] Życzkowski, K.; Horodecki, P.; Sanpera, A.; Lewenstein, M. *Phys. Rev. A* **1998**, *58*, 883.

[150] Życzkowski, K. *Phys. Rev. A* **1999**, *60*, 3496.

[151] Roman, S. *Advanced linear algebra*; New York: Springer-Verlag, 1992.

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

[153] Eisert, J.; Briegel, H. J. *Phys. Rev. A* **2001**, *64*, 022306.

[154] Shi, Y. *quant-ph/0201079* **2002**.

[155] Verstraete, F.; Dehaene, J.; Moor, B. D. *Phys. Rev. A* **2003**, *68*, 012103.

[156] Demkowicz-Dobrzanski, R.; Buchleitner, A.; Kus, M.; Mintert, F. *Phys. Rev. A* **2006**, *74*, 052303.

[157] Lamata, L.; Leon, J.; Salgado, D.; Solano, E. *Phys. Rev. A* **2007**, *75*, 022318.

[158] Li, D.; Li, X.; Huang, H.; Li, X. *Phys. Rev. A* **2007**, *76*, 052311.

[159] Wang, X.-W.; Yang, G. J. *Phys. Rev. A* **2008**, *78*, 024301.

[160] Kruszynska, C.; Kraus, B. *Phys. Rev. A* **2009**, *79*, 052304.

[161] Hiesmayr, B. C.; Huber, M.; Krammer, P. *Phys. Rev. A* **2009**, *79*, 062308.

[162] Sharma, S. S.; Sharma, N. K. *Phys. Rev. A* **2009**, *79*, 062323.

[163] Horodecki, M.; Horodecki, P.; Horodecki, R. *Phys. Rev. Lett.* **2000**, *84*, 4260.

[164] Dür, W.; Cirac, J. I.; Lewenstein, M.; Bruß, D. *Phys. Rev. A* **2000**, *61*, 062313.

[165] Smolin, J. A. *Phys. Rev. A* **2001**, *63*, 032306.

[166] Roberts, G. O.; Gelman, A.; Gilks, W. R. *Ann. Appl. Probab.* **1997**, *7*, 110.

[167] Paris, M.; Řeháček, J. *Quantum state estimation (Lecture notes in physics)*; Springer, 2004.

[168] Klumper, A.; Schadschneider, A.; Zittartz, J. *J. Phys. A* **1991**, *24*, L955.

[169] Fannes, M.; Nachtergaele, B.; Werner, R. F. *Comm. Math. Phys.* **1992**, *144*, 443.

[170] Perez-Garcia, D.; Verstraete, F.; Wolf, M. M.; Cirac, J. I. *Quantum Inf. Comput.* **2007**, *7*, 401.

[171] Tóth, G.; Wieczorek, W.; Gross, D.; Krischek, R.; Schwemmer, C.; Weinfurter, H. *Phys. Rev. Lett.* **2010**, *105*, 250403.

[172] Jeffreys, H. *Theory of probability*; Oxford University Press, 3rd Ed., 1961.

[173] Burnham, K. P.; Anderson, D. R. *Model selection and multimodel inference: a practical information-theoretic approach*; Springer-Verlag, 2002.

[174] Zucchini, W. *J. Math. Psych.* **2000**, *44*, 41.

[175] Kullback, S.; Leibler, R. A. *Ann. Math. Stat.* **1951**, *22*, 79.

[176] Akaike, H. *IEEE Transactions on Automatic Control* **1974**, *19*, 716.

[177] Takeuchi, K. *Suri-Kagaku (Mathematic Sciences)* **1976**, *153*, 12.

[178] Liddle, A. R. *Mon. Not. Roy. Astron. Soc. Lett.* **2007**, *377*, L74.

[179] Lougovski, P.; van Enk, S. J. *Phys. Rev. A* **2009**, *80*, 052324.

[180] Usami, K.; Nambu, Y.; Tsuda, Y.; Matsumoto, K.; Nakamura, K. *Phys. Rev. A* **2003**, *68*, 022314.

[181] Sugiura, N. *Commun. Stat. Theo. Math. A* **1978**, *7*, 13.

[182] Choi, K. S.; Goban, A.; Papp, S. B.; van Enk, S. J.; Kimble, H. J. *Nature* **2010**, *468*, 412.

[183] Kaznady, M. S.; James, D. F. V. *Phys. Rev. A* **2009**, *79*, 022109.

[184] Stoica, P.; Eykhoff, P.; Janssen, P.; Söderström, T. *International Journal of Control* **1986**, *43*, 1841.

[185] Stone, M. *J. Roc. Stat. Soc. B* **1977**, *39*, 44.