Experimental Entanglement Witness Family Measurement and Theoretical Aspects of Quantum Tomography

Dai Jibo

2015

Experimental Entanglement Witness Family Measurement and Theoretical Aspects of Quantum Tomography

DAI JIBO

B. Sc. (Hons.) and B. Eng. (Hons.), National University of Singapore

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DRGREE OF DOCTOR OF PHILOSOPHY

Centre for Quantum Technologies

National University of Singapore

2015

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Dai Jibo

15 April 2015

Acknowledgments

First and foremost, I would like to thank my supervisor Prof. Berthold-Georg Englert for accepting me as a Ph. D. student at the Centre for Quantum Technologies (CQT), and tirelessly guiding and supporting me throughout my candidature. Nothing in this thesis would be possible without you firstly providing me with the opportunity to learn and work under your supervision. I am deeply grateful for your trust in me that is always the source of motivation for me in face of obstacles. Thank you so much for the time you spend on me, for the invaluable guidance on my Ph. D. project, insights in physics, and wisdom of life that you have shared with me, which have helped me tremendously along the way and encouraged me never to give up. You are the role model that I deeply admire and will always try to follow.

I'd like to thank Prof. Feng Yuan Ping and Prof. Gong Jiangbin for willingly writing the recommendation letters for me and encouraging me to apply for Ph. D. studies in CQT. Thank you also for teaching me in my undergraduate studies and guiding me in my final year project, for which the good time that I had while studying physics as an undergraduate student is one of the reasons that urged me to pursue further studies in physics eventually. And I want to thank Prof. Kwek Leong Chuan for interviewing me and recommending me into the CQT Ph. D. program. I also wish to thank Prof. Christian Kurtsiefer and Prof. Gong Jiangbin for taking time out of their busy schedule to serve in my thesis advisory committee, and providing me with help and support when I need them.

For the experimental part of the thesis, I am very much indebted to Dr. Leonid Krivitsky at the Institute of Data Storage (DSI), A*STAR. Thank you for your patience in teaching me and for all the wonderful experimental techniques that you shared with me. I am greatly thankful for the time that you spend with me in the

lab, as well as during our numerous enjoyable discussions. Thank you for never giving up on me, and persistently teaching me new experimental skills. The experience of being a member in your team is truly memorable and enriching.

I would also like to thank Asst/Prof. Hui Khoon NG for guiding me on a substantial part of the thesis. Thank you for the numerous help you provided me with and for the many effective discussions we had. Thank you also for the opportunities that you entrusted on me, to supervise the high school students on science projects. The experience of it is both enjoyable as well as rewarding. Moreover, thank you for critically reading this thesis and giving me many invaluable comments and suggestions on how to improve it.

Throughout my Ph. D. studies, I also received a lot of help from my colleagues in CQT. A special thanks to Dr. TEO Yong Siah, for being always supportive and helpful when I am faced with difficulties. I wish to extend my sincere appreciation to all my colleagues who helped me in one way or another. While it may not be possible to name all of them, I would like to thank Dr. SHANG Jiangwei, Dr. ZHU Huangjun, LI Xikun, Max SEAH Yi-Lin, TAN Wei Hou, Dr. HAN Rui and so on. I would also like to thank Dr. Dmitry KALASHNIKOV in DSI for sharing with me his experimental expertise.

I'd like to especially acknowledge LEN Yink Loong, who worked together with me on most of the project in this thesis. Thank you for spending time with me on the experiment in the dark lab. Thank you for helping me with loads of checking and calculations. Thank you for many of the lunch-time talks and discussions that are both enlightening and thought-provoking.

I would like to acknowledge the financial support from Centre for Quantum Technologies, a Research Centre of Excellence funded by the Ministry of Education and the National Research Foundation of Singapore. I am grateful to all the administrative staff at CQT for providing numerous timely help and a favorable environment at CQT where I can learn and work comfortably.

Last but not least, I wish to thank again all the professors, my colleagues, my friends and my family, for your help given to me, care and love shone on me have always been the source of inspiration for me. Thank you.

Abstract and Summary

Quantum state tomography is a central and recurring theme in quantum information science and quantum computation. In a typical scenario, a source emits a certain desired state which carries the information, or is required for the computational task. Quantum state tomography is needed for the verification and identification of the state emitted by the source. In the first part of the thesis, we focus on the efficient detection of entanglement, a key resource in many quantum information processing tasks. We report an experiment in which one determines, with least tomographic effort, whether an unknown two-photon polarization state is entangled or separable. The method measures whole families of optimal entanglement witnesses at once. We introduce adaptive measurement schemes that greatly speed up the entanglement detection. The witness family measurement enables informationally complete (IC) quantum state tomography if the individual family gives inconclusive results. On average, only about three families need to be measured before the entanglement is detected and the IC state tomography is hardly necessary.

However, in a realistic experiment, not only the quantum state to be reconstructed, but additional parameters in the experimental setup are also unknown, for example, the efficiency of the detectors, the total number of copies emitted, etc. Furthermore, the assumption of a closed quantum system is also only an approximation, and there are often the ignored bath degrees of freedom which interact with the system. The second part of the thesis aims at these aspects. For the former aspect, based on the idea of credible regions, we construct joint optimal error regions for the system state and the other unknown parameters. By marginalizing over the nuisance parameters, one can obtain a marginal likelihood which only depends on the parameter of interest. We illustrate the method and technique with

several examples. Some of them display unusual features in the likelihood function. For the latter aspect, we show how one uses ideas from quantum tomography or state estimation to deduce a reasonable and consistent system-bath state. In typical experimental situations, such a state turns out to be uncorrelated or almost uncorrelated between the system and the bath.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Before the advent of quantum theory, we describe the classical world around us using the Newton's classical mechanics and Maxwell's electromagnetic theory. Both theories are deterministic in the sense that the state of the system now uniquely and completely determines all phenomena about the system in the future. There is the classical chaos, but such random behavior is due to the extreme sensitivity of the subsequent dynamics on the initial conditions and the fact that in practice, we do not have such precise control on these initial conditions. In principle, if we do have such precise control and know the complete knowledge of the state, then everything about the system in the future is completely determined.

However, the development of quantum theory brings challenges to such a point of view: a fundamental feature of quantum theory is that it is probabilistic. The complete knowledge of the state of the system now does not enable us to predict the outcomes of all possible measurements that could be performed on the system. In quantum theory, events are randomly realized and this randomness is an intrinsic feature. One can only predict the probabilities that certain events will be observed if a measurement is done on a system prepared in a certain way. Quantum theory is the mathematical framework that enables us to calculate these probabilities. Born's rule is central in the framework as it provides us with the link between the phenomena observed and the formalism of the quantum theory.

That quantum theory cannot enable us to predict the outcomes of all possible measurements does not imply that quantum theory is ill-defined, or incomplete.

It is not possible to add any further elements to the theory so as to make such predictions. There are various attempts at modifying the theory so as to reinstall determinism in the theory. However, all such attempts cannot make consistent modifications without getting wrong predictions in other situations. The violation of Bell's inequality observed in the laboratory tells us conclusively that, whether you like it or not, randomness is intrinsic in quantum mechanics. To add to that, up to now, there has not been even a single experimental fact that contradicts a quantum-theoretical prediction. Whether it is unsound to one's philosophical ideas of the world or just contrary to one's liking, we have to accept that quantum theory is probabilistic in nature and try to live with such randomness.

Besides the intrinsic randomness, quantum systems also show other non-classical features that do not have analogs in a classical system. Some examples include superposition and entanglement. All these strange features, however, could be utilized in some way to do useful things for us. For example, intrinsic randomness is exploited in cryptographic schemes to make absolute security possible. Superposition is found to be useful in the so-called Deutsch algorithm and other quantum computation tasks. Entanglement plays a crucial role in quantum information protocols such as teleportation. In all these examples, the quantum system is manipulated to perform certain information tasks or computations. In a nutshell, quantum information and quantum computation are about finding ways of utilizing the quantum system in theory, and about gaining better and more precise control on them in the experiments, to perform useful information tasks and computations for us.

In all of these quantum information tasks and computations, quantum state preparation is the first important step for any protocol that makes use of a source of a quantum system. In a typical scenario, a source emits a certain desired state which carries the information, or is needed for the computational task. For instance, a quantum-state teleportation protocol that is carried out using optical equipment requires a source that produces two photons that are prepared in a maximally-entangled quantum state. Ideally, these preparations should be accurate and their implementations should not be too cumbersome. A variety of states should be generated in a controllable manner without the need to consume too much time or

involve too complex a setup with a huge number of pieces of equipment.

In order to verify that the source is indeed producing the desired state, or something close to it, one carries out quantum state tomography on the source. Quantum state tomography is about reconstructing the input state given measurement data collected about it. It is needed for the verification and identification of the state emitted by the source. As we will see later, quantum measurement is not a trivial problem as it is not possible to extract all information needed to reconstruct the state by measuring only a single copy of the system. In general, we need to send many independently and identically prepared copies of such system to a measurement apparatus. The measurement could in general result in different possible outcomes, which are monitored by different detectors placed at appropriate output ports. By counting the relative frequencies of each of the detector clicks, one can then infer the input state using some data processing protocols.

It turns out that most of these desired states possess entanglement, a key resource in many quantum information processing tasks. Hence entanglement verification and detection is also of critical importance in quantum information science and quantum computation. Entanglement witnesses have been introduced such that if we know the state, then we can choose a suitable witness to detect its entanglement. However, given an unknown generic state, one can only randomly select a witness which may or may not detect its entanglement (if there is any in the state). One then has to keep trying new ones until one succeeds. In the case if the state does not possess any entanglement, one will never come to any conclusion. However, if we choose entanglement witnesses which enable quantum state tomography, then measuring a finite number of them will help one reconstruct the input state and then determine whether there is entanglement in the state or not. Hence, in a certain sense, quantum state tomography is a central and recurring theme in quantum information science and quantum computation.

In a realistic quantum state tomography experiment, not only the quantum state to be reconstructed is unknown, but also some additional parameters in the experimental setup, for example, the efficiency of the detectors, the dark counts of the detectors, etc. Furthermore, quantum systems tend to interact with the environ-

ment that is surrounding them, and then decohere. The assumption that we have a closed quantum system is just an approximation, though often a good one. Such interaction generally develops correlations between the system and environment, often in the form of entanglement, which have implications on the subsequent dynamics of the system. This interaction is however often ignored when one describes such quantum state tomography experiments.

It is then the aim of this thesis to study some of the issues raised above. In the first part of the thesis, we focus on state preparation and the efficient detection of entanglement, presenting two experiments that have been performed. The second part of the thesis aims at tackling two theoretical aspects: the issue of additional unknown parameters and that of coupling between the system and environment during quantum state tomography. Below is a more detailed outline of this thesis.

In Chapter 2, we present a short review of quantum mechanics and basic ideas in quantum state tomography that are needed to follow this thesis. For the short review on quantum mechanics, the polarization of light is used as the example and most of the treatment follows very closely to that given in *Lectures on quantum mechanics: basic matters* by Englert [Eng06]. Very often, I remark on things that will strike the experienced reader as rather elementary. This is because over the years, I realized that more and more younger researchers like high school students are also entering this research field and, for them, very little suitable material is there to help them learn and get familiarized with the topics. I find the book by Englert especially to my liking and these elementary remarks in this thesis are meant for the high school students who can get introduced to these basic notions and ideas. For the overview on quantum state tomography, both point estimators and region estimators are briefly introduced that will be used subsequently in this thesis.

Chapter 3 deals with the issue of state preparation. We focus in this chapter on mixed states in particular. They are useful in investigations of quantum computing, studies of the quantum-classical interface, and decoherence channels. We report a controllable method for producing mixed two-photon states via spontaneous parametric down-conversion with a two-type-I crystal geometry. By using variable polarization rotators (VPRs), one obtains mixed states of various purities and de-

grees of entanglement depending on the parameters of the VPRs. The method can be easily implemented for various experiments that require the generation of states with controllable degrees of entanglement or mixedness.

As an application of the source described in Chapter 3, we discuss an experiment done on entanglement witnesses in Chapter 4. Besides testing the utility and robustness of the source introduced in Chapter 3, this experiment also has its fundamental importance on its own as it involves fast and efficient detection of entanglement. In this experiment, one determines, with least tomographic effort, whether an unknown two-photon polarization state is entangled or separable. The method measures whole families of optimal entanglement witnesses. We further introduce adaptive measurement schemes that greatly speed up the entanglement detection.

As we were performing the experiments discussed in Chapter 3 and Chapter 4, we realized that in a typical experiment such as the ones we performed, additional parameters, apart from the state, are also unknown, for example the efficiency of the detectors. However, for simplicity, the majority of the quantum state tomography experiments performed so far assume that the quantum state to be estimated is the only unknown, while other parameters necessary to reconstruct the state are all perfectly known, normally as a result of some form of pre-calibration. However, such pre-calibration is not always feasible. In this chapter, we study quantum state tomography with additional unknown parameters and illustrate the construction of optimal error regions with some examples.

Finally in Chapter 6, we turn our attention to the inevitable interaction between the system and the environment that is omnipresent. Such systems are called open quantum systems. The initial state of a system-environment composite is needed as the input for predictions from any quantum evolution equation, which describes the effects of noise on the system from joint evolution of the system-environment interaction dynamics. The conventional wisdom is to simply write down an uncorrelated state as if the system and environment were prepared in the absence of each other; or one pleads ignorance and writes down a symbolic system-environment state, allowing for possible arbitrary correlations—quantum or classical—between

the system and the environment. Here, we show how one uses ideas from quantum state tomography to deduce a reasonable and consistent initial system-environment state. In typical situations, such a state turns out to be uncorrelated or almost uncorrelated between the system and the environment. This has implications, in particular, on the subject of subsequent non-Markovian or non-completely-positive dynamics of the system, where the non-complete-positivity stems from initial non-trivial correlations between the system and the environment.

We close the thesis with a short conclusion and outlook in Chapter 7.

Chapter 2

Background

In this chapter, we will briefly review some of the elementary concepts that are useful to follow this thesis. In particular, we will first give a short review of quantum mechanics, using the polarization of light as the example, followed by a brief overview on quantum state tomography.

2.1 Quantum mechanics: A brief review

2.1.1 Basic concepts: Events and states

Before one is exposed to quantum mechanics, one's first encounter with physics usually starts with Newton's classical mechanics [New87], which deals with the motion of massive bodies, or simply masses, (that is, their positions, $\boldsymbol{r}_i(t)$, and velocities, $\boldsymbol{v}_i(t)$, where the subscript i refers to the ith mass) under the influence of forces. The equation of motion (also known as the Newton's second law)

$$m_i \frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{v}_i(t) = \boldsymbol{F}_i(t), \tag{2.1}$$

where m_i is the mass of the ith body and $\boldsymbol{F}_i(t)$ is the total force acting on it at time t, governs its motion. Note that, however, Newton's equation of motion is built upon the concepts of masses and forces, which are preexisting, and does not answer why there are masses and forces in the first place. Similarly, in Maxwell's electromagnetic theory [Max73], electric charge is a preexisting concept, and the theory deals with electromagnetic forces exerted on charges, and in turn, how these

(moving) charges modify the electromagnetic fields. Both Newtonian mechanics and Maxwell's electromagetism are deterministic theory. See, for example, Eq. (2.1), where the solution to these differential equations with the given initial conditions $\boldsymbol{r}_i(t=0)$ and $\boldsymbol{v}_i(t=0)$ uniquely determine the subsequent trajectories of the masses.

In quantum mechanics, one deals with the behaviour of atomic systems, and in particular, with the results of measurements on them [Sch01]. Similar to Newtonian classical mechanics, there are also preexisting concepts in quantum mechanics, one of which is that of an *event* [Eng13, Haa90]. Some examples of an event include: the emission of a photon by an atom in its excited state; the landing of a silver atom on a screen; the absorption of a photon by a semiconductor detector, etc. However, contrary to classical mechanics which are deterministic, quantum theory is an intrinsically probabilistic theory. Take for example, the famous Stern-Gerlach experiment [GS22]: We can only predict the percentage of silver atoms landing on the upper (lower) part of screen. But for each individual silver atom, we do not know, or rather, it is unknownable, where it will land. This is because when one deals with atomic measurements, first of all, atomicity means that the microscopic entities have many of their properties carried in certain basic units. Put it simply, there is an electron which cannot be halved. There is no half a unit of charge. As a result of this, we cannot make the electric interaction as arbitrarily small as we like. Secondly, we are also unable to compensate for the disturbance caused during the measurement due to the interactions in the realm of atomic measurement, as we cannot predict in detail what each individual event will do, but only on a statistical level [Sch01]. Put it simply, there is no mechanism that decides the outcome of a quantum measurement [Eng06]. What one can tell is only the probabilities for the occurrence of the various possible outcomes. Quantum mechanics does this job exactly, enabling us to correctly calculate these probabilities.

Let us illustrate these ideas using a simple example: the polarization of the light [Hec01]. As we know, light is a transverse electromagnetic wave, consisting of electric field and magnetic field oscillating in space and time. The direction of the oscillations in space of the electric field defines its polarization axis. A beam of light, travelling along the z-axis, could have its plane of oscillation horizontal (along

x-axis),

$$\boldsymbol{E}(z,t) = \boldsymbol{e}_x E\cos(kz-\omega t) \widehat{=} E(t)\begin{pmatrix}1\\0\end{pmatrix}, \tag{2.2}$$

which we call horizontally polarized light, or vertical (along y-axis),

$$\boldsymbol{E}(z,t) = \boldsymbol{e}_y E\cos(kz-\omega t) \widehat{=} E(t)\begin{pmatrix}0\\1\end{pmatrix}, \tag{2.3}$$

which we call vertically polarized light. A device that could sort out the polarization is the so-called polarizing beam splitter (PBS). When a horizontally polarized light is sent to a PBS, it will be transmitted, whereas being reflected for a vertically polarized light. But the electric field could also be oscillating at an angle, say $\pm 45^\circ$ with respect to the horizontal, that is

$$\boldsymbol{E}(z,t) = \frac{1}{\sqrt{2}}(\boldsymbol{e}_x \pm \boldsymbol{e}_y) E\cos(kz-\omega t) \widehat{=} E(t)\frac{1}{\sqrt{2}}\begin{pmatrix}1\\\pm 1\end{pmatrix}. \tag{2.4}$$

We call such light diagonally (for 45°) polarized or anti-diagonally (for -45°) polarized. What happens when diagonal or anti-diagonal polarized light hits the PBS? If we now put two detectors on the two ports of the PBS, we see that half of the intensity is transmitted and half reflected. Note that we can use the so-called Jones vector as a compact notation to represent the state of polarization of the light, so that an arbitrary polarization is given by

$$\boldsymbol{E}(z,t) = \boldsymbol{e}_x E_x(t) + \boldsymbol{e}_y E_y(t) \widehat{=} \begin{pmatrix}E_x(t)\\E_y(t)\end{pmatrix}, \tag{2.5}$$

where $E_x(t)$ and $E_y(t)$ are the *amplitude* of the electric field along the x and y-axis respectively. Note that the electric field is real, but the Jones vector represents the (relative) amplitude and the (relative) phase of the electric field in x and y directions. Hence, the amplitude could be complex here so as to include circularly polarized light (or more generally elliptically polarized light). Further, also note that by taking the modulus square of the electric field amplitude, one gets the correct *intensity*, observed in the respective detectors in the transmitted arm or reflected

arm of the PBS.

Now what happens if we keep dimming the light source until there is only a single photon in each pulse?[1] Then we notice that at any time, since a photon cannot be further split, only one of the detectors will click. But it is completely random and unpredictable which one will click for the next incoming photon.[2] When one carries out this experiment, a typical sequence that one would get looks like this: HHVHVVVHVH.... If one waits long enough, then on average, one gets half of the photons transmitted, and half reflected. But for each individual photon, all we can say is the *probability* that it will be transmitted or reflected. Following Dirac, we will write a ket $|\mathrm{H}\rangle$ to symbolize a photon with horizontal polarization, and ket $|\mathrm{V}\rangle$ for a photon with vertical polarization, with a vector representation similar to the Jones vector given by

$$|\mathrm{V}\rangle \mathrel{\widehat{=}} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |\mathrm{H}\rangle \mathrel{\widehat{=}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{2.6}$$

with the *completeness* relation given by

$$|\mathrm{V}\rangle\langle\mathrm{V}| + |\mathrm{H}\rangle\langle\mathrm{H}| \mathrel{\widehat{=}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} \tag{2.7}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathrel{\widehat{=}} \mathbb{1}, \tag{2.8}$$

where $\mathbb{1}$ is the two-dimensional identity operator.

Any other polarization state can be written as a linear combination of these two kets. In general, our description of the photon is symbolized by the *state vector*, usually denoted as $|\ \rangle$ (or $\langle\ |$), called the Dirac's *ket* (or *bra*). That is, we have

$$|\ \rangle = |\mathrm{V}\rangle\alpha + |\mathrm{H}\rangle\beta, \tag{2.9}$$

where α, and β are the *probability amplitudes*. The bra is the complex transposition

[1] The concept of a photon is more subtle than that, see, for example, Ref. [Lou00]. But for the purpose of the current discussion, let us take a photon to mean the smallest packet of energy that will trigger only a single click from the detector.

[2] So far, I have only tried to make it sound plausible that the randomness is intrinsic without proof. For the proof, see many standard quantum mechanics textbooks on *Bell's inequality*.

(denoted as †) of the ket, that is $\langle\ |\ =\ |\ \rangle^\dagger$, which is to be represented by a row vector. The complex transposition is to take the transpose of the matrix (denoted as T), followed by complex conjugation (denoted as $*$), that is,

$$M^\dagger = \left(M^{\mathrm{T}}\right)^*, \tag{2.10}$$

for a matrix M.

We also generalize the dot product between two vectors in real space,

$$\boldsymbol{e}_x \cdot \boldsymbol{e}_y = 0, \tag{2.11}$$

to the inner product between the bra and the ket, which is also called a *bracket*, given by

$$\langle \mathrm{H}|\mathrm{V}\rangle = 0, \tag{2.12}$$

expressing the *orthogonality* between these two kets.

Just like taking the modulus square of the electric field amplitude gives us the intensity, taking the modules square of the probability amplitudes gives us the correct probabilities that the photon will be detected at the H-port, or V-port of the detectors. Since the probabilities must add up to one, we need $|\alpha|^2 + |\beta|^2 = 1$, the so-called *normalization* condition.

Some terminology is in order. We say that the set $\{|\mathrm{V}\rangle, |\mathrm{H}\rangle\}$ forms an orthogonal basis for the case of photon polarization, since any polarization of the photon can be written as a linear combination of the two basis kets which are orthogonal to each other. If all the basis kets are normalized, then it is an *orthonormal* basis. More generally, a more complex quantum system may need the specification of d such basis kets, say $\{|k\rangle\}_{k=1,2,\cdots,d}$, for the ket space of dimension d, with the completeness and orthonormality properties given by

$$\sum_{k=1}^{d} |k\rangle\langle k| = \mathbb{1}, \tag{2.13}$$

and

$$\langle k|j\rangle = \delta_{kj}, \quad k, j = 1, 2, \cdots, d, \tag{2.14}$$

where $\mathbb{1}$ is the d-dimensional identity operator.

Such orthonormal basis is not unique and in fact there are infinitely many of them. For example, besides $\{|\mathrm{V}\rangle, |\mathrm{H}\rangle\}$, diagonal and anti-diagonal polarization $\{|\mathrm{D}\rangle, |\mathrm{A}\rangle\}$ or left- and right-circular polarization $\{|\mathrm{L}\rangle, |\mathrm{R}\rangle\}$ also form a basis.

The sets of kets and bras form Hilbert spaces of dimension d that are dual to each other. Introducing an orthonormal basis, one can expand any arbitrary ket in this basis by

$$|\ \rangle = \mathbb{1}|\ \rangle = \sum_{k=1}^{d} |k\rangle\langle k|\ \rangle = \sum_{k=1}^{d} \langle k|\ \rangle |k\rangle, \tag{2.15}$$

where the inner product $\langle k|\ \rangle$ represents the overlap of $|\ \rangle$ with the basis ket $|k\rangle$. In the example of photon polarization, the dimension of the Hilbert space is two, since any polarization can be decomposed into linear combination of two independent ones. Such a system of two-dimensional Hilbert space is called a *qubit*, and the most popular orthonormal basis is perhaps $\{|\mathrm{V}\rangle, |\mathrm{H}\rangle\}$.

2.1.2 Measurement: Born's rule

Continuing the discussion of polarization of the photon, suppose that we send photons with the polarization state $|\ \rangle = |\mathrm{V}\rangle\alpha + |\mathrm{H}\rangle\beta$ to a PBS, then the probability that the next photon is detected at the V-port of the PBS is given by $|\alpha|^2$, that is

$$p(\mathrm{V}) = |\alpha|^2 = |\langle \mathrm{V}|\ \rangle|^2 = \langle \mathrm{V}|\ \rangle\langle\ |\mathrm{V}\rangle. \tag{2.16}$$

We now introduce the so-called *trace* operation, defined by

$$\mathrm{tr}\{|a\rangle\langle b|\} = \langle b|a\rangle, \tag{2.17}$$

with the *linearity* property given by

$$\mathrm{tr}\{A + B\} = \mathrm{tr}\{A\} + \mathrm{tr}\{B\}, \tag{2.18}$$

and

$$\mathrm{tr}\{cA\} = c \times \mathrm{tr}\{A\}, \tag{2.19}$$

for all operators A and B, and all scalars c.

This enables us to re-write Eq. (2.16) as

$$p(\mathrm{V}) = \mathrm{tr}\{\rho\Pi_{\mathrm{V}}\}, \tag{2.20}$$

where

$$\rho = |\ \rangle\langle\ |, \tag{2.21}$$

known as the *statistical operator*, which summarizes our knowledge about the state, and

$$\Pi_{\mathrm{V}} = |\mathrm{V}\rangle\langle\mathrm{V}|, \tag{2.22}$$

which summarizes our knowledge about the measurement. Similarly, we have

$$p(\mathrm{H}) = \mathrm{tr}\{\rho\Pi_{\mathrm{H}}\}, \tag{2.23}$$

with

$$\Pi_{\mathrm{H}} = |\mathrm{H}\rangle\langle\mathrm{H}|, \tag{2.24}$$

for the probability that the next photon is detected at the H-port. We call such a measurement projective measurement as it projects the photons onto the V/H basis. In general, we have

$$p_i = \mathrm{tr}\{\rho\Pi_i\}, \tag{2.25}$$

known as the *Born's rule*. The operators Π_is are known as the *probability operators*.

In the case considered above, the probability operators belong to a class of special operators known as projectors. However, to be general, we allow for non-projective measurement, or generalized measurement [NC10]. Regardless of the details of the measurement, such as the exact physical nature of the measurement, or the state of the system after the measurement, a consistent measurement theory must satisfy the following two criteria. Firstly, the probability of occurrence, p_i, for the outcome of probability operator Π_i must be non-negative. Secondly, the probabilities for all outcomes must sum to unity, assuming no losses.[3] Hence, we call a set of positive

[3]The effect of losses can be accounted for by introducing one more probability operator in the POM.

operators $\{\Pi_i\}$ the *probability operator measurement* (POM), if they satisfy the *positivity* property

$$\Pi_i \geq 0, \quad \text{for all } i, \tag{2.26}$$

and the completeness property

$$\sum_i \Pi_i = \mathbb{1}. \tag{2.27}$$

A particular probability operator Π_i is also called one element of the POM. The POM element can only enable us to calculate the respective probabilities that one is interested in. However, in certain cases, one also would like to know the state of the system after the measurement. For that purpose, one needs to know the details of the implementation of the measurements, so as to construct the measurement operator, also known as the *Kraus operator* [Kra83],

$$\Pi_i = K_i^\dagger K_i. \tag{2.28}$$

The state of the system after the measurement is given by

$$|\ \rangle \to \frac{K_i|\ \rangle}{\langle\ |\Pi_i|\ \rangle}. \tag{2.29}$$

For the situation considered above, we have

$$K_\mathrm{H} = |\mathrm{VAC}\rangle\langle\mathrm{H}|, \quad K_\mathrm{V} = |\mathrm{VAC}\rangle\langle\mathrm{V}|, \tag{2.30}$$

where $|\mathrm{VAC}\rangle$ symbolizes the vacuum state, since the photon is absorbed after the measurement. Depending on which of the detector clicked, the statistical operator is then updated according to

$$\rho \to \frac{K_i \rho K_i^\dagger}{p_i}. \tag{2.31}$$

As a final remark, notice that in quantum mechanics, a measurement of a certain property could be realized in many different bases. For example, for the measurement of the polarization of light, by using optical devices which modifies the polarization such as half-wave plate (HWP) or quarter-wave plate (QWP), one could do a projective measurement in the D/A basis or L/R basis. A photon which is

surely transmitted at the H-port of the PBS will be completely unpredictable when it is measured in the D/A or L/R basis. This effect, namely, if precise knowledge of one of the observables implies that all possible outcomes of measuring the other are equally probable, is known as *complementarity* [SEW91].

2.1.3 Mixed state: Purity

What we described so far is the concept of a *pure state*. That is, the light has a certain pure and well-defined polarization state. It is a state of the photon that there exists a chosen basis, such that if measurement is done in this basis, then we know for sure that only one of the detectors would click. One can also think of it as someone has prepared all the photons in that particular state. How do we then describe a preparation such that a certain fraction of the photons are prepared in a particular state, and others in a different state, say half of the photons in H-polarized state and half in V-polarized state? To describe such a state, let us imagine that it is sent to a measurement apparatus realizing a particular POM, $\{\Pi_i\}$. The measurement result on the fraction of photons in the H-polarized state is predicted by $\mathrm{tr}\{|\mathrm{H}\rangle\langle\mathrm{H}|\Pi_i\}$, whereas that on the fraction of photons in the V-polarized state is predicted by $\mathrm{tr}\{|\mathrm{V}\rangle\langle\mathrm{V}|\Pi_i\}$. Since we have half of each of them, the overall result is predicted by using

$$\begin{aligned} p_i &= \frac{1}{2}\mathrm{tr}\{|\mathrm{H}\rangle\langle\mathrm{H}|\Pi_i\} + \frac{1}{2}\mathrm{tr}\{|\mathrm{V}\rangle\langle\mathrm{V}|\Pi_i\} \\ &= \mathrm{tr}\left\{\left(\frac{1}{2}|\mathrm{H}\rangle\langle\mathrm{H}| + \frac{1}{2}|\mathrm{V}\rangle\langle\mathrm{V}|\right)\Pi_i\right\}. \end{aligned} \tag{2.32}$$

It then follows that the appropriate description of such a mixed state is given by the statistical operator

$$\rho = \frac{1}{2}|\mathrm{H}\rangle\langle\mathrm{H}| + \frac{1}{2}|\mathrm{V}\rangle\langle\mathrm{V}|. \tag{2.33}$$

More generally, the generic form of statistical operator is

$$\rho = \sum_i |i\rangle g_i \langle i|, \tag{2.34}$$

where $|i\rangle$ is the ket of the ith type system, and g_i is its weight in the mixture. The

statistical operator, also known as the density operator, density matrix, or simply, the "state", is the generalization of the ket or bra. It is *hermitian* by construction, satisfying the hermiticity property that $\rho^\dagger = \rho$. In Eq. (2.34), one can interpret the state *as if* the state is made up of mixtures of $|i\rangle\langle i|$, each with the weight g_i. Note that however, this decomposition is generally not unique: There are infinitely many *as-if realities* for a given state, with pure states being the exception. Following the standard terminology, we have a unique *mixture* made up of different *blends*. An arbitrary *ensemble* of systems could consist of purely identical copies of a quantum system, or mixtures of different ones. Then, the first kind of ensemble is known as *pure states*, and the second kind as *mixed states*.

Two remarks are in order: First, we refer to the dimension D of the *state space* (the space of all statistical operators) as the number of entries of the *density matrix*, i.e. the representation of statistical operator in any basis. For example, bipartite qubit systems have a D =16-dimensional state space. We have $D = d^2$, where d is the dimension of the Hilbert space introduced previously. Secondly, we call states with n non-zero eigenvalues as *rank-n* states. For instance, pure states are rank-one states, mixtures of two orthogonal pure states are rank-two states.

To quantify the amount of mixedness in a given state ρ, a simple measure is given by its *purity* [NC10],

$$P = \mathrm{tr}\left\{\rho^2\right\}. \tag{2.35}$$

The values of the purity are bounded between $1/d$ and one. It equals to one if and only if ρ is pure, and equals to $1/d$ when ρ is *completely mixed*, i.e. $\rho = \mathbb{1}/d$. Obviously, the completely mixed state is a rank-d state, or a full rank state. States which are not full rank are rank-deficient.

As an example, let us consider a qubit. One particularly convenient way to parametrize a qubit is to write

$$\rho = \frac{1}{2}\left(\mathbb{1} + \boldsymbol{s}\cdot\boldsymbol{\sigma}\right), \tag{2.36}$$

where $\boldsymbol{\sigma} = \{\sigma_x, \sigma_y, \sigma_z\}$ are the familiar *Pauli matrices*, sometimes denoted as $\{X, Y, Z\}$, and $|\boldsymbol{s}| \leq 1$. The geometry of the qubit state space could be visual-

ized as the *Bloch ball*, with pure states on the surface with $|\boldsymbol{s}| = 1$, and mixed states inside with $|\boldsymbol{s}| < 1$ [NC10].

2.1.4 Bipartite system: Entanglement

For a system composed of two subsystems, called a *bipartite system*, the extension of symbols for states is straightforward, with the tensor product $\otimes$ between the two subsystems' states. For instance, $|\mathrm{H}\rangle_\mathrm{A} \otimes |\mathrm{H}\rangle_\mathrm{B}$, or simply $|\mathrm{HH}\rangle$, identifies a pair of horizontally polarized photons from subsystems A and B. Just like $\{|\mathrm{V}\rangle, |\mathrm{H}\rangle\}$ forms a basis for a single system, we have $\{|\mathrm{VV}\rangle, |\mathrm{VH}\rangle, |\mathrm{HV}\rangle, |\mathrm{HH}\rangle\}$ for a basis for the bipartite system, with straightforward meanings. An arbitrary bipartite (pure) state is then given by

$$|\ \rangle = \alpha|\mathrm{VV}\rangle + \beta|\mathrm{VH}\rangle + \gamma|\mathrm{HV}\rangle + \delta|\mathrm{HH}\rangle, \tag{2.37}$$

with $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$.

For example, we can have (with $\alpha = \beta = \sqrt{\frac{1}{2}}$, and $\gamma = \delta = 0$)

$$\begin{aligned} |\ \rangle &= \sqrt{\frac{1}{2}}|\mathrm{VV}\rangle + \sqrt{\frac{1}{2}}|\mathrm{VH}\rangle \\ &= |\mathrm{V}\rangle \otimes \left(\sqrt{\frac{1}{2}}|\mathrm{V}\rangle + \sqrt{\frac{1}{2}}|\mathrm{H}\rangle\right), \end{aligned} \tag{2.38}$$

describing a bipartite system with subsystem A in the vertical polarization state, and subsystem B in the diagonal polarization state. If we now wish to make reference to one of the subsystem only, we obtain the *reduced statistical operator* of one of the subsystem by taking the *partial trace* over the other subsystem,[4] that is

$$\rho_\mathrm{A} = \mathrm{tr}_\mathrm{B}\{\rho_\mathrm{AB}\}. \tag{2.39}$$

The partial trace tr_B is defined by

$$\mathrm{tr}_\mathrm{B}\{|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|\} = |a_1\rangle\langle a_2|\mathrm{tr}\{|b_1\rangle\langle b_2|\}, \tag{2.40}$$

where $|a_1\rangle$ and $|a_2\rangle$ are any two vectors in the state space of subsystem A, and $|b_1\rangle$

[4] Why the partial trace? See, for example, Ref. [NC10]

and $|b_2\rangle$ are any two vectors in the state space of subsystem B. For the state given by Eq. (2.38), we have $\rho_{\mathrm{A}} = \mathrm{tr}_{\mathrm{B}}\{\rho_{\mathrm{AB}}\} = |\mathrm{V}\rangle\langle\mathrm{V}|$, and $\rho_{\mathrm{B}} = \mathrm{tr}_{\mathrm{A}}\{\rho_{\mathrm{AB}}\} = |\mathrm{D}\rangle\langle\mathrm{D}|$, as one would expect.

The state given by Eq. (2.38) is an example of a tensor product structured state, called a *product state*. How about the state (with $\beta = \sqrt{\frac{1}{2}}$, $\gamma = -\sqrt{\frac{1}{2}}$, and $\alpha = \delta = 0$)

$$|\Psi_-\rangle = \frac{1}{\sqrt{2}}(|\mathrm{VH}\rangle - |\mathrm{HV}\rangle), \tag{2.41}$$

known as the *singlet state*? It turns out that this state (and many others) cannot be written as a tensor product of two subsystem states, that is, there exists *no* values of α_1, β_1, α_2, and β_2 such that

$$\frac{1}{\sqrt{2}}(|\mathrm{VH}\rangle - |\mathrm{HV}\rangle) = (\alpha_1|\mathrm{V}\rangle + \beta_1|\mathrm{H}\rangle) \otimes (\alpha_2|\mathrm{V}\rangle + \beta_2|\mathrm{H}\rangle)\,. \tag{2.42}$$

We call such a state *entangled*, as it seems to be impossible to make reference to only one subsystem without making reference to the other. This non-classical state possesses quantum *entanglement*.

The meaning of the entanglement is revealed further by looking at measurement results done on this bipartite system. One shows that

$$p(\mathrm{HH}) = \mathrm{tr}\left\{|\Psi_-\rangle\langle\Psi_-|\,(\Pi_{\mathrm{H}} \otimes \Pi_{\mathrm{H}})\right\} = 0, \tag{2.43}$$

where the tensor product of operators, $\Pi_{\mathrm{H}} \otimes \Pi_{\mathrm{H}}$, similarly describes the probability operator that subsystem A is projected into H-polarized state and subsystem B is also projected into H-polarized state. Similarly, one shows that

$$p(\mathrm{HV}) = \frac{1}{2}, \;\; p(\mathrm{VH}) = \frac{1}{2}, \;\; \text{and} \;\; p(\mathrm{VV}) = 0. \tag{2.44}$$

This means that when this bipartite system is measured in the H/V basis, half of the time, subsystem A will be detected at the H-port and the other half at the V-port, completely random. One could also see this by looking at the reduced statistical

operator of subsystem A, which is

$$\rho_{\mathrm{A}} = \mathrm{tr}_{\mathrm{B}} \left\{ |\Psi^-\rangle\langle\Psi^-| \right\} = \frac{\mathbb{1}}{2}, \tag{2.45}$$

the completely mixed state. But whenever subsystem A is detected at H-port, subsystem B is always detected at V-port and vice versa, meaning that there exists *correlations* between the measurement results on these two subsystems.

At this point, it is important to distinguish the state given by Eq. (2.41) with that given by

$$\rho = \frac{1}{2}|\mathrm{HV}\rangle\langle\mathrm{HV}| + \frac{1}{2}|\mathrm{VH}\rangle\langle\mathrm{VH}|, \tag{2.46}$$

as the latter state also results in the same correlation. However, such correlation are classical in nature as one can prepare the state given by Eq. (2.46) at separate locations by classical communications. That is, imagine that Alice at lab A prepares the state $|\mathrm{V}\rangle$ half of the time, and $|\mathrm{H}\rangle$ half of the time. But whenever she prepares $|\mathrm{H}\rangle$, she calls Bob at lab B to prepare $|\mathrm{V}\rangle$, and vice versa. This form of classical correlation will disappear if the joint state that they prepare is measured in a different basis. On the contrary, the quantum correlation possessed by the state in Eq. (2.41) persists even if measurement was done in any other orthogonal basis and the state can only be prepared if subsystem A and B were to interact some time at the same location. This kind of intimate *quantum correlation* between two subsystems, which is non-classical in nature, is known as *entanglement*. We obtain less information about the individual subsystem, but gain in knowledge about them as a pair.

More generally, for a k-partite system, one calls the ket entangled if it is a superposition ket, which cannot be factorized completely into tensor products of kets of all individual subsystems. Then, the generalization of entanglement to mixed states are as follows. States which can be written as

$$\rho = \sum_i g_i \rho_{1,i} \otimes \rho_{2,i} \otimes \cdots \otimes \rho_{k,i} \tag{2.47}$$

are known as *separable*, whereas states which are not separable are *entangled*.

Physically speaking, a separable state describes a mixture of systems which can

be obtained by preparing all individual subsystems of the k-partite separately and in well-defined states. In other words, all the subsystems retain their individuality, and have no quantum correlations with each other. The set of all separable states is a *convex set*. That is, mixtures of separable states will remain separable, which is obvious. However, the set of entangled states are not convex. A simple counter-example is an equal mixture of $|\Psi_+\rangle = \frac{1}{\sqrt{2}}(|\text{VH}\rangle + |\text{HV}\rangle)$ with $|\Psi_-\rangle$, which results in $\rho = |\text{VH}\rangle\frac{1}{2}\langle\text{VH}| + |\text{HV}\rangle\frac{1}{2}\langle\text{HV}|$. This clearly is a separable state.

Entanglement is one active topic of current research, where the quantum correlation between subsystems is at the heart of many interesting fields, like quantum key distribution [Eke91], quantum teleportation [BPM$^+$97], and demonstration of various concepts of quantum mechanics [AGR81, CS78]. While it is the our aim of this thesis to study various applications and usefulness of quantum entanglement, we are concerned with how to efficiently verify whether an unknown state is entangled or separable. For this purpose, we introduce the idea of an entanglement witness.

Due to the special closure or convex property of the separable states described above, there must exist [HHH96, Ter00] a hermitian operator W, called the *entanglement witness* or *witness operator*, such that if the state ρ_0 is entangled, then

$$\begin{cases} \text{tr}\{\rho W\} \leq \mu \text{ for all separable states,} \\ \text{tr}\{\rho W\} > \mu \text{ for some entangled states, including } \rho_0. \end{cases} \tag{2.48}$$

μ is called the *threshold* of the witness operator, and when the second inequality above is satisfied, one says that this witness operator *detects* the state ρ. As a concrete example, for a bipartite qubit system, $W = |\Psi_+\rangle\langle\Psi_+|$ is a witness with threshold value $\mu = 1/2$. The concept of a witness was first used by M., P., and R. Horodecki [HHH96], and the term "witness" was introduced by Terhal [Ter00]; for reviews that cover all important aspects of entanglement witnesses, see Refs. [GT09, HHHH09].

Here, we mention briefly one particular important kind of witness operator known as an *optimal witness* W_{Opt}: No other witnesses can detect *all* the entangled states detected by W_{Opt}, *plus* some other states [LKCH00]. Geometrically, a witness defines a hyperplane in the state space, which separates a partial set of entangled

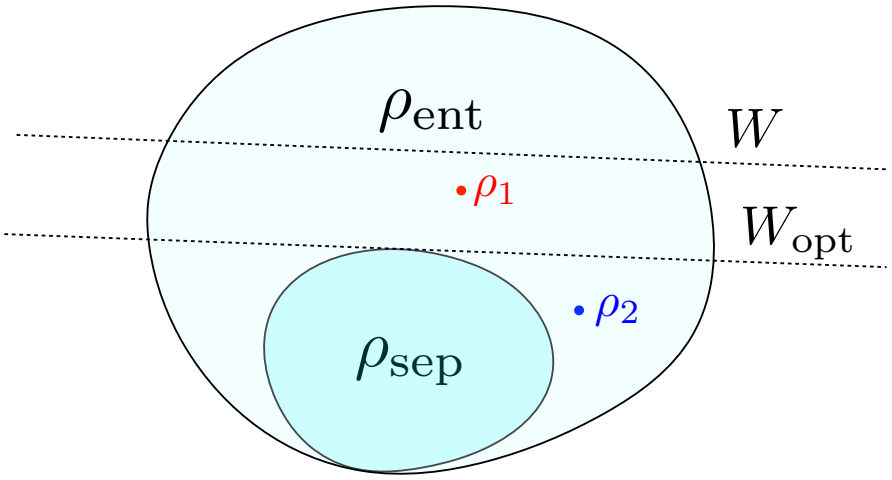

Figure 2.1: Geometry of states and entanglement witnesses: The set of separable state ρ_{sep} is convex, whereas the set of entangled state ρ_{ent} is not. An entanglement witness W defines a hyperplane in the state space which separates the separable states and a partial set of entangled states. An optimal entanglement witness W_{opt} touches the convex set of separable states. In this figure, the state ρ_1 is an entangled state which cannot be detected by W, but can be detected by W_{opt}. The entangled state ρ_2, however, could not be detected by W_{opt}. In order to detect the entanglement in ρ_2, one needs to measure another suitably chosen entanglement witness.

states from all other states. Then, an optimal witness defines a hyperplane in the state space which touches the convex set of separable states, see Fig. 2.1 for a simple illustration.[5] In the figure, the state ρ_1 is an entangled state which cannot be detected by W, but can be detected by W_{opt}. The entangled state ρ_2, however, could not be detected by W_{opt}. In order to detect the entanglement in ρ_2, one needs to measure another suitably chosen entanglement witness.

Alternatively, if the density operator of the state is known, then there are well-known methods to determine whether the state is entangled or separable. For the case of two qubits, we use the Peres-Horodecki criterion [Per96, HHH96]: if the *partial transposition* of a two qubit state is also a valid state, then the original two

[5]The figure shown is for demonstration only: The actual convex state space in its boundary is not as simple as drawn, but generally complex and abstract.

qubit state is separable. Given a general state ρ which acts on $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$,

$$\rho = \sum_{ijkl} p_{ijkl} |i\rangle\langle j| \otimes |k\rangle\langle l|, \tag{2.49}$$

the partial transposition (with respect to subsystem B) is defined as

$$\rho^{\mathrm{T_B}} = \sum_{ijkl} p_{ijkl} |i\rangle\langle j| \otimes |l\rangle\langle k|. \tag{2.50}$$

To understand this criterion, first we note that if a state on $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$ is separable, then it is obvious that its partial transposition is a positive operator, whatever the dimensions of the Hilbert space of A and B. We see this in

$$(\rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}})^{\mathrm{T_B}} = \rho_{\mathrm{A}} \otimes \rho_{\mathrm{B}}^{\mathrm{T}}, \tag{2.51}$$

which is clearly positive, since the transposition does not change the eigenspectrum of the operator ρ_{B}. However, the reverse is true only for a Hilbert space of $\mathcal{C}_2 \otimes \mathcal{C}_2$, and $\mathcal{C}_2 \otimes \mathcal{C}_3$, that is for a two-qubit system, or for a qubit-qutrit system. We will make use of this criterion in the thesis since the experiment discussed later involves a two-qubit state. However, given an unknown state, how to determine the density operator of the state? This is the subject of *quantum state tomography* and we will answer this question in Section 2.2 later.

Just like purity quantifies the amount of mixedness in a given state, we also need to be quantitative about the amount of entanglement possessed by a given state. Indeed, there are a plethora of such measures, stemming from various (geometrical, algebraic, operational) considerations [BŻ06]. For our work, we choose to use the *concurrence* C, and its square, the *tangle* T, which for bipartite qubit, has the advantage of having available analytical expressions, as our basic measures of entanglement [Woo98, CKW00]. The analytical expression for concurrence is

$$C(\rho) = \max\{0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\}, \tag{2.52}$$

where the λ_js are the eigenvalues, in decreasing order, of the Hermitian matrix $R \equiv \sqrt{\sqrt{\rho}\tilde{\rho}\sqrt{\rho}}$. Here, $\tilde{\rho} = \Sigma\rho^{\mathrm{T}}\Sigma$, where ρ and $\Sigma = Y \otimes Y$, Y being the y-Pauli

operator, all are expressed in the *standard basis* $\{|\mathrm{VV}\rangle, |\mathrm{VH}\rangle, |\mathrm{HV}\rangle, |\mathrm{HH}\rangle\}$, i.e.

$$\Sigma = Y \otimes Y = \begin{array}{c} \\ |\mathrm{V}\rangle \\ |\mathrm{H}\rangle \end{array} \begin{array}{c} \langle \mathrm{V}| \quad \langle \mathrm{H}| \\ \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \end{array} \otimes \begin{array}{c} \\ |\mathrm{V}\rangle \\ |\mathrm{H}\rangle \end{array} \begin{array}{c} \langle \mathrm{V}| \quad \langle \mathrm{H}| \\ \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \end{array}$$

$$= \begin{array}{c} \\ |\mathrm{VV}\rangle \\ |\mathrm{VH}\rangle \\ |\mathrm{HV}\rangle \\ |\mathrm{HH}\rangle \end{array} \begin{array}{c} \langle \mathrm{VV}| \quad \langle \mathrm{VH}| \quad \langle \mathrm{HV}| \quad \langle \mathrm{HH}| \\ \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \end{array}. \tag{2.53}$$

For all separable states, their concurrences are zero, and for maximally entangled states, $C = 1$.

Finally, we mention that just like the set $\{|\mathrm{VV}\rangle, |\mathrm{VH}\rangle, |\mathrm{HV}\rangle, |\mathrm{HH}\rangle\}$ forms a basis for a bipartite qubit system, there are also entangled kets which form an orthonormal basis. The often used one is the so-called Bell basis, which consists of the following four *Bell states*

$$|\Psi_-\rangle = \frac{1}{\sqrt{2}}(|\mathrm{VH}\rangle - |\mathrm{HV}\rangle), \tag{2.54}$$

$$|\Psi_+\rangle = \frac{1}{\sqrt{2}}(|\mathrm{VH}\rangle + |\mathrm{HV}\rangle), \tag{2.55}$$

$$|\Phi_-\rangle = \frac{1}{\sqrt{2}}(|\mathrm{VV}\rangle - |\mathrm{HH}\rangle), \tag{2.56}$$

$$|\Phi_+\rangle = \frac{1}{\sqrt{2}}(|\mathrm{VV}\rangle + |\mathrm{HH}\rangle). \tag{2.57}$$

These states are maximally entangled states in the two-qubit state space, having a concurrence equal to one.

2.1.5 Dynamics

So far, we have been concerned with only kinematics of the quantum system (with the exception of the measurement), that is how a quantum system is described. We now turn our attention to the dynamics of the quantum system, that is how a

quantum system evolves in time. We focus on closed quantum system first and look at how the basis bras evolve with time. At time t, we have a complete orthonormal basis $\langle \ldots, t|$, where the ellipses stands for any quantum property of the system (polarization, position, momentum, etc). At time $t+\tau$, we have another set of basis bras $\langle \ldots, t+\tau|$. These two sets must be related to each other by some *unitary* transformation, that is, we have

$$\langle \ldots, t+\tau| = \langle \ldots, t|U_t(\tau), \tag{2.58}$$

with $U_t^\dagger U_t = \mathbb{1}$. It then follows that the dynamics of the quantum system is governed by unitary evolution

$$\rho \to U\rho U^\dagger. \tag{2.59}$$

This is true for a closed quantum system. If the system is open to interact with the environment, there will be flow of information between the system and the environment. The evolution of the system will no longer by described by unitary transformations. However, one can still treat the system plus the environment composite as forming a closed quantum system. The joint system-environment state then evolves according to

$$\rho_{\mathrm{SE}} \to U_{\mathrm{SE}}\rho_{\mathrm{SE}}U_{\mathrm{SE}}^\dagger, \tag{2.60}$$

where the unitary transformation, U_{SE}, acts on the system-environment as a whole. Here, U_{SE} is related to the total Hamiltonian of the system-environment composite $H_{\mathrm{T}} = H_{\mathrm{S}} + H_{\mathrm{E}} + H_{\mathrm{I}}$, where the three terms represent the Hamiltonian of the system, that of the environment, and that of the interaction respectively, see Fig. 2.2. Knowing the system-environment state at $t = 0$, as well as the total unitary $U_{\mathrm{SE}}(t)$, one can then work out the system-environment state at time t. The system-only state is obtained as usual, by taking the partial trace over the environment, so that

$$\rho_{\mathrm{S}}(t) = \mathrm{tr}_{\mathrm{E}}\left\{U_{\mathrm{SE}}(t)\rho_{\mathrm{SE}}(0)U_{\mathrm{SE}}^\dagger(t)\right\}. \tag{2.61}$$

In most of the cases, one assumes that initially the system and environment is

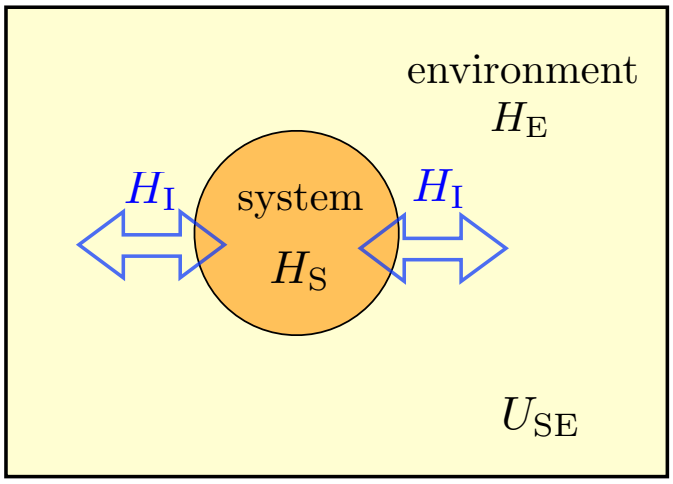

Figure 2.2: Quantum system with Hamiltonian H_{S} is inevitable to interact with the environment with Hamiltonian H_{E}. There will be flow of information between the system and the environment due to the interaction Hamiltonian H_{I}. The evolution of the system will no longer by described by unitary transformations. However, one can still treat the system plus the environment composite as forming a closed quantum system. The evolution of the joint system-environment state is then governed by the total unitary U_{SE}.

uncorrelated and in a product state, that is

$$\rho_{\mathrm{SE}}(0) = \rho_{\mathrm{S}}(0) \otimes \rho_{\mathrm{E}}(0). \tag{2.62}$$

Such an assumption results in master equation descriptions for the system-only state [NC10, BP06].

2.2 Quantum state tomography

2.2.1 Introduction

After the short review on quantum mechanics, we are now prepared for discussions on quantum state tomography. Quantum mechanics enables us to correctly predict the probabilities of various outcomes through the Born's rule, that is, given the state ρ, and the measurement $\{\Pi_i\}$, what are the probabilities $\{p_i\}$. However, in many

cases, one would encounter the reverse problem. How could we infer the unknown state ρ if we know the measurement results?

This seems to be a rather innocent question, however, the answer is completely non-trivial. If one knows the correct probabilities p_is, then the answer is indeed simple and straightforward. One just inverts the Born's rule. However, unless one has an infinite amount of resources, one would never be able to get the exact correct probabilities. This is due to the Heisenberg uncertainty relation [Hei27, Hei30] and the complementarity principle which forbids us from extracting all necessary information to reconstruct the state ρ from a single copy of the system. Any measurement on a generic quantum system inevitably disturbs it, so that after one measurement is done, almost no further information could be extracted from it about its state *before* the measurement. Take for example, the situation we considered in Section. 2.1.2, with the Kraus operators given by Eq. (2.30), the photon is destroyed after the measurement and no further information could be obtained. In order to reconstruct its state, one needs to measure an ensemble of independently and identically prepared such photons. However, as one only has a finite number of copies, what one gets in the experiment is only the relative frequencies of each outcome, not the true probabilities, because there will always be fluctuations in the data. Quantum state tomography, or quantum state estimation, is about estimating the state from these measurement data.

The scenario is hence depicted in Fig. 2.3. In general, we have an unknown input state ρ to be estimated. The state is sent to a measurement apparatus described by a POM $\{\Pi_i\}$, with K outcomes. Each of the detectors D_i corresponds to a particular outcome Π_i. What is observed in the experiment is a sequence of detector clicks. One then knows the total number of clicks of each detector n_i. The relative frequencies are then computed using

$$f_i = \frac{n_i}{N}, \quad \text{with} \quad N = \sum_{i=1}^{K} n_i, \tag{2.63}$$

and the task of quantum state tomography is to find an estimator, $\hat{\rho}$, for the unknown input state ρ_{true}, given these detector clicks $\{n_i\}$.

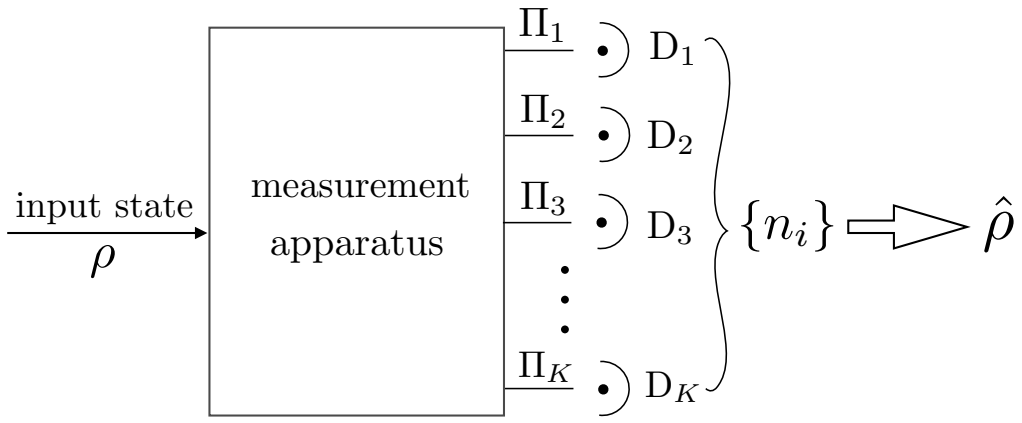

Figure 2.3: In general, we have an unknown input state ρ to be estimated. The state is sent to a measurement apparatus described by a POM $\{\Pi_i\}$, with K outcomes. Each of the detectors D_i corresponds to a particular outcome Π_i. What is observed in the experiment is a sequence of detector clicks. One then knows the total number of clicks of each detector n_i. Quantum state tomography is to reconstruct an estimator $\hat{\rho}$ for the input state from the measurement data $\{n_i\}$.

While the probabilities are uniquely determined by ρ according to the Born's rule, the converse may not be true. We remind ourselves that an arbitrary density matrix has d^2-1 independent parameters taking into consideration of its hermiticity and unit-trace properties. Hence, to fully characterize a state, we need to perform measurements to acquire at least d^2 independent *outcomes* to obtain enough information about the state.[6] A Set of measurement operators which provides enough information to characterize an unknown state is called an *informationally complete* (IC) POM. For *minimal* IC POM, exactly d^2 outcomes are measured. Hence, the probabilities uniquely determine the ρ if and only if the POM is IC. If the POM is informationally incomplete, then there will generally be more than one state that could give rise to the same data. In order to single out one estimator, additional criteria have to be invoked. We will discuss one such criterion later. Generally, due to the statistical nature of quantum measurement outcomes, measurements on many identical copies of the state are needed in order to obtain a precise and reliable state identification.

[6]We need d^2 (instead of d^2-1) independent outcomes for normalization of the probabilities because the total number of measured copies are usually not known *a priori*.

At this point, one may object that even in classical inference theory, the same data fluctuation is also present, which limits the accuracy of the estimator. Such objections are indeed valid. However, there is an added restriction in quantum state estimation, that is the estimator should be a positive estimator in order to make sense. There should not be any measurement that gives a negative probability. Hence, if one simply inverts the Born's rule, the resulting estimator may not be a positive estimator. In classical estimation, all resulting estimators are valid. In a nutshell, quantum state estimation is then classical estimation plus quantum constraints on top of it. All the statistical tools that we know about classical inference are still valid in the quantum regime. However, the quantum constraints often make the problem much more difficult.

As a concrete example, let us consider the so-called *trine measurement* performed on a qubit [Pv04]. The classical analog is that of a three-sided die, characterized by $\{p_1, p_2, p_3 = 1 - p_1 - p_2\}$. Classically, the valid probability space is an equilateral triangle, shown in Fig. 2.4. The three vertices correspond to the three extremal points $(p_1, p_2, p_3) = (1, 0, 0)$, $(0, 1, 0)$ or $(0, 0, 1)$ respectively. However, there are quantum constraints in the case of the trine measurement, and the valid probability space are restriced to

$$p_1^2 + p_2^2 + p_3^2 \leq \frac{1}{2}, \tag{2.64}$$

which limit the valid states to the circle inscribed in the triangle.

Finally, we remark that quantum state tomography is a central and recurring theme in quantum information science and quantum computation. Quantum state preparation is the first important step for any protocol that makes use of a source of quantum systems. For instance, a quantum-state teleportation protocol that is carried out using optical equipment requires a source that produces two-photons that are each prepared in a maximally-entangled quantum state. In a typical scenario, a source emits a certain desired state which carries the information, or is needed for the computational task. In order to verify the integrity of the quantum state that appropriately describes the source prepared, one carries out quantum-state tomography on the source. Quantum state tomography is a rich field by itself that is still under research. In the next part, we will first look at point estimator, which is

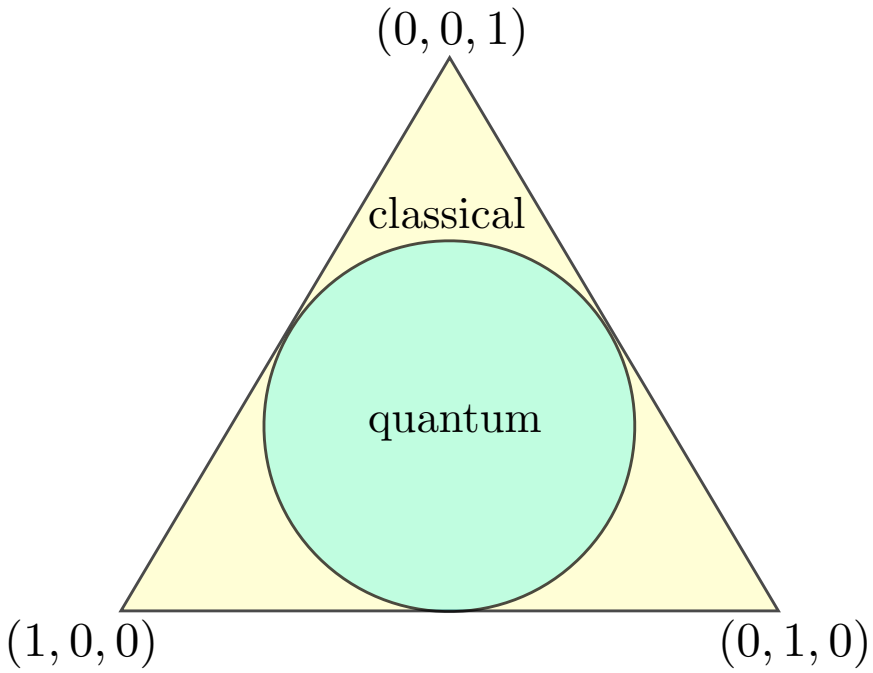

Figure 2.4: The classical analogy of the trine measurement is that of a three-sided die, characterized by $\{p_1, p_2, p_3 = 1 - p_1 - p_2\}$. The three vertices correspond to the three extremal points $(p_1, p_2, p_3) = (1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ respectively. Classically, any point in the equilateral triangle is a valid probability state of the die. However, there are quantum constraints in the case of the trine measurement, which limit the valid states to the circle inscribed in the triangle.

reporting a single estimator in the state space. When the sample size is large where the relative frequencies approach the probabilities, reporting a point estimator is justified. In this case where the central limit theorem applies, the estimation errors attached with the estimator could be found using the Fisher information. We will then move on to region estimator, which is reporting a set of estimators in the state space. Such region estimators are especially important when the sample size is small and one should not be so conclusive as to report only a single estimator in the state space.

2.2.2 Point estimator

There are many well-studied quantum tomographic methods which enable one to reconstruct a single estimator from the measurement data. The choice of which one to use is still a subject under intense debate and discussion [SKR+15, SNE14]. Inevitably, the choice may also depend on the system under consideration as well as the application in mind. In this section, for the purpose of this thesis, we will quickly review three such methods: the *maximum likelihood* (ML) estimation method [Fis22, Hel76a], the Bayesian mean (BM) estimation method [CL08, Hel76b, SBC01], and the *maximum likelihood maximum entropy* (MLME) estimation method [TZE+11, TSE+12], while very briefly mention some of the other techniques.

Maximum likelihood estimation

Suppose we performed an experiment on a prepared "true" state ρ_{true}, and obtained the frequencies $\{f_j\}$ for the POM $\{\Pi_j\}$. By denoting $p_j = \text{tr}\{\Pi_j \rho\}$ as the probability of getting the outcome j for the state ρ, the *likelihood* L for given ρ to produce the observed data is then

$$L(\{n_j\}|\rho) = \prod_j p_j^{n_j}. \tag{2.65}$$

The likelihood function is the probability (apart from the constant multinomial factor for a give data) for observing the data given the state ρ. Given no prior knowledge about the system, one estimates the state which maximizes the likelihood, ρ_{ML}, as the true state ρ_{true}. In a nutshell, ML estimation is to look for the state for which the data is most likely to occur. For an IC POM, this *maximum-likelihood* (ML) *estimator* is unique. Ideally, ρ_{ML} should coincide with ρ_{true}, and satisfy $\text{tr}\{\rho_{\text{ML}}\Pi_j\} = p_j = f_j$. Realistically however, there are unavoidable fluctuations in the outcomes due to statistical nature of quantum phenomena, and possibly imperfections in real experimental set-up. It is then almost certain that ρ_{ML} will not coincide with ρ_{true}.

ML estimator has the advantage that the positivity constraint of the state is automatically taken care of, by searching only among the valid density operators. Unlike linear inversion, it will never give rise to an invalid statistical operator. How-

ever, ML estimators are often rank-deficient since the true maximum of the likelihood often lies outside the state space. Hence, ML estimators are also biased. This unfortunate terminology seems to suggest that ML estimator is"bad" in some sense, which is not really true. What matters is the consistency of the estimator. ML estimators are *consistent* estimators in the sense that when the measurements are performed over a reasonably large number of copies of the system, ρ_{ML} will be close to ρ_{true}, and $\rho_{\mathrm{ML}} \to \rho_{\mathrm{true}}$ when $N \to \infty$. When the number of copies measured is large, the likelihood function becomes very narrowly peaked. The ML estimator is then one's best guess about the unknown state, since any other state will have a negligibly smaller likelihood value as compared with the ML estimator. Efficient computer algorithms to perform maximum-likelihood quantum state estimation are readily available, as can be obtained from references [Pv04, ŘHKL07].

Bayesian mean estimator

In a typical experiment in tomography, it is desirable to measure a large number of copies N, so that the resulting state estimator is as close to the true state as possible in the asymptotic limit. However, there are situations where the value of N is far from this limit, due to various reasons. Depending on the nature and physical carrier of the quantum system, the state could be difficult to generate and one then only has limited resources. For example, to perform tomography on quantum systems where the Hilbert space dimension d is large, typically correlated many-body systems, the number of POM outcomes d^2 for IC measurement increases rather fast. Not only that, for each outcome, one has to measure a sufficiently large number of copies. Therefore in these situations, in order to obtain a good tomographic accuracy of the estimator, N would have to increase rather quickly with d. In addition, the generation of these correlated systems, out of a multi-photon source for example, generally becomes more difficult and rarer as d increases. So the time needed to measure a fixed number of copies of these systems increases. Thus, realistic experiments on these complex systems would involve relatively fewer copies.

In the regime of small N, the likelihood has a broad peak, and neighboring

quantum states around the maximum give approximately the same likelihood as the maximal value. As such, it is statistically reasonable to consider all these neighboring states also as plausible states of the physical source. The natural resulting estimator is then to report the average of all these estimators, weighed by their respective likelihood values. However, as the state space is continuous rather than discrete, one needs to perform some form of integration rather than simply "adding". The idea of marginalization can thus be employed when the observer incorporates his or her prior distribution ($\mathrm{d}\rho$) about the true state of the source into state estimation.

The resulting estimator constructed from this procedure is the Bayesian mean estimator, that is

$$\hat{\rho}_{\mathrm{B}} = \frac{\int(\mathrm{d}\rho)L(\{n_i\}|\rho)\rho}{\int(\mathrm{d}\rho)L(\{n_i\}|\rho)}. \tag{2.66}$$

As can be noticed, one disadvantage of BM estimation is that it is difficult to handle the operator integral. This integral is to be performed over the admissible state space, which means that the parameters can vary only up to the positivity constraint for ρ. As the boundary of the state space is highly complicated, it is impossible to do the integrations by hand. One is forced to do these numerically. One way of doing the integral is to use a Monte Carlo method [SSN$^+$15a, SSN$^+$15b].

Maximum-likelihood-maximum-entropy (MLME) estimation

If the measurement does not form an IC POM, one could then have many estimators which are consistent with the experimental data with equally high likelihood. To still have a unique identification of the state, additional constraints need to be enforced. One such method is known as the *maximum-likelihood-maximum-entropy* (MLME) scheme, in which after the maximum-likelihood estimation stage, one chooses the estimator that maximizes the von Neumann entropy $S(\rho) = -\mathrm{tr}\{\rho \log_2 \rho\}$ [TZE$^+$11, TSE$^+$12]. For a given set of frequencies $\{f_j\}$, this MLME estimator ρ_{MLME} is unique. As the entropy is a measure of uncertainty in a physical system [NC10], the MLME estimator corresponds to the least-bias and most conservative guess for true state consistent with the measurement data.

Finally, we remark that there are also other estimation schemes such as linear inversion [Fan57], minimax estimators [NE12], hedged maximum likelihood estimators [BK10a], etc. We refer to the interested reader to references.

Whichever the technique to choose, one needs to know how accurate the estimation is and how close the estimator is to the target state. To measure how "close" the estimator is to the target ρ_{true}, the *fidelity*, F, which generalizes the overlap between two kets, is used [Joz94, NC10]. For any ρ and ρ_0, the fidelity is given by

$$\begin{aligned} F &= \text{tr}\Big\{\sqrt{\sqrt{\rho}\rho_0\sqrt{\rho}}\Big\} \\ &= \text{tr}\{|\sqrt{\rho}\sqrt{\rho_0}|\}. \end{aligned} \tag{2.67}$$

Obviously, $F = 1$ when the two states are identical, and $F = 0$ when they correspond to blends of orthogonal states, i.e. $\rho = \sum_j |\phi_j\rangle g_j \langle\phi_j|$, $\rho_0 = \sum_j |\psi_j\rangle f_j \langle\psi_j|$, where $\langle\psi_j|\phi_k\rangle = 0\ \forall\{j,k\}$. When both states are pure, F reduces to the absolute value of their overlap, $F = |\langle\psi|\phi\rangle|$.

Another commonly used distance measure between two states is the *trace distance* [NC10]. It is given by

$$D(\rho,\sigma) \equiv \frac{1}{2}\text{tr}\left\{|\rho-\sigma|\right\}. \tag{2.68}$$

If two statistical operators are close in trace distance, then any measurement performed on those states will result in probability distributions which are close together in the classical sense of trace distance. In another words, no measurement will be able to distinguish these two states very well. It is also worth mentioning that the trace distance and the fidelity are closely related, despite their different forms. In the case of pure states, they are completely equivalent to each other. For many purposes, it does not matter which one is used to quantify the distance.

2.2.3 Region estimator

Whether one prefers the ML estimator, Bayesian mean estimator, or any other estimator mentioned above, it is a single state which is a point in the state space. The

data unavoidably have statistical fluctuations and therefore, we should supplement them with error bars (in the case of a one-dimensional problem). Strictly speaking, these are error regions in the state space for higher-dimensional problems. Many recipes, which are often ad hoc in nature, have been used. They usually rely on having a lot of data, involve data resampling or consider all data that one might have observed. For example, the confidence regions recently studied are a set of regions, one region for each data, whether observed or not, from the measurement of N copies. To find a confidence region for the actual data that is observed, one must specify the whole set of confidence regions for other data that might have, but is not observed, because the confidence level is a property of the whole set.

For us, we would like to construct error regions that are based on only the data actually observed. For this purpose, we find the concept of the bounded-likelihood regions introduced in Ref. [SNS$^+$13] natural and fitting, as an extension and generalization of the maximum likelihood estimator. For a given data, instead of asking which estimator gives the largest likelihood value, we ask which region gives the largest likelihood for that region, $L(D|\mathcal{R})$. All regions that enter the competition should of course have the same size, $\mathcal{S}_\mathcal{R}$. This maximization problem gives the bounded-likelihood regions as the result. The basic idea is then to report a region in space states which consists of all states ρ for which the point likelihood exceeds a threshold value. See, for example, Fig. 2.5 for an illustration of the bounded-likelihood region, here plotted for the case of a single parameter θ. $\hat{\theta}_{\mathrm{ML}}$ is the maximum likelihood point estimator. The red line is the bounded-likelihood region $\mathcal{R}_\lambda$, for the threshold value λ.

The bounded-likelihood regions are also the smallest credible regions, that is, given a certain desired credibility, this region is the smallest in size. The notion of the size and credibility is central in the construction of such regions. The size of the region is quantified by the prior probability of the region, that is, the probability that the true state is in the given region *before* the data is collected. The credibility, on the other hand, is the posterior probability, that is, the probability that the actual state is in the region given that the data have been obtained. The bounded-likelihood regions then have straightforward and clear interpretations: It is the

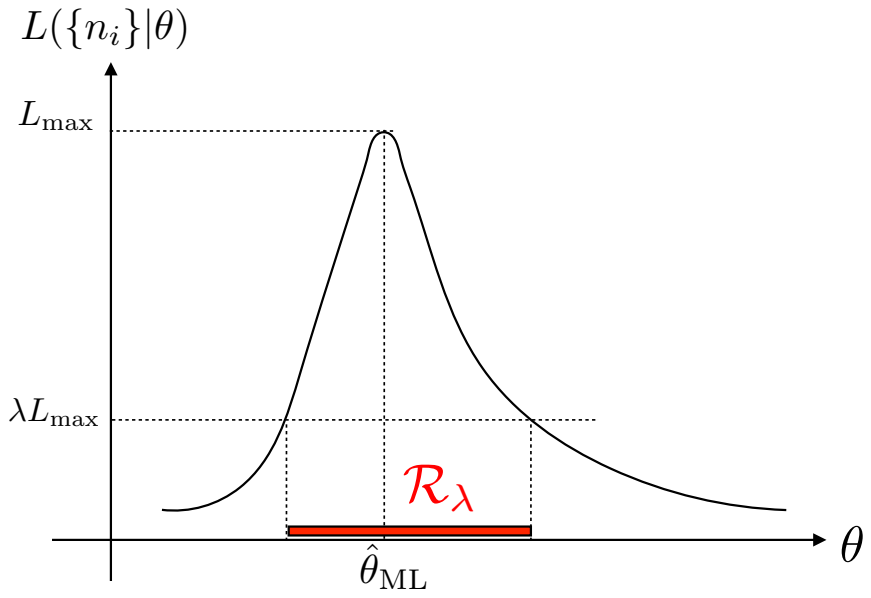

Figure 2.5: An illustration of the bounded-likelihood region, here plotted for the case of a single parameter θ. $\hat{\theta}_{\mathrm{ML}}$ is the maximum likelihood point estimator. The red line is the bounded-likelihood region $\mathcal{R}_\lambda$, for the threshold value λ.

region for which the observed data are most likely among all regions of the same size. It is also the region that contains the actual state with high probability. If one fixes the desired credibility, then this is the smallest region that one can report.

Chapter 3

Controllable Generation of Mixed Two-Photon States

In this chapter, we report a controllable method for producing mixed two-photon states via spontaneous parametric down-conversion with a two-type-I crystal geometry. By using variable polarization rotators (VPRs), one obtains mixed states of various purities and degrees of entanglement depending on the parameters of the VPRs. The generated states are characterized by quantum state tomography. The experimental results are found to be in good agreement with the theory. The method can be easily implemented for various experiments that require the generation of states with controllable degrees of entanglement or mixedness.

3.1 Introduction

Quantum information is a promising field that utilizes the nonclassical aspects of physical systems to perform sophisticated tasks of computations and communications. Quantum entanglement plays an important role in the implementation of these tasks. Arguably, the most famous entangled states are the Bell states, which are pure and maximally entangled. They are used in quantum cryptography [Eke91], quantum teleportation [BPM$^+$97], and the demonstration of various concepts of quantum mechanics [CS78, AGR81]. However, apart from the highly entangled pure states, there exists a vast uncharted region of state space, where states can be simultaneously mixed and entangled [Ter00, HHH96]. Mixed states are useful

in investigations of quantum computing [CGK98], studies of the quantum-classical interface [KE04], and decoherence channels [ZB05]. These applications motivate an interest in the generation and characterization of mixed states.

Up to now, several methods have been suggested for generating photonic mixed states. In Ref. [LHLLK06], Werner states are produced with controlled addition of white noise to the Bell states. In Ref. [WAB$^+$05], more sophisticated schemes for producing broad classes of states are presented. Such schemes employ an incoherent temporal mixing of state amplitudes, several decoherers, or a hybrid technique. However, as commented in Ref. [WAB$^+$05], such schemes require many additional optical components, and are challenging in practice.

In this chapter, we report a controllable method for producing mixed states, which requires only few additional optical components added to a Bell-state generation set-up. We use variable polarization rotators (VPRs), placed in the pump and signal beams of a conventional Spontaneous Parametric Down-Conversion (SPDC) set-up with two type-I crystals. Varying the parameters of the VPRs enables us to obtain mixed states of various purities and degrees of entanglement, which are then characterized by quantum state tomography. We remark that VPRs have been used earlier by Gogo *et al.* in studies of quantum erasure [GSB05] without, however, presenting a detailed study and systematic characterization of the generated states.

3.2 Mixed-state generation with VPR

The basic principle underlying this method of producing mixed states is to generate incoherent mixtures of orthogonal Bell states with controllable weights. Consider a SPDC set-up with two type-I crystals with orthogonal axes, pumped by a laser polarized at $-45°$ [KWW$^+$99], also see Fig. 3.1. The first beta-barium borate (BBO) crystal produces pairs of horizontally polarized photons, while SPDC at the second crystal produces vertically polarized photons. As the coherent length of the pump beam is much longer than the thickness of the crystal, it is indistinguishable whether a given pair is produced in the first or the second crystal, resulting in a coherent superposition of horizontally and vertically polarized photon pairs. The produced

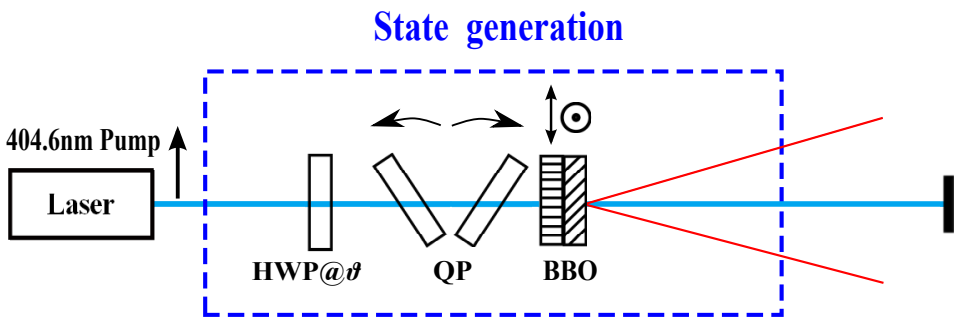

Figure 3.1: Schematic for the generation of the Bell states. A continuous-wave diode laser pumps two type-I BBO crystals with optic axes on orthogonal planes, and the SPDC occurs in the non-collinear frequency-degenerate regime. When the HWP is set at $\pm 22.5°$, it changes the vertically polarized pump photons to $\mp 45°$, and thereby produces the Bell states. One can set the HWP at an arbitrary angle ϑ to generate a class of rank-1 states given in Eq. (3.8), more about this later. Quartz plates (QP) are used to control the relative phase between the generated states from the two crystals. For more details, see [Len14].

state is then the Bell state

$$|\Phi_-\rangle = \left(|\mathrm{H_sH_i}\rangle - |\mathrm{V_sV_i}\rangle\right)/\sqrt{2}, \tag{3.1}$$

where H and V denote horizontal and vertical polarizations of down-converted photons, respectively, and the subscripts s-signal, and i-idler (hereafter omitted for simplicity) label the spatial modes. If the pump is polarized at $+45°$, the orthogonal Bell state $|\Phi_+\rangle = (|\mathrm{HH}\rangle + |\mathrm{VV}\rangle)/\sqrt{2}$ is generated. By blending these two Bell states with different weights, one obtains the incoherent mixture

$$\begin{aligned}\rho &= \alpha|\Phi_-\rangle\langle\Phi_-| + (1-\alpha)|\Phi_+\rangle\langle\Phi_+| \\ &= \frac{1}{2}\Big(|\mathrm{HH}\rangle\langle\mathrm{HH}| + |\mathrm{VV}\rangle\langle\mathrm{VV}|\Big) + \Big(\frac{1}{2}-\alpha\Big)\Big(|\mathrm{HH}\rangle\langle\mathrm{VV}| + |\mathrm{VV}\rangle\langle\mathrm{HH}|\Big),\end{aligned} \tag{3.2}$$

where α $(0 \leq \alpha \leq 1)$ is the weight of the state $|\Phi_-\rangle\langle\Phi_-|$ in the mixture.

To generate such mixed states, one switches the polarization of the pump back and forth between $-45°$ and $+45°$ and averages over a sufficient interval of time to sample both polarization states. Such a switching can be realized by inserting a VPR with electrically driven retardance in the pump beam. By applying a square waveform across the VPR, one achieves fast flippings between the two orthogonal linear polarizations. The parameter α is equal to the duty cycle (DC $= t/\tau$, where

t is the duration of the high voltage, and τ is the period of the waveform) of the applied square waveform.

The purity $P = \mathrm{tr}\{\rho^2\}$ of such states is given by

$$P = 2\Big(\alpha - \frac{1}{2}\Big)^2 + \frac{1}{2}. \tag{3.3}$$

When α is zero or one, one of the Bell states is produced, and the purity is one. When α is 0.5, the two Bell states are mixed with the same proportion, and the purity has the minimum value of 0.5. Thus, changing α enables one to obtain states with different purities.

Different amounts of entanglement can also be obtained by varying the parameter α. The tangle T is a measure of quantum-coherence properties of a quantum state [Woo98, CKW00]. It has a value of zero for separable states, and one for maximally entangled states. For the states given by Eq. (3.2), the tangle is related to α by

$$T = (1 - 2\alpha)^2. \tag{3.4}$$

Generated states can be characterized by polarization correlation analysis. For the Bell states, the polarization correlation gives a visibility of 100% when measured in the $\pm 45^\circ$ basis. For the states given by Eq. (3.2), the visibility varies with α according to

$$V = |1 - 2\alpha| = \sqrt{T}. \tag{3.5}$$

In particular, when $\alpha = 0.5$, the generated state is

$$\rho = \frac{1}{2}\Big(|\mathrm{HH}\rangle\langle\mathrm{HH}| + |\mathrm{VV}\rangle\langle\mathrm{VV}|\Big), \tag{3.6}$$

which results in zero visibility of the polarization correlation in $\pm 45^\circ$ basis. Note that for the state given by Eq. (3.2), $V = \sqrt{T}$, however, this relation is not true in general.

With a single VPR in the pump beam and α set to 0.5, one produces mixed states with minimal purity P, $= 0.5$, see Eq. (3.3). However, for a completely mixed state, $P = 1/d$, where d is the dimension of the density matrix ($d = 4$ for a two-photon

state). The completely mixed state can be written as:

$$\rho_{\mathrm{M}} = \frac{1}{4}\Big(|\mathrm{HH}\rangle\langle\mathrm{HH}| + |\mathrm{VV}\rangle\langle\mathrm{VV}| + |\mathrm{HV}\rangle\langle\mathrm{HV}| + |\mathrm{VH}\rangle\langle\mathrm{VH}|\Big) = \frac{\mathbb{1}}{4}. \tag{3.7}$$

Thus, one has to include in the mixture the other two Bell states $|\Psi_-\rangle = (|\mathrm{VH}\rangle - |\mathrm{HV}\rangle)/\sqrt{2}$, and $|\Psi_+\rangle = \mathrm{VH}\rangle + |\mathrm{HV}\rangle)/\sqrt{2}$ with equal weights. These two Bell states can be generated in the same set-up by inserting a VPR with a half-wave retardance in either the signal or idler beam. In this case the VPR transforms the states $|\Phi_-\rangle \longrightarrow |\Psi_-\rangle$, and $|\Phi_+\rangle \longrightarrow |\Psi_+\rangle$. By switching between zero and half-wave retardance of the VPR, incoherent mixtures of $|\Phi_-\rangle\langle\Phi_-|$ and $|\Psi_-\rangle\langle\Psi_-|$, or $|\Phi_+\rangle\langle\Phi_+|$ and $|\Psi_-\rangle\langle\Psi_-|$, are produced. Hence, by using two VPRs with DCs set to 0.5, one placed in the pump beam, and the other in the signal or idler beam, one can obtain the completely mixed state given by Eq. (3.7).

A Werner state [Wer89] can then be produced by mixing this completely mixed state with the singlet state. Notice that an important family of states, namely the Bell diagonal mixed states, can also be generated if the waveforms and duty cycles of the two VPRs are independently controlled. More generally, one can obtain mixtures of non-maximally entangled states

$$|\Psi_\vartheta\rangle = \cos\vartheta|\mathrm{HH}\rangle + \sin\vartheta|\mathrm{VV}\rangle \tag{3.8}$$

by rotating the polarization of the pump beam to an arbitrary angle. This enables the generation of even more states such as the Collins−Gisin states [CG04], given by

$$\rho_{\mathrm{CG}}(\alpha,\vartheta) = \alpha|\Psi_\vartheta\rangle\langle\Psi_\vartheta| + (1-\alpha)|\mathrm{HV}\rangle\langle\mathrm{HV}|. \tag{3.9}$$

3.3 Experimental Set-up

3.3.1 State Preparation

The experimental set-up is shown in Fig. 3.2. Two BBO crystals of 0.8 mm thickness with orthogonal optic axes are cut at 29.5° for noncollinear frequency-degenerate

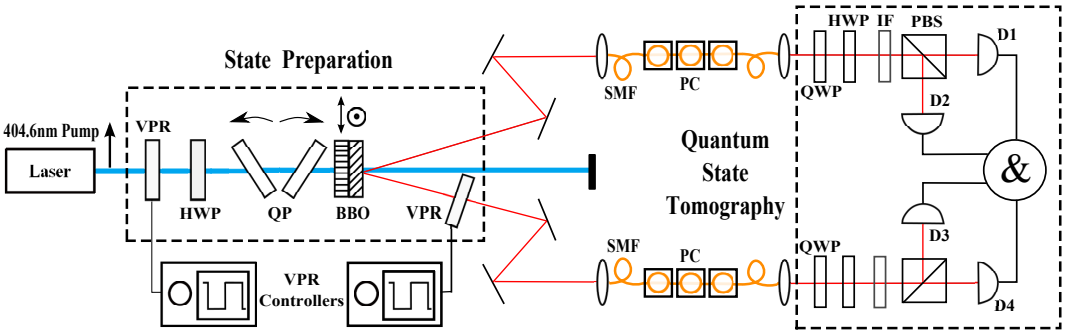

Figure 3.2: Experimental set-up. Two type-I BBO crystals with orthogonal axes are pumped by a cw diode laser. The SPDC is operated in the noncollinear frequency-degenerate regime. Mixed states are generated by inserting variable polarization rotators (VPRs) in the pump and signal beams. QP, are quartz plates used to control the phase of the produced states. The SPDC photons are coupled into single-mode fibers (SMF) with lenses (L). PC are polarization controllers, IF - interference filters. Quarter- and half-wave plates (QWP, HWP) and polarizing beam splitters (PBS) are used for quantum state characterization. D1-4 are single photon detectors, whose outputs are processed by a coincidence circuit (&).

phase matching. The BBOs are pumped by a frequency stabilized continuous wave diode laser at a wavelength of 404.6 nm (Ondax, LM series). The first crystal produces pairs of vertically polarized photons from the pump's horizontal polarization component, while the second crystal produces pairs of horizontally polarized photons from the pump's vertical polarization component. With noncollinear SPDC, the down-converted photons travel in different directions with a cone opening angle of 4°. Down-converted photons of narrow spatial bandwidth are collected into single-mode optical fibers (SMF) using aspherical lenses (L) with a focal length of 11 mm, placed at a distance of 800 mm from the BBOs. The down-converted photons are filtered with interference filters (IF) centered at 810 nm, with 10 nm full width at half maximum. Polarization rotations in the optical fibers are compensated for by manual polarization controllers (PC).

With a half-wave plate (HWP) in the pump beam at 22.5°, the vertical polarization of the pump is rotated to −45°, and the resulting state is

$$|\,\rangle = \frac{|\mathrm{HH}\rangle - e^{\mathrm{i}\phi}|\mathrm{VV}\rangle}{\sqrt{2}}, \tag{3.10}$$

where ϕ is the relative phase. This phase can be set to zero or integer multiples of 2π by manipulating the ellipticity of the pump with two quartz plates (QP) which

are placed after the HWP, thus producing $|\Phi_-\rangle = (|\text{HH}\rangle - |\text{VV}\rangle)/\sqrt{2}$.

Two different types of VPR have been used in the experiment as we only have one piece of each. The first one is a photoelastic modulator (HINDS Instrument, I/FS50). It is based on the photoelastic effect, in which a mechanically stressed optical device exhibits birefringence proportional to the resulting strain. The mechanical stress is being applied by a piezoelectric transducer. They are resonant devices where the precise oscillation frequency is determined by the properties of the transducer assembly. In our case, the frequency is at 50 kHz and the modulator switches between zero and half-wave retardance. However, it only has a fixed DC of 0.5, which cannot be tuned. The second one is a liquid crystal retarder (LCR) (Meadowlark, LRC-200). The liquid crystal molecules in the retarder have an ordered orientation, which together with the stretched shape of the molecules creates an optical anisotropy. When an electric field is applied, the molecules align to the field and the level of birefringence is controlled by the tilting of the LC molecules, which is controlled by the voltage applied. In our experiment, it is configured so that zero retardance is obtained at the "low" voltage of $V_\text{L} = 1.39\,\text{V}$. The pump polarization entering the crystal remains at $-45°$, and $|\Phi_-\rangle$ is generated. For the "high" voltage of $V_\text{H} = 1.64\,\text{V}$, half-wave retardance is obtained. The pump polarization entering the crystal changes to $+45°$, and $|\Phi_+\rangle$ is generated. A square waveform with the frequency of 1 Hz is applied to the LCR. The averaging is done over 3-minutes time intervals. By varying the DC of the square waveform, different mixed two-photon states are obtained, see Eq. (3.2).

To produce the completely mixed state, given by Eq. (3.7), the photoelastic modulator is placed in the pump beam, and the LCR with DC = 0.5 is placed in the signal beam. The operation voltages of the LCR for down-converted photons are adjusted to $V_\text{L} = 1.84\,\text{V}$ and $V_\text{H} = 6.95\,\text{V}$. The frequency of the LCR square waveform is kept at 1 Hz, and the averaging is done over 3-minutes time intervals.

3.3.2 State characterization

To fully characterize the states, quantum state tomography is performed. In both signal and idler arms, we install quarter-, and half-wave plates (QWP, HWP) and a

polarizing beam splitter (PBS) [JKMW01]. The transmission and reflection ports of both PBSs are directed to single-photon detectors (Silicon Avalanche Photodiodes, quantum efficiency $\sim 50\%$, Qutools Twin QuTD). Coincidence events are registered using the Time-To-Digital Converter (quTAU) with a time window of 5ns. The coincidences between any two of the detectors are recorded. By manipulating the wave plates in front of the PBSs, we perform measurements in nine different bases, which give an overcomplete measurement of 36 outcomes. Using the technique of maximum likelihood estimation [Pv04], the state is inferred from the collected data.

3.4 Results

For the calibration of the set-up, the polarization correlation analysis in the $\pm 45^\circ$ basis is performed with the Bell state $|\Phi_-\rangle$. The constant "low" voltage is applied across the LCR in the pump beam. The QWPs in the signal and idler arms are fixed at 0°. The HWP in the idler arm is fixed at -22.5° and the HWP in the signal arm is rotated. This causes the coincidence rate between the counts of two detectors in the transmitted ports of the PBSs to vary sinusoidally.

The visibility is defined as

$$V = \left| \frac{N_{+45^\circ} - N_{-45^\circ}}{N_{+45^\circ} + N_{-45^\circ}} \right|, \tag{3.11}$$

where N_{+45° and N_{-45° are the coincidence rates when the HWP in the signal arm is oriented at $+22.5^\circ$ and -22.5°, respectively. The obtained value of visibility $97.3 \pm 1.4\%$ indicates high-quality of the produced Bell state.

The analysis is then extended to the mixed states prepared with the LCR operating at different DCs, and with the photoelastic modulator. The obtained experimental visibilities are in good agreement with the theoretical expectation of Eq. (3.5), see Fig. 3.3. One observes a degradation of the visibility as α increases from 0 to 0.5. After which, the visibility is gradually restored as α increases from 0.5 to 1.

Next, quantum state tomography on the generated states is performed. The state

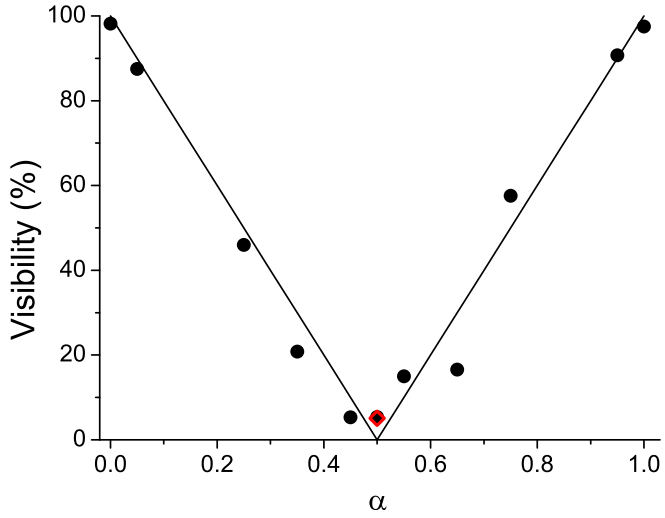

Figure 3.3: Dependence of the visibility of the polarization correlation measurements in the $\pm 45^\circ$ basis on the DC of the LCR (solid circle), and of the photoelastic modulator (red open diamond). The solid line is the theoretical prediction. The error bars are smaller than the symbols.

is reconstructed and the fidelities, F, of the reconstructed density matrices with the states given by Eq. (3.2) are calculated. Using the definition $F = \mathrm{tr}\{|\sqrt{\rho}\sqrt{\sigma}|\}$, where ρ is the maximum likelihood estimator, and σ is the target state, fidelities $F > 97\%$ are consistently observed, see Table. 3.1. The last two entries with PEM and $\mathbb{1}/4$ refer to the state generated using only the photoelastic modulator in the pump beam, and the completely mixed state generated using two VPRs, respectively.

α	Fidelity
0	0.9857±0.0005
0.05	0.9886±0.0005
0.25	0.9814±0.0006
0.35	0.9750±0.0012
0.45	0.9776±0.0007
0.50	0.9782±0.0006

α	Fidelity
0.55	0.9724±0.0012
0.65	0.9727±0.0008
0.75	0.9792±0.0007
0.95	0.9786±0.0012
1	0.9785±0.0003
PEM	0.9890±0.0005
$\mathbb{1}/4$	0.9942±0.0003

Table 3.1: Fidelities of the reconstructed states with the target states for various DCs of the LCR. $F > 97\%$ are consistently obtained for all the states. The last two entries with PEM and $\mathbb{1}/4$ refer to the state generated using only the photoelastic modulator in the pump beam, and the completely mixed state generated using two VPRs, respectively.

The matrix elements of reconstructed density matrices are shown graphically in Fig. 3.4, for α=0.05, 0.25, and 0.5. As an example, for $\alpha = 0.25$, the reconstructed density matrix expressed in the standard basis is

$$\rho_{\alpha=0.25} \widehat{=} \begin{pmatrix} 0.545 & 0.049+0.012\mathrm{i} & -0.013+0.038\mathrm{i} & -0.232+0.110\mathrm{i} \\ 0.049-0.012\mathrm{i} & 0.008 & -0.005+0.008\mathrm{i} & 0.015+0.004\mathrm{i} \\ -0.013-0.038\mathrm{i} & -0.005-0.008\mathrm{i} & 0.013 & -0.039+0.006\mathrm{i} \\ -0.232-0.110\mathrm{i} & 0.015-0.004\mathrm{i} & -0.039-0.006\mathrm{i} & 0.434 \end{pmatrix}.$$

The reconstructed state has a tangle of $T = 0.2476 \pm 0.0049$, a purity of $P = 0.6295 \pm 0.0022$, and a fidelity with the target state of $F = 0.9814 \pm 0.0006$.

As one can see from Fig. 3.4, an increase of α from 0 to 0.5 causes vanishing of the off-diagonal elements, which indicates the decreasing amount of entanglement of the mixed states. The calculated values of the tangle for the reconstructed states, are in good agreement with Eq. 3.4, see Fig. 3.5(a). The state with almost zero

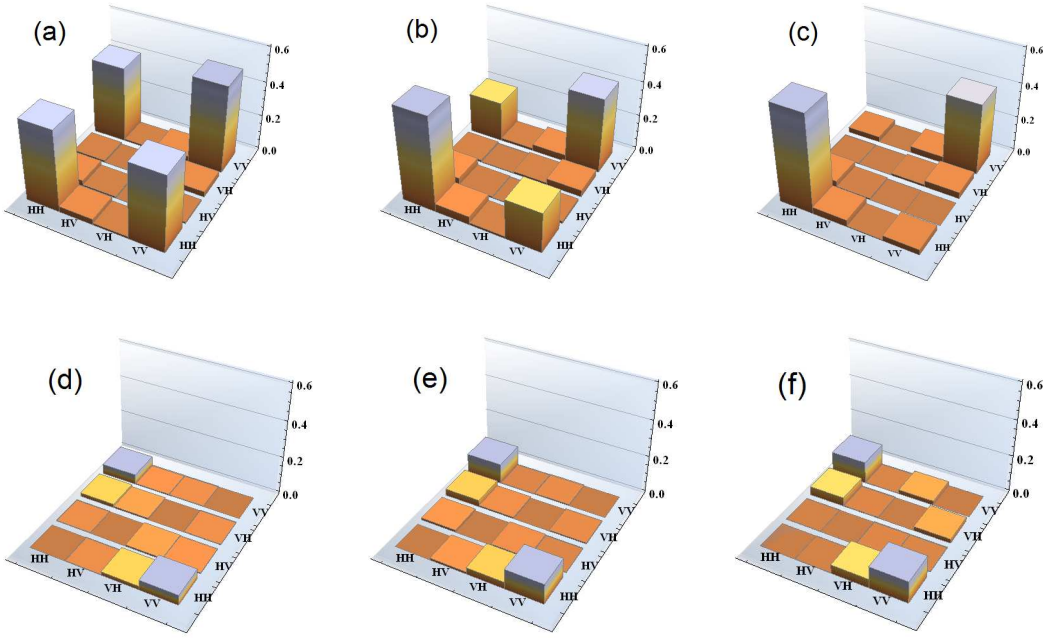

Figure 3.4: Absolute values of (a,b,c) real and (d,e,f) imaginary parts of the density matrices representing the reconstructed states for (a,d) 0.05DC; (b,e) 0.25DC and (c,f) 0.50DC. The vanishing of the off-diagonal elements is clearly seen.

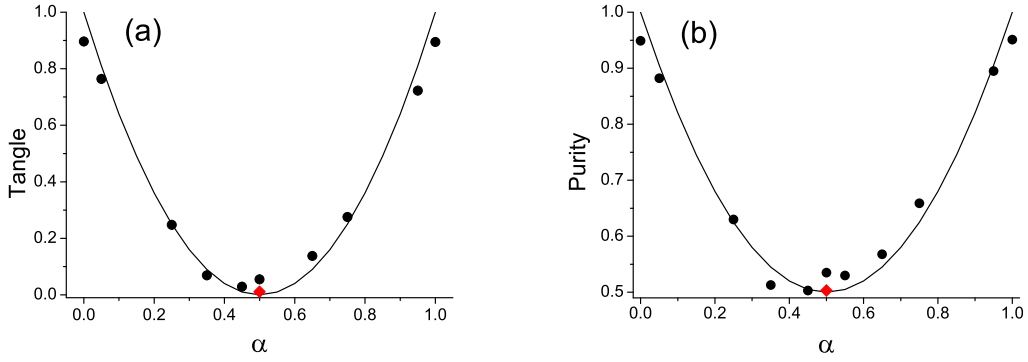

Figure 3.5: Dependence of (a) tangle and (b) purity of reconstructed states on the DC of the LCR (solid circles), and of the photoelastic modulator (red diamonds). The solid curves are the theoretical predictions. The error bars are smaller than the symbols used for both figures.

tangle is generated with the photoelastic modulator. It performs better than the LCR at α=0.5, as the wavelength of the pump laser is at the edge of the working range of the LCR.

Lastly, one verifies that the method indeed produces states with different amounts of mixedness. The dependence of the purity of the states on α is shown in Fig. 3.5(b). The obtained results are close to those predicted by Eq. (3.3).

To demonstrate generation of the completely mixed state, the set-up is modi-

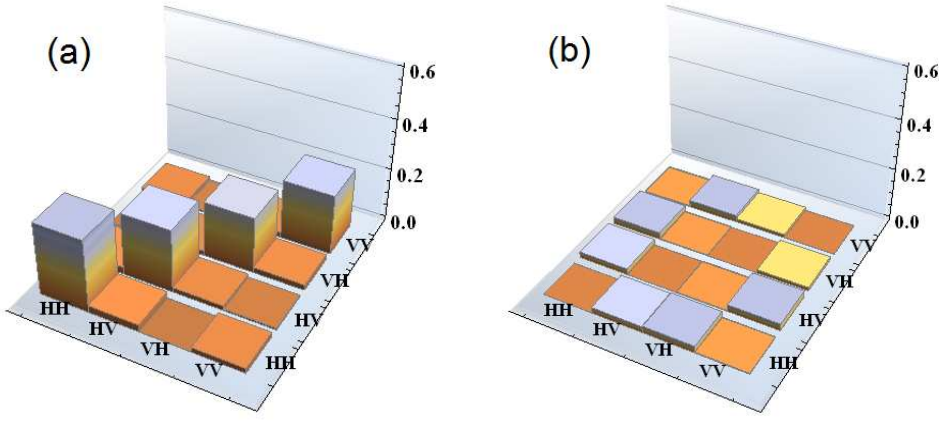

Figure 3.6: Absolute values of (a) real and (b) imaginary parts of the reconstructed density matrix for the completely mixed state.

fied by inserting VPRs in the pump and in the signal beams, see Sec. 3.1. The reconstructed density matrix of the state is shown graphically in Fig. 3.6. The reconstructed state has a tangle of zero, a purity of $P = 0.2615 \pm 0.0007$, and a fidelity with the target state of $F = 0.9942 \pm 0.0003$. Such a state is separable and highly mixed.

3.5 Conclusion

In summary, an accessible way of generating controllable mixtures of two Bell states with high fidelities is reported. There is no need for extra optical components, except the VPR. Furthermore, the method does not require any reconfiguration of the standard set-up for Bell states generation, as the VPR can be easily turned on and off, to generate mixed and pure states, respectively. In addition, it offers the important advantage of interferometric stability. The presented method will facilitate state generation and can be used in various quantum-information processing applications. We have used it for an experimental realization of qubit-pair tomography with witness bases which will be described in the next chapter [ZTE10, DLT$^+$14].

Chapter 4

Witness-Family Measurements

In this Chapter, we report an experiment in which one determines, with least tomographic effort, whether an unknown two-photon polarization state is entangled or separable. The method measures whole families of optimal entanglement witnesses. We introduce adaptive measurement schemes that greatly speed up the entanglement detection. The experiments are performed on states of different ranks, and we find good agreement with results from computer simulations.

4.1 Introduction

Entangled states play an important role in the manipulation of quantum information, be it for present-day quantum key distribution or future quantum computation. One may need to verify if a certain quantum state—perhaps emitted by a source of quantum-information carriers or obtained as the output of a quantum computation—is entangled or not. For this purpose, the expectation value of an *entanglement witness* is telling: The state is surely entangled if a negative value is obtained. A positive value, however, is inconclusive—the unknown state could be entangled or separable, the witness cannot tell, but other witnesses might be able to; recall Fig. 2.1 in Chapter 2.

How many witnesses, then, does one need to measure until a conclusion is reached? The answer to this question is given in Ref. [ZTE10]: If one solely relies on the expectation values of the witnesses one by one, one may never get a conclusive answer; if, however, the expectation values of suitably chosen witnesses

are jointly used for an estimation of the quantum state, $D^2 - 1$ witnesses suffice for a D-dimensional quantum system. The idea is to choose these witnesses so that their expectation values altogether enable IC quantum state tomography of the state. Even though each of them does not give any conclusive answer individually, jointly, they give full information of the state so that no more measurement is necessary.

The number can be further reduced by exploiting all the information gathered when determining the expectation values of the witnesses measured in succession. Traditional witness measurement only utilizes the expectation value of the witness operator. However, by measuring in the eigenbasis of the witness operator, one obtains many relative frequency values which can be exploited, rather than just a single expectation value. In the case of a two-qubit state ($D = 4$), for instance, one never has to measure more than six witnesses, rather than $15 = 4^2 - 1$. This can be demonstrated by an experiment such as the one proposed in Ref. [ZTE10]; we are here reporting its laboratory realization.

4.2 Witnesses and witness families

A hermitian observable W is an entanglement witness if $\mathrm{tr}\{\rho_{\mathrm{sep}} W\} \geq 0$ for *all* separable states ρ_{sep} and $\mathrm{tr}\{\rho_{\mathrm{ent}} W\} < 0$ for at least one entangled state ρ_{ent}. For each entangled state, there are some witnesses that detect it ("< 0"), but many other witnesses will give an inconclusive result ("≥ 0").

We shall concern ourselves with two-qubit systems—in the experiment, they are polarization qubits of a down-converted photon pair—and focus on optimal decomposable witnesses [LKCH00] of the form $W = \big(|\mathrm{w}\rangle\langle\mathrm{w}|\big)^{\mathrm{T}_2}$, where $|\mathrm{w}\rangle$ is the ket of an entangled pure two-qubit state, and $^{\mathrm{T}_2}$ denotes the partial transposition on the second qubit. This witness is optimal in the sense that no other witness can detect some entangled states in addition to the states already detected by W. The generic example is $|\mathrm{w}\rangle = |00\rangle \cos(\frac{1}{2}\alpha) + |11\rangle \sin(\frac{1}{2}\alpha)$ with $\sin\alpha \neq 0$, on which all other $|\mathrm{w}\rangle$s can be mapped by local unitary transformations.

For any α, the eigenkets of the resulting witness,

$$W^{(\alpha)} = |00\rangle\frac{1+\cos\alpha}{2}\langle 00| + |11\rangle\frac{1-\cos\alpha}{2}\langle 11| + |\Psi_+\rangle\frac{\sin\alpha}{2}\langle\Psi_+| - |\Psi_-\rangle\frac{\sin\alpha}{2}\langle\Psi_-|\,, \tag{4.1}$$

are the same: the two product kets $|00\rangle$ and $|11\rangle$ as well as the two Bell kets $|\Psi_\pm\rangle = (|01\rangle \pm |10\rangle)/\sqrt{2}$. This whole family of entanglement witnesses can, therefore, be measured by the projective measurement of their common eigenstates—the *witness basis* of the family.

For a separable state, $\mathrm{tr}\{\rho_{\mathrm{sep}}W^{(\alpha)}\} \geq 0$ for *all* α. Hence, we can apply a stricter criterion by searching among all the $W^{(\alpha)}$s for which the minimum value of $\mathrm{tr}\{\rho W^{(\alpha)}\}$ is obtained. In another words, this search of the minimum maximizes our chance of entanglement detection. Upon denoting the expectation values $p_1 = \mathrm{tr}\{\rho|00\rangle\langle 00|\}, p_2 = \mathrm{tr}\{\rho|11\rangle\langle 11|\}, p_3 = \mathrm{tr}\{\rho|\Psi_+\rangle\langle\Psi_+|\}$ and $p_4 = \mathrm{tr}\{\rho|\Psi_-\rangle\langle\Psi_-|\}$, we have

$$\begin{aligned}
&\min_\alpha\left\{\mathrm{tr}\{\rho W^{(\alpha)}\}\right\}\\
=&\min_\alpha\left\{p_1(\cos\alpha)^2 + p_2(\sin\alpha)^2 + (p_3-p_4)\sin\alpha\cos\alpha\right\}\\
=&\min_\alpha\Bigg\{\frac{p_1+p_2}{2} + \underbrace{\left(\frac{p_3-p_4}{2}\right)\sin(2\alpha) + \left(\frac{p_1-p_2}{2}\right)\cos(2\alpha)}_{\frac{1}{2}\sqrt{(p_1-p_2)^2+(p_3-p_4)^2}\sin\left(2\alpha+\tan^{-1}\left(\frac{p_1-p_2}{p_3-p_4}\right)\right)}\Bigg\}\\
=&\frac{p_1+p_2}{2} - \frac{1}{2}\sqrt{(p_1-p_2)^2+(p_3-p_4)^2}.
\end{aligned} \tag{4.2}$$

Hence, this requirement implies the witness-family criterion [ZTE10]

$$\mathcal{S} \equiv 4p_1p_2 - (p_3-p_4)^2 \geq 0\,. \tag{4.3}$$

In the experiment, we estimate these probabilities from the observed frequencies, that is, by measuring the relative frequencies for the two product states, f_1 and f_2, and those for the Bell states, f_3 and f_4. Consequently, once the frequency data are obtained from the measurement, a negative value of $\mathcal{S}$ reveals that the unknown state ρ_{true} is entangled.

Three remarks are in order. (i) The expectation value of one witness $W^{(\alpha)}$ is a linear function of the f_js, whereas $\mathcal{S}$ is a quadratic function. This is reminiscent of, yet different from, the "nonlinear entanglement witness" of Ref. [KKK10], which requires a joint measurement on two copies of the unknown state. Our *witness-family measurement* uses only one copy at a time. (ii) Also, the nonlinear witnesses of Ref. [GL06] are quite different; their evaluation requires complete or almost complete knowledge of the state. (iii) The witnesses $W^{(\alpha)}$ can also be measured by other schemes, such as that of Barbieri *et al.* [BDMDN$^+$03] who extracted the expectation value of $W^{(\pi/2)}$ from local measurements that examine the two qubits individually. By contrast, we perform a joint measurement on both qubits, which implements the projective measurement in the eigenbasis of $W^{(\alpha)}$ and so realizes the most direct measurement of the witness. More about (ii) and (iii) in Section. 4.8.

Our witness-family measurement provides estimates for *three* two-qubit observables (four f_js with unit sum), whereas the expectation value of a single witness is only one number. For example, for the witness basis given above, one gets $\{f_1, f_2, f_3, f_4\}$ from the eigenbasis measurement. This is equivalent to measuring the set of observables, $\{Z\mathbb{1} + \mathbb{1}Z, ZZ, XX + YY\}$, as one can show that

$$\left\langle Z\mathbb{1} + \mathbb{1}Z \right\rangle = 2(f_1 - f_2), \tag{4.4}$$

$$\left\langle ZZ \right\rangle = f_1 + f_2 - (f_3 + f_4), \tag{4.5}$$

and

$$\left\langle XX + YY \right\rangle = 2(f_3 - f_4). \tag{4.6}$$

To obtain the other observables, one measures another suitably chosen witness family. As discussed in Ref. [ZTE10], these data can be exploited for quantum-state reconstruction after measuring six witness families, related to each other by the six local unitary transformations of Table 4.1: Each witness family provides the expectation values of one (of six) single-qubit observables and of two (of nine) two-qubit observables. Therefore, a measurement of all six witness families constitutes an informationally complete (IC) measurement for full tomography of an unknown two-qubit state; thereby, all six single-qubit parameters and six two-qubit parameters

Family	U_1	U_2	Observables
1	$\mathbb{1}$	$\mathbb{1}$	$Z\mathbb{1}+\mathbb{1}Z$, ZZ, $XX+YY$
2	$\mathbb{1}$	X	$Z\mathbb{1}-\mathbb{1}Z$, ZZ, $XX-YY$
3	$C^\dagger$	C	$X\mathbb{1}+\mathbb{1}Z$, XY, $YZ+ZX$
4	$C^\dagger$	XC	$X\mathbb{1}-\mathbb{1}Z$, XY, $YZ-ZX$
5	C	$C^\dagger$	$Y\mathbb{1}+\mathbb{1}X$, YX, $ZY+XZ$
6	C	$XC^\dagger$	$Y\mathbb{1}-\mathbb{1}X$, YX, $ZY-XZ$

Table 4.1: The six witness families that enable full tomography of the two-qubit state. The single-qubit unitary operators U_1 and U_2 transform the first family into the other five families. The Pauli operator X permutes $|0\rangle$ and $|1\rangle$; the Clifford operator C permutes the three Pauli operators cyclically.

are obtained once, while three two-qubit parameters, that is, ZZ, XY and YX, are determined twice, see Table 4.1. The single-qubit unitary operators U_1 and U_2 transform the first family into the other five families. The Pauli operator X permutes $|0\rangle$ and $|1\rangle$; the Clifford operator C permutes the three Pauli operators cyclically, that is $CX = YC$, $CY = ZC$, and $CZ = XC$. It is given explicitly by

$$C = \frac{1}{2}\left(\mathbb{1} - \mathrm{i}X - \mathrm{i}Y - \mathrm{i}Z\right). \tag{4.7}$$

These six witness families will enable full tomography of the two-qubit state. This offers the possibility of measuring an IC set of witness families such that, if all families give an inconclusive result ($\mathcal{S} \geq 0$), a full state estimation can be performed for identifying ρ_{true}. With ρ_{true} then at hand, its separability can be determined straightforwardly by, for example, checking the Peres-Horodecki criterion [Per96, HHH96].

4.3 Three Schemes

4.3.1 Scheme A: Random sequence

Since ρ_{true} is unknown, there is no preference for a particular one of the six families to start with. Hence, one starts with a randomly chosen family and checks the inequality (4.3). If the result is inconclusive, one then chooses the next family at random from the remaining five families, and so forth until a conclusive result is obtained. If all six families give inconclusive results, ρ_{true} is estimated from the data

to establish if it is entangled or separable.

4.3.2 Scheme B: Adaptive measurements

Alternatively, we can perform the witness-family measurements in an adaptive manner: We choose the next family to be measured in accordance with the data obtained from previous measurements. Each time a witness family is measured, a set of four frequencies is obtained and these *informationally incomplete* data are used to partially estimate ρ_{true} by, for example, jointly maximizing the likelihood and the entropy—the MLME strategy of Ref. [TZE$^+$11, TSE$^+$12]. The MLME estimators ρ_{MLME} tend to be highly-mixed states and are thus hard to detect by entanglement witnesses; therefore, if the measurement of a witness family detects the entanglement of the MLME estimator by the criterion (4.3), measuring that family has a good chance of detecting the entanglement of ρ_{true}. The value of $\mathcal{S}$ is used for comparing the unmeasured witness families with the MLME estimator, with f_j replaced by the jth MLME probability $\text{tr}\{\rho_{\text{MLME}} U^\dagger \Pi_j U\}$ where $U = U_1 \otimes U_2$ is one of the six unitary operators of Table 4.1, and Π_j projects to the jth ket in the witness bases.

The family that gives the smallest value of $\mathcal{S}$ is then measured next; this judicious choice of family reduces the average number of witness families that need to be measured before the entanglement is detected. Instead of fixing the above six families, based on the MLME estimator, one can also choose from all thinkable families in each step. However, it is not worth the trouble as such optimization hardly improves the entanglement detection, see Section. 4.8.

4.3.3 Scheme C: Maximum-likelihood set

All state estimators, including ρ_{MLME}, that maximize the likelihood compose a convex set, the *maximum-likelihood (ML) set*. They all give the same estimated probabilities for the witness families already measured. When the number of qubit pairs measured per witness family is large (10^4 pairs suffice in practice), ρ_{true} is very likely contained in the ML set. Then, if there is no separable state in the ML set, we can conclude that ρ_{true} is entangled. For finite data, this conclusion is correct within a certain

error margin, as is the case for all conclusions drawn from entanglement witness measurements. For this non-separability check, we compute the maximal values of the likelihood for both the entire state space ($\mathcal{L}_{\max}$) and the entire space of separable states ($\mathcal{L}^{\text{sep}}_{\max}$) [ŘH03]. If we find that $\mathcal{L}_{\max} > \mathcal{L}^{\text{sep}}_{\max}$, we infer that ρ_{true} is entangled. To further economize the adaptive scheme, this check is performed before looking for the unmeasured witness family with the smallest value of $\mathcal{S}$.

4.4 Simulations

To investigate the efficiencies of the three schemes, we perform computer simulations of witness-family measurements with both pure and full-rank mixed two-qubit entangled states. We generate a unitarily invariant ensemble of random entangled states by the procedure of Ref. [ZS01]: For each random state, we first compose an auxiliary matrix $\mathcal{A}$ of dimensions 1×4 (for a pure state) or 4×4 (for a full-rank state), with the random complex entries chosen from a normal distribution with zero mean and unit variance; then, the 4×4 matrix representing the random state is $\mathcal{A}^\dagger\mathcal{A}/\text{tr}\{\mathcal{A}^\dagger\mathcal{A}\}$. Most of the random states generated this way are entangled, since there are many more entangled states in such an ensemble than separable states; we check the concurrence of each state to ensure that only entangled states are used in the simulation.

Figure 4.1 shows the histograms that summarize the cumulative distribution in the percentage of entangled states detected against the number of witness families needed using schemes A, B and C. We observe that the average number of families is largest for scheme A and smallest for scheme C. For instance, with only three witness families measured, scheme C detects about 95% of the rank-1 entangled states whereas scheme A will need the measurement of five families to reach the same detection rate. When using either scheme A or scheme B, about 2% of the random pure states and about 67% of the full-rank mixed states are undetected by the six families without performing full tomography. The additional separability check in scheme C reduces the percentage of undetected pure states to virtually zero, and one needs no more than five witness families to detect entanglement for the rest of the pure states. The improvement is even more dramatic for the mixed

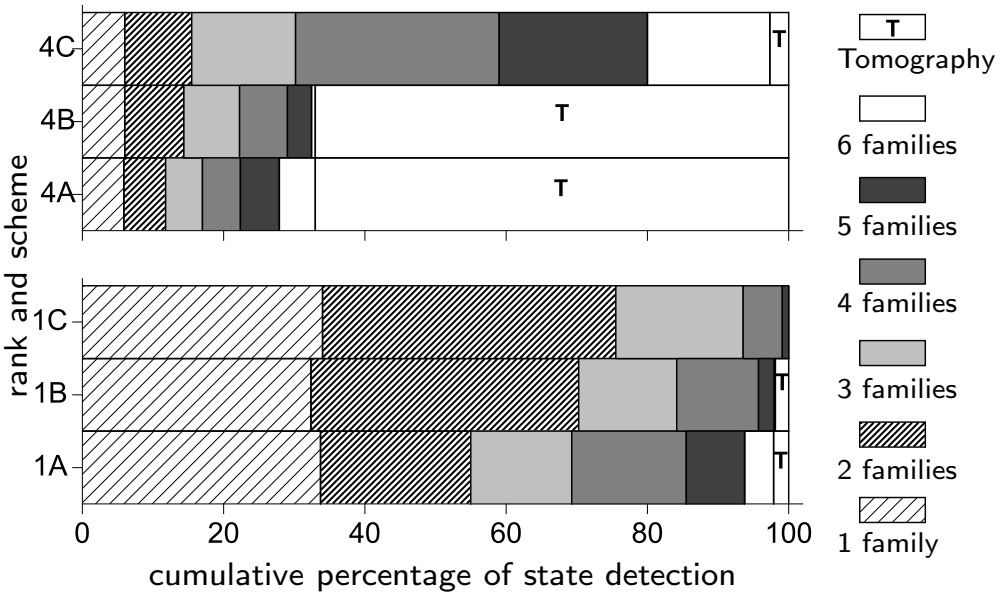

Figure 4.1: Simulation results on the measurement of the set of six informationally complete entanglement witness families for 10^4 randomly chosen two-qubit entangled states: pure states (bottom) and full-rank mixed states (top). The cumulative histograms compare between measurements performed with scheme A, scheme B, and scheme C.

states, with a reduction from about 67% to about 2.7%. We also observe that the mean number of witness families needed to detect entanglement for mixed states is larger than that for pure states. This is as expected, since mixed states generally have weaker entanglement and are, therefore, harder to detect.

4.5 Experiment

We experimentally test the entanglement detection and tomographic scheme with three classes of states of different ranks. The first class of states are the pure states $\rho_{\text{true}}^{(1)} = |\vartheta\rangle\langle\vartheta|$, with $|\vartheta\rangle = |00\rangle \sin\vartheta + |11\rangle \cos\vartheta$ for $0 < \vartheta < \pi$, $\vartheta \neq \pi/2$. The second class of states are rank-two states of the form $\rho_{\text{true}}^{(2)} = |\Phi_+\rangle\mu\langle\Phi_+| + |\Phi_-\rangle(1-\mu)\langle\Phi_-|$ for $0 \leq \mu \leq 1$ and $\mu \neq 1/2$, where $|\Phi_\pm\rangle = (|00\rangle \pm |11\rangle)/\sqrt{2}$. The third class of states are the Werner states $\rho_{\text{true}}^{(3)} = |\Psi_-\rangle\lambda\langle\Psi_-| + (1-\lambda)/4$ for $1/3 < \lambda \leq 1$. The experiment uses the polarization qubits of a down-converted photon pair at 810 nm with, for example, ket $|10\rangle$ standing for the photon in mode 1 horizontally polarized and the photon in mode 2 vertically polarized.

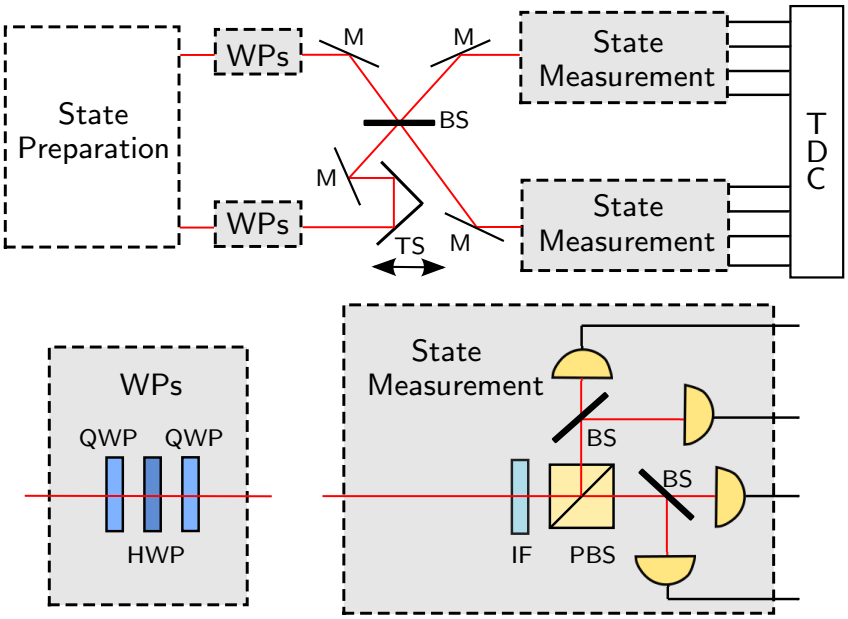

Figure 4.2: Experimental set-up. The polarization-entangled two-photon states are prepared by the method described in Ref. [DLT+13]. Upon emerging from the source, the two photons are guided with mirrors (M) to interfere at a 50:50 beam splitter (BS), with the temporal overlap controlled by a translation stage (TS). After passing through interference filters (IF), the photons are sorted by polarizing beam splitters (PBS) and registered by one of the photo-detectors, four on each side. The detector outputs are addressed to a time-to-digital converter (TDC), and coincidences between counts of any two detectors are recorded. Two sets of wave plates (WPs), each composed of a half-wave plate (HWP) and two quarter-wave plates (QWP), implement the polarization changes that correspond to the unitary operators of Table 4.1.

The experimental set-up is shown in Fig. 4.2. The rank-one states and rank-two states are produced as described in Refs. [KWW+99, DLT+13]. The rank-four states are produced by adding a controlled admixture of white noise to the singlet state by varying the coincidence time window of the detection electronics [LHLLK06]. For greater details, see Ref. [Len14].

Owing to unavoidable imperfections, the actual states emitted by the source are not the ideal $\rho_{\text{true}}^{(1,2,3)}$ stated above but full-rank approximations of them. For example, as reported in Ref. [DLT+13], fidelities above 97% are consistently achieved for the rank-two states, and the experimental rank-one and rank-four states are of similar quality [Len14].

To implement the most direct measurement of the witness bases, we make use of

a Hong-Ou-Mandel (HOM) [HOM87] interferometer. The signal and idler photons are directed to interfere at a 50:50 beam splitter (BS) and could either emerge on the same side of the BS (bunching) or emerge on different sides of the BS (anti-bunching). Due to this interference effect, the four eigenkets, that is, two product kets and two Bell kets, of the witness operator in Eq. (4.1) can be distinguished. If the input state were $|\text{HH}\rangle$, $|\text{VV}\rangle$, or $|\Psi^+\rangle$, the two photons will emerge on the same sides of the BS, whereas they will anti-bunch for the input state $|\Psi^-\rangle$. Two additional PBS could then be installed at the two output ports of the BS, see Fig. 4.3. The four eigenkets of the witness operator could then be measured due to the following distinctive features summarized in Table 4.2.

However, in order to obtain the different signatures of the four eigenkets, one must be able to distinguish between a single photon event and a two-photon event. The detectors in our lab is not able to meet this requirement. Hence, we installed four additional 50:50 BSs, that is, one at each output port of the two PBSs to further split the photons, see Fig. 4.2. Due to this further splitting, a total of eight detectors need to be used. The coincidences between the appropriate pairs of them enable us to distinguish and measure the relative frequencies of each of the four eigenkets; for more details, see Ref. [Len14].

To get good interference effect, the two photons need to be indistinguishable. If they are distinguishable even in principle, for example, having different polarizations, arrival times, or any other properties, the interference effect will be weak or disappear completely. Hence, we need to optimize the interferometer, carefully controlling the polarizations of the photons, and the temporal as well as spatial overlap

Eigenket	Photon counts (LH,LV,RH,RV)
$\|\text{HH}\rangle$	(2,0,0,0) or (0,0,2,0)
$\|\text{VV}\rangle$	(0,2,0,0) or (0,0,0,2)
$\|\Psi^+\rangle$	(1,1,0,0) or (0,0,1,1)
$\|\Psi^-\rangle$	(1,0,0,1) or (0,1,1,0)

Table 4.2: Signatures of the four eigenkets of the witness operator. For example, if the signal photon and idler photon were in the state $|\text{HH}\rangle$, then either the LH-detector, or the RH-detector will register two photon counts, with each registering photon counts half of the time, while all the other detectors will register no photon counts.

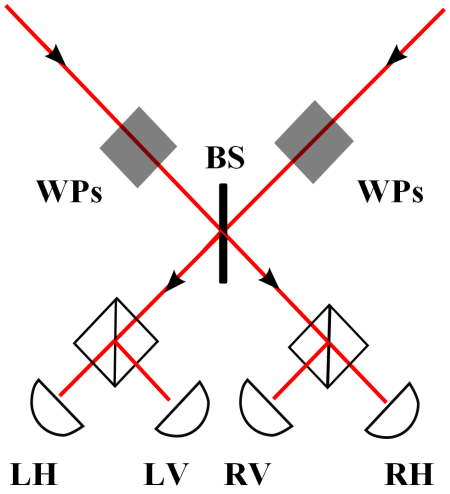

Figure 4.3: Realization of a witness-basis measurement using HOM interferences, with the signatures given in the Table 4.2. As an example, when the detectors at LH and LV ports both register photons simultaneously, this corresponds to a measurement signature for the $|\Psi^+\rangle$ eigenket. The wave plates WPs are used to change the witness family for subsequent measurements.

of the signal and idler photons. For this optimization, we first remove the other BSs in the output ports of the PBSs; see Fig. 4.2. Next, we address the outputs of the two detectors in the transmission ports of the two PBSs to a coincidence unit, where we record the coincidence rate for the input two-photon state $|\text{HH}\rangle$. A translation stage with a step size of 500 nm is used to control the temporal overlap between the photons, and the spatial overlap is controlled by adjusting the mirrors that direct the two beams to interfere on the HOM BS.

The interferometer is optimized at where the coincidence rate is minimal, as shown in Fig. 4.4. The visibility of the HOM dip is $95 \pm 3\%$. Imperfections in the BS ratios, the WPs, and the polarization controllers limit the maximum experimental achievable visibility of the dip. Nevertheless, the visibilities of the HOM dips obtained in our experiment exceed 90% for all the data collected for the construction of the histograms shown later in Fig. 4.5.

On the way from the source to the HOM BS, the photons pass through sets of wave plates (WPs) that change the polarization in accordance with one of the six local unitary transformations of Table 4.1. Each of the local unitary operators of

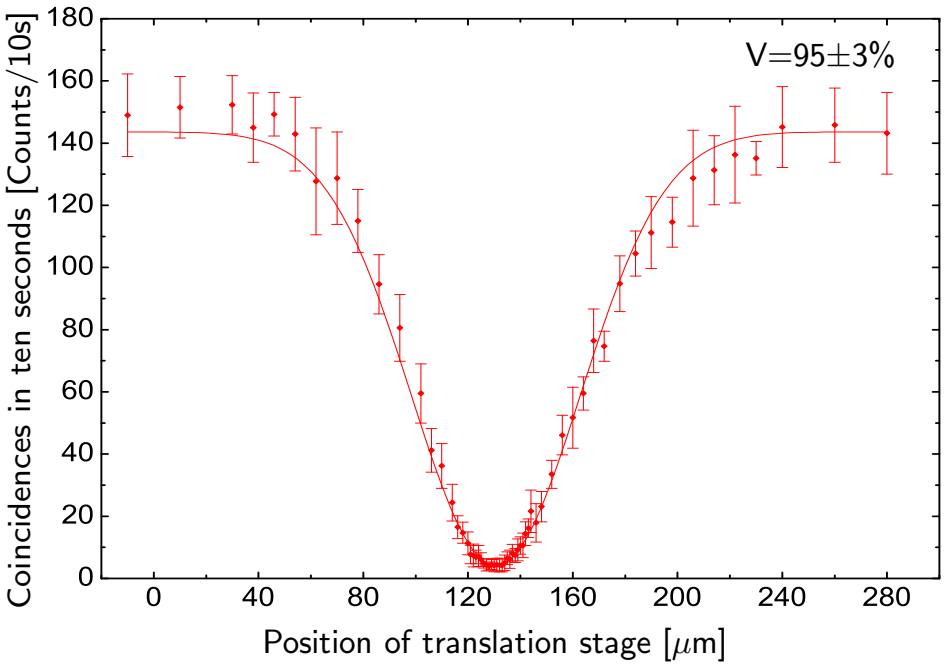

Figure 4.4: An example of a HOM dip obtained in our experiment for the state $|\text{HH}\rangle\langle\text{HH}|$. The visibility, V, of the HOM dip above is $95 \pm 3\%$; other HOM dips observed for different polarization states are similar to this one.

Table 4.1 is implemented by a HWP between two QWPs,

$$U = U_{\text{QWP}}(\alpha) U_{\text{HWP}}(\beta) U_{\text{QWP}}(\gamma)\,, \tag{4.8}$$

where the QWP with angle γ is the first in the sequence; see Fig. 4.2.

The matrix representation for a HWP, apart from an irrelevant global phase factor, is (see, e.g., [EKW01])

$$\begin{aligned} U_{\text{HWP}}(\theta) &= \begin{pmatrix} |\text{V}\rangle & |\text{H}\rangle \end{pmatrix} \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix} \begin{pmatrix} \langle\text{V}| \\ \langle\text{H}| \end{pmatrix} \\ &\widehat{=} \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}, \end{aligned} \tag{4.9}$$

where θ is the angle between its major axis and the vertical direction, and we have

$$U_{\mathrm{QWP}}(\theta) \mathrel{\widehat{=}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 - \mathrm{i}\cos(2\theta) & -\mathrm{i}\sin(2\theta) \\ -\mathrm{i}\sin(2\theta) & 1 + \mathrm{i}\cos(2\theta) \end{pmatrix} \tag{4.10}$$

for a QWP. Further, the matrices for the Pauli operator X that permutes $|\mathrm{V}\rangle$ and $|\mathrm{H}\rangle$ and the Clifford operator C that permutes the three Pauli operators cyclically are

$$X \mathrel{\widehat{=}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad C \mathrel{\widehat{=}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -\mathrm{i} \\ 1 & \mathrm{i} \end{pmatrix}, \tag{4.11}$$

respectively.

The angles α, β, and γ, for which the various Us are realized, are reported in Table 4.3. As an example, we consider

$$U = XC \mathrel{\widehat{=}} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\mathrm{i} \\ 1 & \mathrm{i} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \mathrm{i} \\ 1 & -\mathrm{i} \end{pmatrix} \tag{4.12}$$

and verify that $\alpha = 0$, $\beta = 0$, and $\gamma = -\pi/4$ are correct choices. Indeed, they are:

$$\begin{aligned} U_{\mathrm{QWP}}(0)U_{\mathrm{HWP}}(0)U_{\mathrm{QWP}}(-\pi/4) &\mathrel{\widehat{=}} \frac{1-\mathrm{i}}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix} \\ &= \mathrm{e}^{-\mathrm{i}\pi/4} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \mathrm{i} \\ 1 & -\mathrm{i} \end{pmatrix}, \end{aligned} \tag{4.13}$$

since the global phase factor is irrelevant.

U	α	β	γ
$\mathbf{1}$	0	0	0
X	0	$\pi/4$	0
C	0	$\pi/4$	$-\pi/4$
$C^\dagger$	$-\pi/4$	0	0
XC	0	0	$-\pi/4$
$XC^\dagger$	$\pi/4$	0	0

Table 4.3: Wave plate settings to realize each of the unitary operators of Table 4.1. The angles α, β, and γ are the settings of the QWP, HWP and QWP respectively, shown in Fig. 4.2, such that the corresponding U is obtained from Eq. (4.8).

After interfering at the HOM BS, in each output port of the interferometer, the photons pass through an interference filter with a central wavelength at 810 nm and a full width at half maximum of 10 nm and are then sorted by a polarizing beam splitter (PBS). To discriminate between one-photon and two-photon events, another 50:50 BS is installed into each output port of the PBS, and eight single-photon avalanche photodiodes detect the photons. A time-to-digital converter (TDC) records the arrival times of the photons, and coincidences between any two of the eight detectors are obtained from the analysis of the time stamp record of the TDC. For each rank, we studied 21 different states.

4.6 Results

We set the source to a particular state and performed witness-family measurements for one minute per family, and so measured about 10^4 photon pairs for each state and family. The data were analyzed for all three schemes.

For the three classes of states, only one of the witness families gives a conclusive result. Hence, for scheme A, the number of families needed for entanglement detection is equally likely to be one to six. The data for schemes B and C were analyzed as follows. The first family is chosen at random among the six families. If the measurement of this family gives a negative value of $\mathcal{S}$, the state is detected to be entangled and no further measurement is needed. If the result is inconclusive, then one uses the adaptive scheme to choose the next family, until a conclusive result is obtained. The result is n, the number of witness families that have to be measured in order to detect the entanglement; see Table 4.4 for illustrations.

The results are shown in Fig. 4.5, where one observes a significant improvement over the non-adaptive scheme A, which needs about 3.5 families on average. One rarely needs more than four witness families to detect the entanglement in these states; full tomography is never necessary since all rank-one, rank-two, and rank-four states are detected by one of the witness families.

We also performed simulations for these three specific classes of two-qubit states used in the experiment. For each class, data from 10^4 photon pairs (per witness

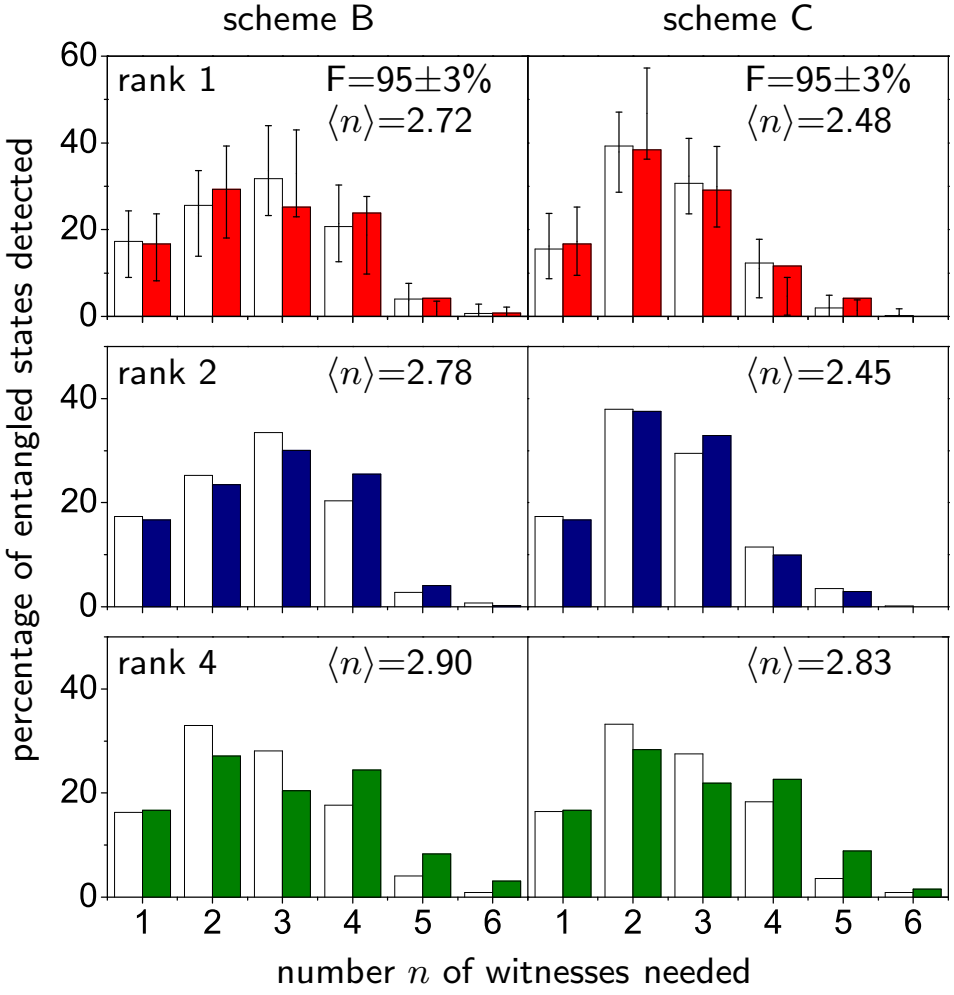

Figure 4.5: A comparison of schemes B (left column) and C (right column) for rank-one states (top row), rank-two states (middle row), and rank-four states (bottom row). The histograms report the percentage of entangled states detected against the number n of witness families needed *without* performing state estimation; $\langle n \rangle$ is the average value. Both the simulation data (left empty bars) and the experimental data (right full bars) show that, for the three kinds of quantum states considered, scheme C provides further improvement over scheme B: It requires fewer families on average and the distributions are narrower. The similarity of the two histograms for the rank-one states is confirmed by their large fidelity F; similar values are obtained for the other histograms. — Here, the simulation uses only states of the kind generated by the state preparation in the set-up of Fig. 4.2, whereas no such restriction applies to the randomly-chosen states for Fig. 4.1.

family) for 10^3 states were simulated with Monte Carlo techniques, for the schemes A, B, and C. As expected, the simulation showed that for measurements done with the random order of scheme A, the number of families needed for entanglement detection is distributed almost evenly from one to six, with a mean of about 3.5, and fewer families need to be measured in the adaptive schemes. The striking similarity between the results of the simulation and the experiment indicates that there are no significant systematic errors in the experimental data. The small differences are due to statistical fluctuations in the random choice of the first family and various the imperfections in the experiment as mentioned before. For example, the simulations for Fig. 4.5 use the ideal states, whereas the experiment has limited fidelities to the ideal states.

For the rank-one states, the fidelity $F = \sum_j (p_j p'_j)^{1/2}$ between two probability distributions $\{p_j\}$ and $\{p'_j\}$ compares the histograms from simulation and experiment, and the large value of F is reassuring. The error bars for the rank-one states are obtained by bootstrapping the actual data one hundred times; these error bars show the variation in the histograms that repeated measurements of this kind would display. The fidelity values and the error bars for the other histograms are of similar sizes and not displayed.

For the three classes of states, all entangled states are successfully detected

State	Families and their $\mathcal{S}$ values			n
$\rho^{(2)}_{\text{true}}$	family 3	family 4	family 2	3
$\mu = 0.15$	0.46 ± 0.03	0.06 ± 0.02	-0.23 ± 0.04	
$\rho^{(3)}_{\text{true}}$	family 1	—	—	1
$\lambda = 1$	-0.83 ± 0.03			
$\rho^{(1)}_{\text{true}}$	family 4	family 2	—	2
$\vartheta = \pi/4$	0.21 ± 0.03	-0.87 ± 0.03		

Table 4.4: Examples demonstrating how Fig. 4.5 is derived. The first family is chosen at random among the six families. If the measurement of this family gives a negative value of $\mathcal{S}$, then the state is detected to be entangled and no further measurement is necessary. However, if the result is inconclusive, then one uses the adaptive scheme to choose the next family, until a conclusive result is obtained. The figure of merit is n, the number of witness families that have to be measured in order to detect the entanglement.

without the need to perform full tomography. There are, however, other entangled states that would escape detection, and all separable states can only give inconclusive results. Regarding entangled states, we recall that about 2% of the random pure states and about 67% of the random mixed states are not detected by the six witness families without the separability check of scheme C. Hence, to confirm the efficiency and accuracy of the tomographic scheme, we also collected IC data even if the state is known to be entangled before all six witness-family measurements are done so as to check the reliability of the tomography. Using the technique of ML estimation [BDPS00, ŘHKL07], we then infer the state from the data and calculate the fidelity $\mathrm{tr}\{|\sqrt{\rho_{\mathrm{true}}}\sqrt{\rho_{\mathrm{est}}}|\}$ between the true and the estimated state. The average fidelities are $98.3 \pm 0.7\%$, $97.4 \pm 1.4\%$, and $98.7 \pm 1.1\%$ for the respective true states of rank one, two, and four. Indeed, if all the six family measurements fail to detect the entanglement in the state, one can use the tomographic information to reliably reconstruct the unknown input state and then determine its separability numerically.

4.7 Conclusions

We performed an experiment to verify the witness-family-based entanglement detection scheme introduced in Ref. [ZTE10]. In going beyond that proposal, we also introduced adaptive schemes that use the information acquired in previous measurements to reduce the average number of witness families that need to be measured. A few-witness way of detecting entanglement for photon-polarization qubits was thus demonstrated. Further, we showed that the witness-family measurements enable reliable quantum state tomography. With the necessary changes and within the limits set by what is experimentally feasible, the witness-family approach is also applicable to qubits of other physical kinds than photon polarization and to higher-dimensional or multi-partite systems.

4.8 Further Comments

4.8.1 The experiment by Barbieri *et al.*

Upon denoting the Pauli operators of the jth qubit by X_j, Y_j, and Z_j, the witness of Eq. (4.1) is

$$W^{(\alpha)} = \frac{1}{4}\big[1 + (Z_1 + Z_2)\cos\alpha + Z_1Z_2 + (X_1X_2 + Y_1Y_2)\sin\alpha\big]\,. \tag{4.14}$$

In the experiment of Barbieri *et al.* [BDMDN$^+$03] the two polarized photons are measured separately and, therefore, a direct measurement of $W^{(\alpha)}$ is not possible. Instead, the expectation value of $W^{(\pi/2)}$ is inferred from three auxiliary measurements that detect the common eigenstates (i) of X_1 and X_2, (ii) of Y_1 and Y_2, and (iii) of Z_1 and Z_2.

Together, the data acquired in these measurements establish the nine expectation values

$$\begin{aligned} &\langle X_1\rangle\,,\ \langle X_2\rangle\,,\ \langle X_1X_2\rangle\,,\\ &\langle Y_1\rangle\ ,\ \langle Y_2\rangle\ ,\ \langle Y_1Y_2\rangle\,,\\ &\langle Z_1\rangle\,,\ \langle Z_2\rangle\,,\ \langle Z_1Z_2\rangle\,, \end{aligned} \tag{4.15}$$

of which only the three values in the right column are used for $\big\langle W^{(\pi/2)}\big\rangle$. The other six values are not exploited. While this is wasteful, it does not matter in the context of Ref. [BDMDN$^+$03] where one knows beforehand that the unknown state is of the form

$$\rho = \frac{1}{4}\big[1 - p(X_1X_2 + Y_1Y_2 + Z_1Z_2)\big] \tag{4.16}$$

with $0 \le p \le 1$. The value of parameter p is then provided by $\big\langle W^{(\pi/2)}\big\rangle = (1-3p)/4$.

All witnesses in the first and the second family of Table 4.1 are made available by the five expectation values in the bottom row and the right column in (4.15). It follows that the inequality

$$\begin{aligned} 1 + \langle Z_1Z_2\rangle^2 \ \ge\ & 2\big|\langle X_1X_2\rangle\langle Y_1Y_2\rangle - \langle Z_1Z_2\rangle + \langle Z_1\rangle\langle Z_2\rangle\big| \\ & + \langle X_1X_2\rangle^2 + \langle Y_1Y_2\rangle^2 + \langle Z_1\rangle^2 + \langle Z_2\rangle^2 \end{aligned} \tag{4.17}$$

holds for all separable states. There are two more inequalities of the same structure for the expectation values of the first and the second row, respectively, and the third column in (4.15).

If all three inequalities are obeyed, further measurements are needed. They would detect the common eigenstates of X_1 and Y_2, of Y_1 and X_2, and so forth. Each such measurement gives the expectation values of two single-qubit operators already contained in (4.15) and adds one new two-qubit expectation value to the list. More inequalities analogous to (4.17) become available in the course. Eventually, when all six two-qubit expectation values that are missing in (4.15) are determined, full tomography is achieved.

Clearly, the sequence in which the fourth, fifth, ... ninth measurements are carried out, can be optimized by an adaptive strategy. There is also the option of checking, at each stage, whether there are separable states in the convex set of maximum-likelihood estimators, and inferring that the unknown state is entangled if there are none.

In summary, the measurement scheme of Barbieri *et al.* can be used for an indirect measurement of witness families. When state reconstruction is necessary, this indirect measurement needs nine settings, each providing three expectation values of the fifteen independent ones; each of the six single-qubit expectation values is determined thrice. By contrast, the direct measurement achieves full tomography with six settings, whereby three of the nine two-qubit expectation values are determined twice (see Table II in Ref. [ZTE10]). Although one could conclude that the direct measurement is less wasteful and should be preferred over the indirect measurement, one must remember that the direct measurement is not feasible when the two polarized photons are at different locations. Then, the Barbieri *et al.* scheme does the job.

4.8.2 The nonlinear witnesses of Gühne and Lütkenhaus

The inequality $\langle W^{(\alpha)} \rangle \geq 0$ holds for all separable states; this is, of course, the witness property. In Ref. [GL06], Gühne and Lütkenhaus show that the null bound

can be replaced by various state-dependent positive bounds,

$$\left\langle W^{(\alpha)}\right\rangle \geq \left|\left\langle G^{(\alpha)}\right\rangle\right|^2, \tag{4.18}$$

where, for example, the non-Hermitian operators

$$\begin{aligned} G_1^{(\alpha)} = {} & \frac{1}{\sqrt{8}}\sin\Big(\frac{\alpha}{2}+\frac{\pi}{4}\Big)(1+X_1X_2+Y_1Y_2+Z_1Z_2) \\ & + \frac{1}{\sqrt{8}}\cos\Big(\frac{\alpha}{2}+\frac{\pi}{4}\Big)(Z_1+Z_2+\mathrm{i}X_1Y_2-\mathrm{i}Y_1X_2) \end{aligned} \tag{4.19}$$

or

$$\begin{aligned} G_2^{(\alpha)} = {} & \frac{1}{\sqrt{8}}\sin\Big(\frac{\alpha}{2}+\frac{\pi}{4}\Big)(X_1+X_2+\mathrm{i}Y_1Z_2-\mathrm{i}Z_1Y_2) \\ & + \frac{1}{\sqrt{8}}\cos\Big(\frac{\alpha}{2}+\frac{\pi}{4}\Big)(X_1Z_2+Z_1X_2+\mathrm{i}Y_1-\mathrm{i}Y_2) \end{aligned} \tag{4.20}$$

are possible choices for $G^{(\alpha)}$. The lower bounds in Eq. (4.18) are second-degree polynomials of one-qubit and two-qubit expectation values; there are also lower bounds that are polynomials of fourth or higher order.

We can evaluate $\left\langle G_1^{(\alpha)}\right\rangle$ as soon as the three witness families 1, 3 (or 4), and 5 (or 6) have been measured; similarly, $\left\langle G_2^{(\alpha)}\right\rangle$ is available after measuring the four families 3–6. This illustrates that partial tomography is needed before the more stringent lower bounds of Eq. (4.18) are at hand.

Nevertheless, it could be interesting to exploit criteria of this kind (with parameter α optimized) for a possible further reduction of the number of witness families that need to be measured before one can conclude that the unknown state is entangled. This is unexplored territory.

4.8.3 General adaptive schemes

The adaptive schemes B and C that speed up the entanglement detection select the next witness family from the six pre-chosen families specified by the unitary operators in Table 4.1. If, instead, one selects also from other families than the six pre-chosen ones, such a more general adaptive scheme might be more efficient, in the

sense that fewer families need to be measured on average before one can conclude that the unknown state is entangled. Thereby, the selection is still done by opting for the family which is expected to give the smallest value of $\mathcal{S}$ upon measurement. It turns out that the more general adaptive schemes are not worth the trouble. We justify this remark by a study of the generalizations of schemes B and C.

Scheme B'

As in scheme B, we calculate the MLME estimator ρ_{MLME} and exploit its properties when choosing the next family to be measured. If ρ_{MLME} is entangled, $\rho_{\mathrm{MLME}}^{\mathrm{T}_2}$ has one negative eigenvalue, and the eigenket $|\phi\rangle$ to this eigenvalue is entangled. The next witness family is then the one obtained from $W = \left(|\phi\rangle\langle\phi|\right)^{\mathrm{T}_2}$ because this family is best for detecting the entanglement of ρ_{MLME}. Since ρ_{MLME} is the current best guess for the unknown state ρ_{true}, this family has also a good chance of detecting entanglement in ρ_{true}. If the MLME estimator is separable, however, we proceed as in scheme B.

Scheme C'

On top of scheme B', the separability check of scheme C is implemented.

Simulations

We investigate the general adaptive schemes by performing computer simulations for pure and full-rank states and constructing histograms analogous to those in Fig. 4.1. As can be seen from the results in Fig. 4.6, scheme B' improves over scheme B. For both rank-one and rank-four states, the cumulative percentage of states detected is higher. The percentage of undetected rank-four states after six families drops from about 67% to about 25%. While this improvement is substantial, it pales in comparison with the dramatic reduction to about 2.7% when using scheme C. Therefore, when aiming at the most efficient way of detecting entanglement, we have to employ either scheme C or scheme C'.

Now, as we learn from Fig. 4.6, scheme C' is only slightly better than scheme C — if at all. For, the somewhat smaller proportion of rank-four states detected

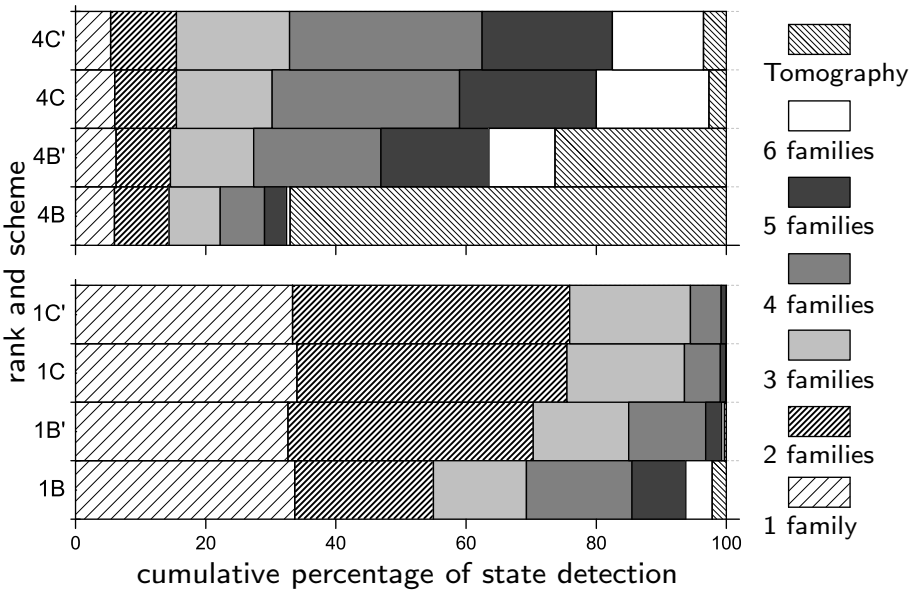

Figure 4.6: Simulation results for 10^4 randomly chosen two-qubit entangled states: pure states (bottom) and full-rank mixed states (top). The cumulative histograms compare between adaptive measurements performed with the six pre-chosen families of Table 4.1 (schemes B and C) and with six arbitrary families (schemes B' and C').

by the first, randomly chosen, family and the somewhat larger proportion of states requiring tomography are surely resulting from statistical fluctuations in the simulation. The other differences between the histograms for schemes C and C' are of similar size. Accordingly, there is no evidence that scheme C' is worth the trouble of its implementation, which requires that the settings of the WPs are calculated and updated in every step for every state. On the other hand, the data obtained in scheme C are optimal for the tomographic reconstruction of ρ_{true}. In summary, then, scheme C serves all purposes very well.

Chapter 5

Quantum State Tompgraphy with Additional Unknown Parameters

The majority of the quantum state tomography experiments performed so far assume that the quantum state to be estimated is the only unknown, while other parameters necessary to reconstruct the state are all perfectly known. However, in a typical experiment such as the ones discussed in the previous chapters, additional parameters apart from the state, for example the efficiency of the detectors, are also unknown. In this chapter, we study quantum state tomography with additional unknown parameters and illustrate the construction of optimal error regions with some examples.

5.1 Introduction

As mentioned previously, quantum information is a promising field that utilizes quantum systems to perform sophisticated tasks of computations and communications. Inevitably, characterization of a source of quantum information carriers, verification of the properties of a quantum channel, and so on, are of crucial importance to quantum information processing tasks. The efforts in providing reliable answers to such challenges culminate in the field of quantum state estimation [Pv04].

In quantum state estimation, assumptions have usually been made to simplify the real problem. Currently, most of the quantum state tomography experiments assume that the quantum state to be estimated is the only unknown, while other parameters necessary to reconstruct the state are all perfectly known. Such parameters could be needed to specify the devices in the measurement, (e.g. the efficiency of the detectors) or the type of unitary transformation the quantum state undergoes during the experiment (e.g. the splitting ratio of the beam splitter, the settings of the wave plates), or perhaps needed to completely characterize the source (e.g. the mean number of photons emitted out by the source per second). For example, one of the many simplification is that the detection of the quantum systems can be normalized to as if the detections were perfect. In performing the Pauli measurements to determine the polarization state of a photon, one first calibrates the relative efficiencies of the detectors, and then normalizes the detector counts, such that one would obtain data that corresponds to perfect detection. Similar normalization technique is also used in estimating the phase of a Gaussian state, where the detectors are purposely made to have equal efficiency (usually by lowering the efficiency of one detector). While such normalization techniques are legit, they are not always possible. For example, one might not be able to obtain a reliable relative efficiencies of the detectors, due to the set-up of the real detectors in the experiment.

In a more realistic setting, we need to consider the situation for which there are additional unknown parameters. Typically, these various parameters are normally calibrated before the experiment (if that is possible). This calibration provides the experimenter with a prior knowledge of these parameter values. In the process of carrying out the experiment, the data not only tell us new knowledge about the state, but also the additional parameters. There exist proposals to reconstruct the quantum state, and calibrate the measurement device, simultaneously—the so-called "self-calibrating" technique [MŘH12]. The essence of this technique is to estimate the state and other additional parameters in the measurement set-up by maximizing the likelihood of obtaining the data, given a (set of) state parameter(s) as well as any additional parameters relevant in the experiment. Various demonstrations of the technique have been presented [QBJ13, SIK$^+$13]. However, they lack error regions to

make them statistically meaningful. Similar to quantum state estimation, one should also supplement the point estimators for these additional parameters with error regions. To tackle the issue of reporting error regions, various methods have been proposed, but they usually rely on having a lot of data , involve data resampling, or consider all data that one might have observed. There are also numerical coarse-graining scheme to reliably estimate the quantum state of a source given a set of noisy measurement outcomes that are not well calibrated [TŘH14].

In this chapter, we would like to consider a more realistic quantum state estimation experiment, in which various additional parameters could be unknown. We get rid of the assumption that the number of copies measured in the experiment is large so that it suffices to report point estimators and would like to report meaningful error regions if just a few copies are measured. We would also like to construct error regions systematically from only the data that are actually observed. We show that the machinery developed in Ref. [SNS$^+$13] can indeed by efficiently applied to various realistic quantum state tomography experiments with additional unknown parameters.

In the construction, one faces the subtlety that the the state parameters (e.g. the Bloch vector of a qubit) and other parameters (e.g. the number of undetected copies from the inefficient detectors) are of different nature. The optimal error region is therefore not a region in the state space, but rather, it comprises of direct sum of the state space and the parameter space. The experimenter in the end might be interested in the value and errors of the state estimator and the different parameters individually. We demonstrate how to separate the different estimators from their joint optimal regions, and attaching error regions (intervals) to them separately. By marginalizing over the nuisance parameters, one obtains a marginal likelihood which only depends on the parameter of interest. The marginal likelihood characterizes one's evidence that he or she has about the parameter of interest. The bounded-likelihood intervals obtained from the marginal likelihood are the optimal error intervals as they are the smallest credible intervals.

Here is a brief outline of this chapter. We set the stage and fix the notations in Section 5.2. We also explain how to report the error interval for the parameter

that one is interested in, by marginalizing over the nuisance parameters. We then illustrate the construction with a simple example on polarization tomography in Section 5.3. In Section 5.4, we focus on quantum phase estimation, and demonstrate how to deal with cases where there are multiple peaks and multiple regions in the parameter space. Finally we close this chapter with discussion and conclusion.

5.2 Setting the stage

To set the stage, we first introduce the general scenario and set the notations used in this paper. In a typical setting, there is an unknown state to be reconstructed based on the data acquired. Several independently and identically prepared copies of the state are measured by an apparatus that realizes a POM. The experimenter's lack of knowledge about the state, $\rho(\boldsymbol{s})$, can be captured by $\boldsymbol{s} = \{s_1, s_2, \cdots, s_n\}$, where n is the number of parameters needed to fully specify the states. It is straightforward to generalize to the preparation of a set of states $\{\rho_k(\boldsymbol{s})\}$. Any other additional parameters that are needed to reconstruct the states are represented by $\boldsymbol{\gamma}$=$\{\gamma_1, \gamma_2, \cdots, \gamma_m\}$, where m is the number of additional parameters in the experiment.

Our aim is now to reconstruct the state, $\rho(\boldsymbol{s})$, as well as estimate the unknown parameters, $\boldsymbol{\gamma}$, simultaneously. By that, we mean we would like to report a region in the space of the state and the parameters that contains the true values with high probability, yet small in size. A notion of size is taken for granted here [SNS$^+$13]. Prior to taking any data, it is most natural to assign equal probabilities to equivalent alternatives. For example, if we split the state space equally and fairly into two, then it is equally likely that any state will be contained in either of the two pieces. What does it mean by splitting equally and fairly? We say that it is fair provided the size of the two pieces is the same. Hence, here we take it seriously that the size of a region is quantified by its prior probability. Note that the above reasoning can also be reversed. If we have already established the prior probability of the regions with some other arguments, perhaps by symmetry or other considerations [SNS$^+$13], then we take it as the natural measure of the size of the region [EGS06].

We denote by $(\mathrm{d}\rho)$ the size of the infinitesimal vicinity of the state ρ, and sim-

ilarly by $(\mathrm{d}\boldsymbol{\gamma})$ that in the infinitesimal vicinity of the additional parameters. Here, the calibration of the additional unknown parameters provide us with the prior ("size $\equiv$ prior probability")

$$(\mathrm{d}\boldsymbol{\gamma}) = \mathrm{d}\gamma_1 \mathrm{d}\gamma_2 \cdots \mathrm{d}\gamma_m w(\gamma_1, \gamma_2 \cdots \gamma_m), \tag{5.1}$$

where $w(\gamma_1, \gamma_2 \cdots \gamma_m)$ is the prior density for the parameters $\boldsymbol{\gamma}$. In most cases, these additional parameters are uncorrelated and the prior density is a product of prior densities for each of the individual parameters, that is

$$w(\gamma_1, \gamma_2 \cdots \gamma_m) = w(\gamma_1) w(\gamma_2) \cdots w(\gamma_m). \tag{5.2}$$

Together with our knowledge about the source which provides the quantum state, it enables us to compute the size $S_{\mathcal{R}}$ of the region $\mathcal{R}$,

$$S_{\mathcal{R}} = \int_{\mathcal{R}} (\mathrm{d}\rho)(\mathrm{d}\boldsymbol{\gamma}) \quad \text{with} \int_{\mathcal{R}_0} (\mathrm{d}\rho)(\mathrm{d}\boldsymbol{\gamma}) = 1. \tag{5.3}$$

Here, one could of course simplify the notation by grouping the additional parameters $\boldsymbol{\gamma}$ with the state parameters and writing only $(\mathrm{d}\rho)$ for both. However we choose to write $(\mathrm{d}\rho)(\mathrm{d}\boldsymbol{\gamma})$ just for the sake of clarity so that we can make reference to the state parameters and any other additional parameters easily. $\mathcal{R}_0$ is the reconstruction space on which the size is normalized to unity.

In the process of carrying out the experiment, new data are obtained. From them, we update our knowledge about the various parameters according to

$$\text{posterior} \propto \text{prior} \times \text{likelihood}, \tag{5.4}$$

where the likelihood here is $L(D|\rho, \boldsymbol{\gamma})$, a likelihood that is conditioned on both ρ and $\boldsymbol{\gamma}$. The posterior is then a property of the state and additional parameters as a whole. Just like the prior probability tells us about the size of the region, the posterior probability here tells us about the credibility of the region.

The optimal error regions can now be constructed. Following Ref. [SNS$^+$13], we find the bounded-likelihood region (BLR), $\mathcal{R}_\lambda$, for different threshold values of λ.

The size of the region is given by

$$s_\lambda = \int_{\mathcal{R}_0} (\mathrm{d}\rho)(\mathrm{d}\boldsymbol{\gamma})\chi_\lambda(\rho,\boldsymbol{\gamma}), \tag{5.5}$$

where

$$\chi_\lambda(\rho,\boldsymbol{\gamma}) = \theta\left(L(D|\rho,\boldsymbol{\gamma}) - \lambda L(D|\hat{\rho}^{\mathrm{ML}},\hat{\boldsymbol{\gamma}}^{\mathrm{ML}})\right). \tag{5.6}$$

Here $\theta(\ )$ denotes Heaviside's unit step function. The BLR is the maximum likelihood region (MLR), i.e., the region for which the given data are more likely than for any other region with the same pre-chosen size.

The BLR $\mathcal{R}_\lambda$ has credibility

$$c_\lambda = \frac{1}{L(D)} \int_{\mathcal{R}_0} (\mathrm{d}\rho)(\mathrm{d}\boldsymbol{\gamma})\chi_\lambda(\rho,\boldsymbol{\gamma})L(D|\rho,\boldsymbol{\gamma}). \tag{5.7}$$

The BLR is also the smallest credible region (SCR), i.e., the smallest region with pre-chosen credibility, where credibility of a region is the probability of finding actual state and parameter in that region, conditioned on the data. Once s_λ is known as a function of λ, we can also obtain c_λ from

$$c_\lambda = \frac{\lambda s_\lambda + \int_\lambda^1 s_{\lambda'}\mathrm{d}\lambda'}{\int_0^1 s_{\lambda'}\mathrm{d}\lambda'}. \tag{5.8}$$

The error regions will be reported by communicating s_λ and c_λ as functions of λ. The experimenter interested in the MLR with the size s of his or her liking, or the SCR of his or her desired credibility c can determine the required values λ. It is then easy to check if a state is inside the specified error region.

At this stage, notice that the SCR is a region containing both the state parameter $\boldsymbol{s}$, as well as additional parameters $\boldsymbol{\gamma}$. It is a property of all of them as a whole. The experimenter in the end might be interested in reporting the state estimator and estimators for the other parameters separately, with their own error regions. Suppose that the experimenter is interested in the parameter γ_i, he or she can first integrate over the nuisance parameters (anything else other than γ_i) to obtain the

marginal likelihood

$$L(D|\gamma_i) = \int_{\mathcal{R}_0} (\mathrm{d}\rho) \prod_{j\neq i} \mathrm{d}\gamma_j w(\boldsymbol{\gamma}) L(D|\rho, \boldsymbol{\gamma}). \tag{5.9}$$

With this marginal likelihood at hand, one can then proceed with determining the ML estimator for the parameter γ_i, as well as BLRs. The BLRs will be the smallest credible region, or rather smallest credible interval for the one-dimensional parameter γ_i. If the experimenter is interested in a higher-dimensional state, he or she can construct the marginal likelihood and optimal error regions analogously.

Finally, just as a comment, the MLR and the SCR are unique for the data observed. Once the data are obtained, there is the MLR and SCR, whereas the MLRs and SCRs associated with unobserved data play no role at all. This is very different from the confidence regions, where one needs to consider all possible data that were not obtained, but could possibly be obtained. Nonetheless, they are not totally unrelated. For discussions on how high-credibility regions offer starting points of construction of confidence regions, see Ref. [CR12].

5.3 Polarization measurement with imperfect detectors

To demonstrate our method, we start with a simple, yet very illustrative example of the polarization measurement on a single qubit. The state of the source is described by a statistical operator ρ. For the moment, suppose only the expectation value of σ_z is of interest, then a projective measurement in the z-basis is sufficient to determine $z = \langle \sigma_z \rangle$.

In a typical and also more realistic situation in the lab, the detector efficiencies will not be unity and it is inevitable that some photons will escape detection. The POM describing the above situation is then $\{\Pi_1, \Pi_2, \Pi_0\}$. Here, Π_1 describes the detection in the transmission port of PBS, given by

$$\Pi_1 = \eta_1 \frac{\mathbb{1} + \sigma_z}{2}, \tag{5.10}$$

where η_1 is the efficiency of the detector in this port. Similarly, Π_2 describes detec-

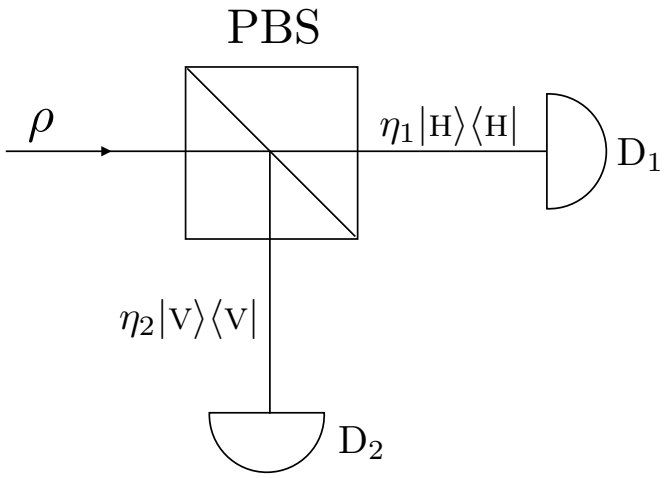

Figure 5.1: Polarization measurement on a single qubit: An unknown state ρ is sent to a polarizing beam splitter (PBS) where only the expectation value of σ_z is of interest. In a typical and also more realistic situation in the lab, the detector efficiencies will not be unity and it is inevitable that some photons will escape detection. The detector D_1 with quantum efficiency η_1 realizes the POM element $\eta_1|\mathrm{H}\rangle\langle\mathrm{H}|$, and The detector D_2 with quantum efficiency η_2 realizes the POM element $\eta_2|\mathrm{V}\rangle\langle\mathrm{V}|$. The POM element Π_0 for the missing counts is not drawn, see text for more details.

tion in the reflection port, which is

$$\Pi_2 = \eta_2 \frac{\mathbb{1} - \sigma_z}{2}, \tag{5.11}$$

and $\Pi_0 = \mathbb{1} - \Pi_1 - \Pi_2$ accounts for the missing copies. For simplicity, we assume that the efficiencies of both detectors are known.

The probability that the emitted photon is horizontally polarized (i.e. detector 1 clicks) is given by

$$p_1 = \mathrm{tr}\{\rho\Pi_1\} = \frac{\eta_1(1+z)}{2}. \tag{5.12}$$

Similarly,

$$p_2 = \mathrm{tr}\{\rho\Pi_2\} = \frac{\eta_2(1-z)}{2}, \tag{5.13}$$

and $p_0 = 1 - p_1 - p_2$. In typical experiments, the data consist of the observed number of clicks for detector 1 and 2 respectively, denoted by n_1 and n_2. However,

the number of missing copies, n_0, is generally unknown, except that one does have a prior knowledge for the n_0, as a result of one's prior knowledge for the total number of measured copies, $N = n_1 + n_2 + n_0$. In this case, one *a priori* expects that N follows the usual Poissonian photon statistics, i.e.

$$P(N) = \frac{(\nu T)^N}{N!} \mathrm{e}^{-\nu T} \ , \ N = 0,\ 1,\ 2 \ldots , \tag{5.14}$$

where the average photon count rate, ν, and the duration of data collection, T, are known.

We can then incorporate the additional parameter, which is n_0 here, into the likelihood function. The likelihood of obtaining a particular set of data, D, given z and n_0, is

$$L(D|z, n_0) = p_1^{n_1} p_2^{n_2} p_0^{n_0} \binom{N}{n_0}, \tag{5.15}$$

where the binomial factor comes from the different possibilities of missing n_0 copies from a total of N copies. Using MLE, upon maximizing the likelihood with respect to z, we get

$$\begin{aligned} \hat{z}^{\mathrm{ML}} =& \frac{n_1(1-\eta_2) + n_2(1-\eta_1)}{(\eta_1 - \eta_2)(n_0 + n_1 + n_2)} \pm \\ & \frac{\sqrt{(\eta_2 n_0 + n_1 - \eta_1(n_0 + n_1) - n_2(1-\eta_2))^2 + 4 n_1 n_2 (1-\eta_2)(1-\eta_1)}}{(\eta_1 - \eta_2)(n_0 + n_1 + n_2)}. \end{aligned} \tag{5.16}$$

The $\pm$ sign of $\hat{z}^{\mathrm{ML}}$ is determined by the constraint $|z| \leq 1$.

To obtain the ML estimator for n_0, we cannot differentiate L with respect to n_0 since n_0 only takes integral values. Instead, suppose we have found $\hat{n}_0^{\mathrm{ML}}$ and $\hat{z}^{\mathrm{ML}}$, such that $L_{\max} = L(D|\hat{z}^{\mathrm{ML}}, \hat{n}_0^{\mathrm{ML}})$ is maximum. When $n_0 = \hat{n}_0^{\mathrm{ML}} + 1$,

$$L_{\max} > L(D|\hat{z}^{\mathrm{ML}}, \hat{n}_0^{\mathrm{ML}} + 1). \tag{5.17}$$

Similarly, when $n_0 = \hat{n}_0^{\mathrm{ML}} - 1$,

$$L_{\max} > L(D|\hat{z}^{\mathrm{ML}}, \hat{n}_0^{\mathrm{ML}} - 1). \tag{5.18}$$

From these two inequalities, one gets

$$\hat{n}_0^{\mathrm{ML}} = \mathrm{floor}\left((n_1+n_2)\frac{1-p_1-p_2}{p_1+p_2}\right). \tag{5.19}$$

In summary, given a set of data, we have to find $\hat{z}^{\mathrm{ML}}$ and $\hat{n}_0^{\mathrm{ML}}$ that satisfy both Eqs. (5.16) and (5.19) as the ML estimators for the two unknowns respectively.

The point estimator that we obtained above should be supplemented by an error region. To construct the region, we use Eq. (5.3), where in this context, there is only one additional parameter, i.e. $\boldsymbol{\gamma} = \{n_0\}$. Since it is discrete, the integral would mean a summation in n_0 with the Poissonian prior that we have about this parameter. Hence, we have,

$$\mathcal{S}_{\mathcal{R}} = \int_{\mathcal{R}} (\mathrm{d}\rho) \sum_{n_0=0}^{\infty} \frac{(\nu T)^{n_1+n_2+n_0}}{(n_1+n_2+n_0)!} \mathrm{e}^{-\nu T}. \tag{5.20}$$

For this three-element POM, i.e. $(\mathrm{d}p) = \mathrm{d}p_1\mathrm{d}p_2\mathrm{d}p_0$, the permissible probabilities, assuming without loss of generality that $\eta_1 \geq \eta_2$, are identified by

$$\begin{aligned} w_{\mathrm{cstr}}(p) \doteq & \theta(p_1)\theta(p_2)\theta(p_0-(1-\eta_1)) \\ & \times \delta\left(\frac{p_1}{\eta_1}+\frac{p_2}{\eta_2}-1\right)\delta(p_1+p_2+p_0-1), \end{aligned} \tag{5.21}$$

where $\delta(\ )$ is Dirac's delta function. The dotted equal sign stands for 'equal up to a multiplicative constant', namely the factor that ensures the unit size of the reconstruction space. For the primitive prior, we then have

$$(\mathrm{d}\rho) = \mathrm{d}p_1\mathrm{d}p_2\mathrm{d}p_0 w_{\mathrm{cstr}}(p) = \frac{\eta_1\eta_2}{2}\mathrm{d}z, \tag{5.22}$$

so that the multiplicative constant in Eq. (5.21) is given by $\frac{1}{\eta_1\eta_2}$ to ensure the correct normalization, since then we have

$$\int_{\mathcal{R}_0} (\mathrm{d}\rho) = \int \mathrm{d}p_1\mathrm{d}p_2\mathrm{d}p_0 \frac{1}{\eta_1\eta_2} w_{\mathrm{cstr}}(p) = \int_{-1}^{1} \frac{\mathrm{d}z}{2} = 1. \tag{5.23}$$

We see here that the primitive prior in the probabilities turns out to be also the

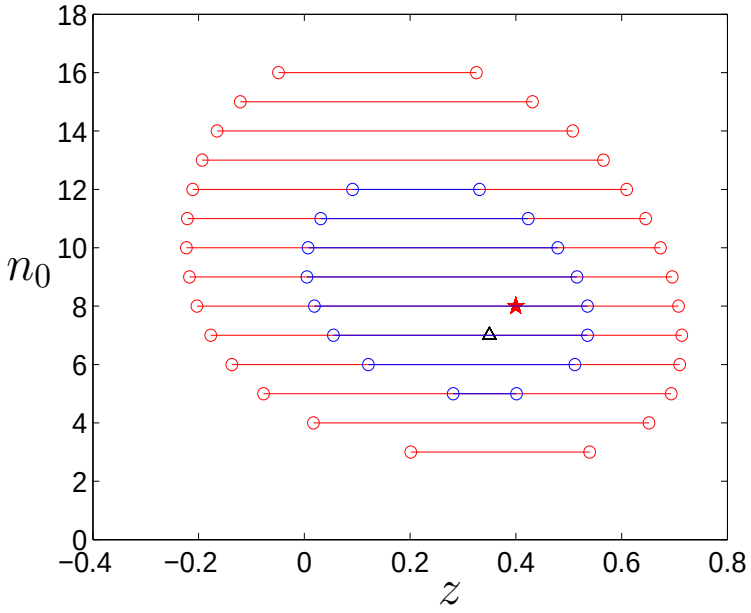

Figure 5.2: Graphs with $(n_1, n_2, \eta_1, \eta_2) = (10, 4, 0.7, 0.5)$. The red star ($\star$) is the true state that is used for the simulation. The black triangle ($\triangle$) is the ML estimator. The collection of the red lines form the SCR for this set of data for $c_\lambda = 0.9$, and the blue lines are for $c_\lambda = 0.5$.

primitive prior in z. Collecting the pieces, finally we have

$$\mathcal{S}_\mathcal{R} = \int_\mathcal{R} \frac{\mathrm{d}z}{2} \sum_{n_0=0}^{\infty} \frac{(\nu T)^{n_1+n_2+n_0}}{(n_1+n_2+n_0)!} \mathrm{e}^{-\nu T}. \tag{5.24}$$

The size of the bounded-likelihood region for z and n_0 as a whole can then be found using

$$s_\lambda = \int_{R_0} (\mathrm{d}\rho) \sum_{n_0=0}^{\infty} p(n_0) \chi_\lambda(z, n_0), \tag{5.25}$$

where

$$\chi_\lambda(z, n_0) = \theta\left(L(n_1, n_2|z, n_0) - \lambda L(n_1, n_2|\hat{z}^{\mathrm{ML}}, \hat{n}_0^{\mathrm{ML}})\right). \tag{5.26}$$

Once the size is determined for different values of λ, the credibility can then be evaluated according to Eq. (5.8).

Figure 5.2 shows SCRs obtained for a simulated experiment in which $(n_1, n_2, \eta_1, \eta_2) = (10, 4, 0.7, 0.5)$, for two different values of c_λ. In the simulation, the true state is $z = 0.4$, and a total of 21 copies are used which resulted from a Poissoan distribution of $\nu = 9.44$ photons per second and $T = 2$ seconds. The actual $\nu = 9.44$ is

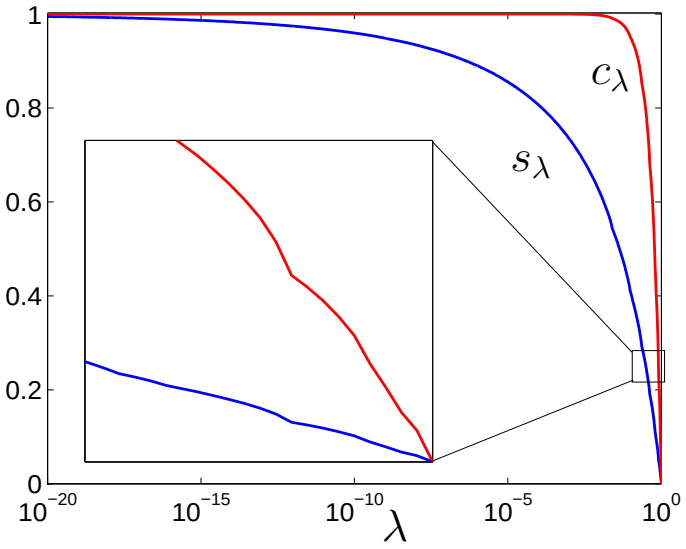

Figure 5.3: Size (the blue curve) and credibility (the red curve) as functions of λ for the primitive prior for the regions in Fig. 5.2. The experimenter interested in the SCR of his desired credibility c can determine the required value of λ and check if a given value of z and n_0 is in the region. The kinks (which are barely noticeable but shown in the inset) in the graph are due to the discrete nature of the parameter n_0.

based on a prior on ν with known mean $\bar{\nu}$ and standard deviation δ_ν (more about this in the following part). The number of missing copies is eight in the simulation. The ML estimator for z and n_0 is $\hat{z}^{\mathrm{ML}} = 0.3506$ and $\hat{n}_0^{\mathrm{ML}} = 7$ respectively. Figure 5.3 shows the size and credibility of the MLRs and SCRs as functions of λ. As λ increases, the size of both MLR and SCR decreases. Notice that there are kinks in the graph. They are due to the discrete nature of the parameter n_0.

The s_λ and c_λ is a property of the region of both z and n_0 together. The experimenter may wish to separate these two parameters and attach error regions to them individually. To do this, one first obtains the *marginal likelihood* for z only, given by

$$\begin{aligned} L(D|z) &= \sum_{n_0=0}^{\infty} L(D|z,n_0)p(n_0) \\ &= \sum_{n_0=0}^{\infty} L(D|z,n_0)\frac{(\nu T)^{n_1+n_2+n_0}}{(n_1+n_2+n_0)!}\mathrm{e}^{-\nu T} \\ &= p_1^{n_1}p_2^{n_2}\frac{(\nu T)^{n_1+n_2}}{(n_1+n_2)!}\mathrm{e}^{-\nu T(p_1+p_2)}. \end{aligned} \tag{5.27}$$

More strictly speaking, the likelihood in Eq. (5.27) is a likelihood conditioned on z as well as ν. In a realistic setting, the experimenter normally calibrates the source so that he or she knows the average value and spread of ν, denoted by $\bar{\nu}$ and δ_ν. This knowledge could be captured as a prior in ν, for example

$$p(\nu, \nu + \mathrm{d}\nu) = \mathrm{d}\nu \beta^{\alpha+1} \frac{\nu^\alpha}{\alpha!} \mathrm{e}^{-\beta\nu}, \quad \alpha, \beta > 0, \tag{5.28}$$

with

$$\alpha = \left(\frac{\bar{\nu}}{\delta_\nu}\right)^2 - 1, \tag{5.29}$$

and

$$\beta = \frac{\bar{\nu}}{(\delta_\nu)^2}. \tag{5.30}$$

Here, the widely used Gaussian distribution should be avoided since it has a tail in the negative region, but ν is a positive quantity. Instead, a Gamma distribution is more appropriate. It is commonly used as a conjugate prior distribution for various types of rate parameters, such as the rate of a Poisson distribution (which is the case here). The Gamma distribution also has the advantage that it is the maximum entropy probability distribution for a random variable X for which $E[X]$ is fixed and greater than zero, and $E[\ln(X)]$ is fixed. It follows that

$$\begin{aligned} p(N|\bar{\nu}, \delta_\nu) &= \int_0^\infty p(\nu, \nu + \mathrm{d}\nu) \frac{(\nu T)^N}{N!} \mathrm{e}^{-\nu} \\ &= \frac{(\alpha + N)!}{\alpha! N!} \beta^{\alpha+1} (\beta + 1)^{-\alpha-N-1}, \end{aligned} \tag{5.31}$$

and the new marginal prior for z now is

$$\begin{aligned} L(D|z) &= \sum_{n_0=0}^{\infty} L(D|z, n_0) p(N|\nu, \delta_\nu) \\ &= p_1^{n_1} p_2^{n_2} \frac{(\alpha + n)!}{\alpha! n!} \beta^{\alpha+1} (\beta + p_1 + p_2)^{-\alpha-n-1}, \end{aligned} \tag{5.32}$$

where $n = n_1 + n_2$, with p_1 and p_2 given by Eqs. (5.12) and (5.13), and α and β given by Eqs. (5.29) and (5.30) respectively.

With Eq. (5.32) at hand, we can now evaluate the size and credibility as a function of λ. We report the results in Fig. 5.4. There are no more kinks in Fig. 5.4

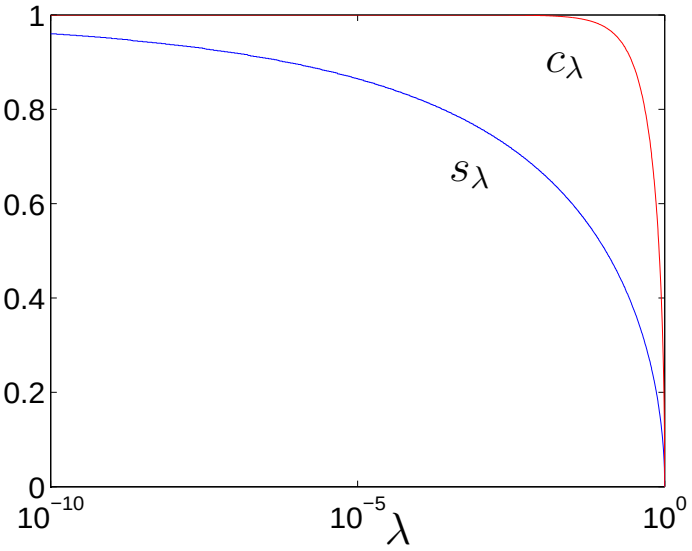

Figure 5.4: Size (the blue curve) and credibility (the red curve) as a function of λ for the bounded-likelihood regions using the marginalized likelihood of Eq. (5.32). There are no more kinks as the discrete parameter has been marginalized over.

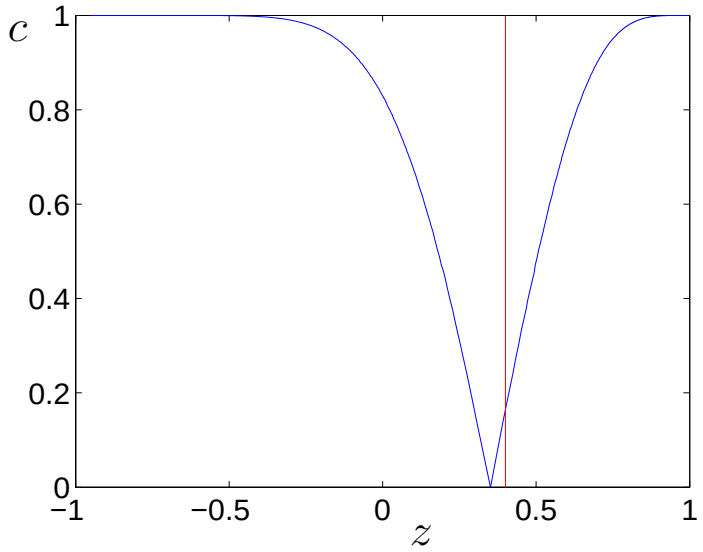

Figure 5.5: The smallest credible interval for the parameter z as a function of credibility. From this figure, one draws horizontal lines to determine the smallest credible interval corresponding to one's desired value of the credibility; see text for details. The red line shows the true state of $z = 0.4$ used in the simulation.

since the the discrete parameter n_0 has been marginalized over. With Fig. 5.4, the experimenter can find the value of λ that corresponds to his or her desired credibility and check whether an arbitrary value of z is in the BLR. Since the parameter of

interest, z, is one-dimensional here, it is more instructive to plot the smallest credible interval as a function of the credibility and this is shown in Fig. 5.5. The vertical axis is the credibility and the blue curve indicates the smallest credible interval. From the figure, one draws a horizontal line at $c = c_0$, where c_0 is the desired value of the credibility. This horizontal line will cut the blue curve at two points, $z_{\min}$, and $z_{\max}$, unless $c_0 = 0$. The smallest credible interval to report is then $(z_{\min}, z_{\max})$. When c goes zero, one obtains the maximum likelihood point estimator. When c_0 goes one, one then needs to report the full interval $(-1, 1)$.

5.4 Estimation of phase in an interferometer

Let us consider an ideal Mach-Zehnder interferometer with lossless beam splitters of 50:50 splitting ratio, and only input port 1 is used. The unknown phase ϕ between the two arms is to be estimated with the help of an auxiliary phase controller which switches the control phase randomly between either 0 or $\pi/2$, see Fig. 5.6. Note that this random switch simply selects either of the two choices 0 or $\pi/2$ half of the time. But when the choice is made, the value of the random auxiliary phase is then known, so that for each copy of the photon, we know the setting of the auxiliary phase control. Effectively, half of the incoming photons are measured with the auxiliary phase set to 0, while the other half with $\pi/2$. The auxiliary phase is needed so as to get a unique estimator for the unknown phase [HMP$^+$96, ŘHZ$^+$99] ; this device is equivalent to the 8-port homodyne detection scheme [NFM91].

The probabilities that the respective detectors will click with each used control phase are,

$$\begin{aligned} p_1 &= \eta_1 \left(\frac{1 + \cos\phi}{4} \right), \quad p_2 = \eta_2 \left(\frac{1 - \cos\phi}{4} \right), \\ p_3 &= \eta_1 \left(\frac{1 - \sin\phi}{4} \right), \quad p_4 = \eta_2 \left(\frac{1 + \sin\phi}{4} \right), \end{aligned} \tag{5.33}$$

where η_j is the given efficiency of the detector D_j.

For the moment, we consider perfect detectors, i.e. $\eta_1 = \eta_2 = 1$. Interesting features in the likelihood function now could arise due to the sinusoidal dependence

on the parameter. For this case, we have

$$L(D|\phi) = \prod_{j=1}^{4} p_j^{n_j}, \tag{5.34}$$

where $D = \{n_1, n_2, n_3, n_4\}$. We estimate the phase ϕ by maximizing the logarithm of the likelihood. Requiring $\dfrac{\partial \log L}{\partial \phi} = 0$ gives

$$\sqrt{A^2 + B^2} \sin\left(\phi + \tan^{-1}\frac{B}{A}\right) = 0, \tag{5.35}$$

where

$$A = \frac{n_2}{1 - \cos\phi} - \frac{n_1}{1 + \cos\phi}, \tag{5.36}$$

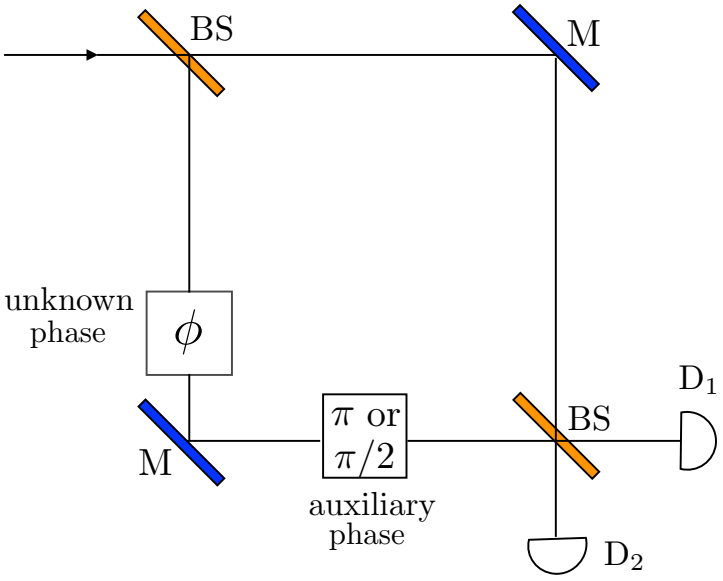

Figure 5.6: An ideal Mach-Zehnder interferometer with lossless beam splitters (BS) of 50:50 splitting ratio and mirrors (M). Only input port 1 is used, and the two output ports are directed to two detectors D_1 and D_2. The unknown phase ϕ between the two arms is to be estimated with the help of an auxiliary phase controller which switches the control phase randomly between either 0 or $\pi/2$. Note that this random switch simply selects either of the two choices 0 or $\pi/2$ half of the time. But when the choice is made, the value of the random auxiliary phase is then known, so that for each copy of the photon, we know the setting of the auxiliary phase control. Effectively, this is a four element POM; see text for details.

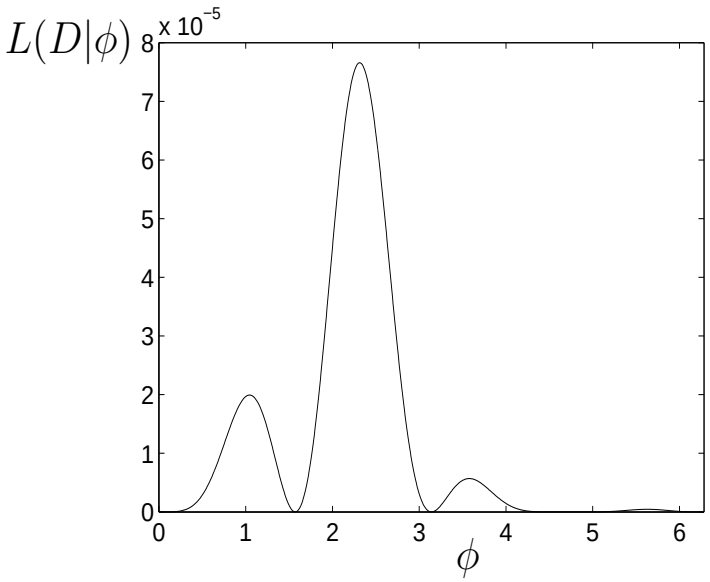

Figure 5.7: Example of a likelihood function exhibiting multiple peaks and multiple regions in the parameter ϕ. One of the peaks is hardly visible.

and

$$B = \frac{n_4}{1+\sin\phi} - \frac{n_3}{1-\sin\phi}. \tag{5.37}$$

Equation (5.35) implies either $A = B = 0$ or $\tan\phi = -B/A$, which tells us that there could be multiple local maxima in the likelihood function plotted against ϕ, depending on the data that is actually obtained in the experiment. We note that each of these local maxima resides in different disjointed regions. For each of the disjointed regions, there is only one local maximum. For illustration, suppose the data $D = \{1, 2, 1, 3\}$ is obtained, Fig. 5.7 shows that the likelihood function now exhibits four peaks in the parameter ϕ (one of the peaks is hardly visible).

Indeed, if we now identify the permissible probabilities for this POM, we have,

$$w_{\text{cstr}}(p) \doteq \theta(p)\delta\left(p_1+p_2-\frac{1}{2}\right)\delta\left(p_3+p_4-\frac{1}{2}\right)\delta\left((p_1-p_2)^2+(p_3-p_4)^2-\frac{1}{4}\right), \tag{5.38}$$

where

$$\theta(p) = \prod_{k=1}^{4}\theta(p_k). \tag{5.39}$$

This is very similar to the crosshair measurement considered in Ref. [SSN+15b].

The only difference is in the replacement of the step function in Eq. (23) of Ref. [SSN+15b]; we have a delta function, which restricts our states to be only the set of pure states on the equator. Such a set is not a convex set, and hence one does not expect the likelihood to be convex in the parameter ϕ. Now, for the primitive prior, we have

$$\begin{aligned}(\mathrm{d}\rho) &= \mathrm{d}p_1\mathrm{d}p_2\mathrm{d}p_3\mathrm{d}p_4 w_{\mathrm{cstr}}(p) \\ &= \mathrm{d}p_2\mathrm{d}p_4\,\delta\left(\left(\frac{1}{2}-2p_2\right)^2+\left(\frac{1}{2}-2p_4\right)^2-\frac{1}{4}\right) \\ &= \frac{1}{2|\cos\phi|}\mathrm{d}p_4 = \frac{\mathrm{d}\phi}{8},\end{aligned} \tag{5.40}$$

so that the multiplicative constant in Eq. (5.38) is $\frac{4}{\pi}$ to ensure the correct normalization. The primitive prior in the probabilities turns out to be the primitive prior in ϕ in disguise.

In general, even if the likelihood function exhibits multiple peaks, the concepts of BLR are still meaningfully defined. The subtle difference is that we now consider variations on the union of the disjoint regions. The search for the MLR is now slightly modified: for given total size of the union of the regions, we want to maximize the sum of the likelihood of the union of the regions, i.e. we would like to have

$$\sum_{i=1}^{k}\int_{\partial R_i}\mathrm{d}\vec{A}_i\cdot\delta\vec{\epsilon}_i\,L(D|\rho,\boldsymbol{\gamma})=0, \tag{5.41}$$

subject to the constraint

$$\sum_{i=1}^{k}\int_{\partial R_i}\mathrm{d}\vec{A}_i\cdot\delta\vec{\epsilon}_i=0, \tag{5.42}$$

where k is the number of disjoint regions in the likelihood function. Here, $\partial\mathcal{R}_i$ is the boundary of the ith region $\mathcal{R}_i$, $\mathrm{d}\vec{A}_i$ is the vectorial surface element of $\partial\mathcal{R}_i$ for the ith region, and $\delta\vec{\epsilon}_i$ is the infinitesimal displacement of ρ and $\boldsymbol{\gamma}$ in the ith region. Since each of the regions are disjointed, the results of the variation show that the MLR and SCR that we are looking for is still the BLR.

The size of the regions can be calculated just as before, using Eq. (5.5). The only difference is that we now need to evaluate the disjoint regions separately. Whenever

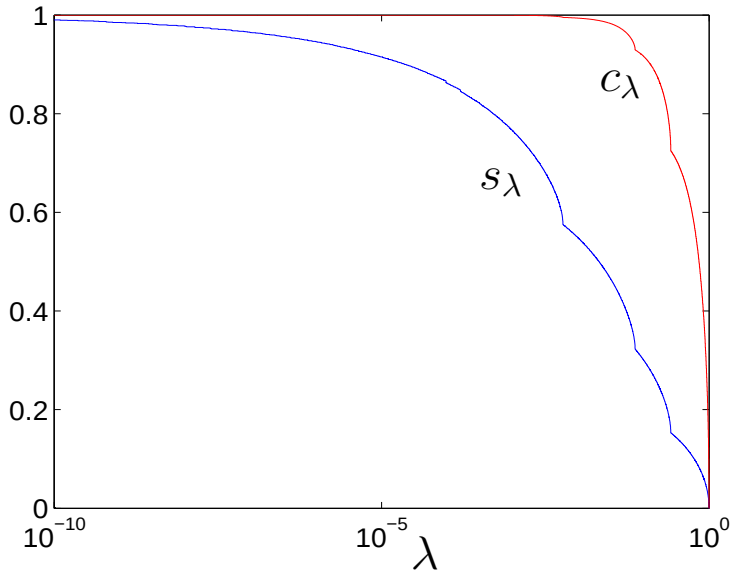

Figure 5.8: Size (the blue curve) and credibility (the red curve) of the BLR. In this case, the BLR consists of a union of regions. The kinks in the graph occurs whenever a further decrease of λ results in more regions being included in the BLRs.

we want the size of a union of regions, we simply add the size of each of the regions together to obtain the total size. Once the size is determined, the credibility is found in the same way, using Eq. (5.8).

Figure 5.8 shows the size and credibility as a function of λ for the data $D = \{1, 2, 1, 3\}$. Each of the disjoint regions has a size of 0.25 as λ goes to zero, since a primitive prior for ϕ is used. The kinks of the graph occur whenever the further decrease of λ results in more regions being included in the BLRs.

Similar to the first example, since the parameter, ϕ, is one-dimensional, it is more instructive to plot the smallest credible interval as a function of the credibility and this is shown in Fig. 5.9. From the figure, one can easily determine the optimal interval to report if one wants a certain value of the credibility, in the same manner as explained in the case of Fig. 5.5. The difference is that there are now multiple branches in the blue curve, so that the horizontally drawn line may cut the blue curve at more than two points, depending on the chosen value of the credibility. For example, if one wishes to have a credibility of $c = 0.95$, then the dashed black line cuts the blue curve six times, resulting in three pairs of $(\phi^i_{\rm min}, \phi^i_{\rm max})$, for $i = 1, 2$ and 3. As a result, one has to report the union of these three disjoint intervals as

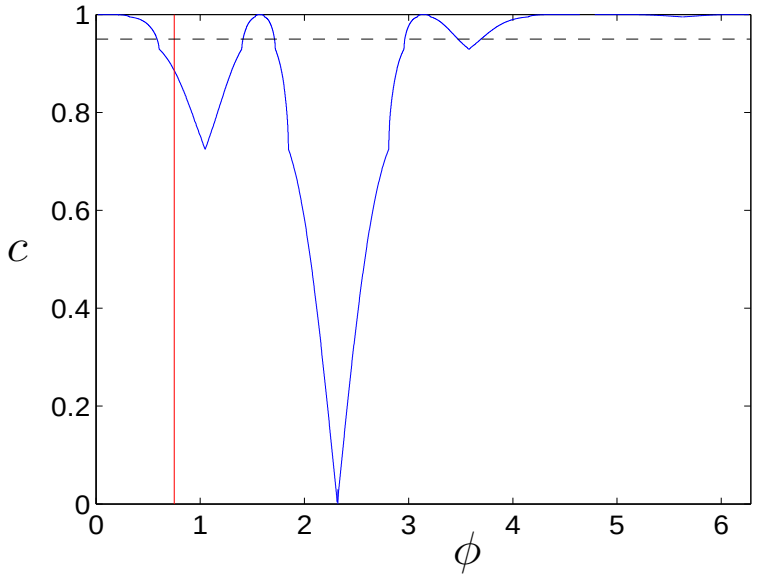

Figure 5.9: The smallest credible interval for ϕ as a function of the credibility. The red line is the true state $\phi = 0.75$ used in the simulation. The black dash shows that if one desires a credibility of $c = 0.95$, then one has to report the union of three intervals.

the smallest credible interval. For 95% of the time, the true state will be inside this interval. If only $c = 0.8$ is needed, then reporting the union of two intervals will suffice.

From this example, the advantage of reporting MLR, rather than a single ML estimator, is manifested obviously. The ML estimator (the one corresponding to the global maximum) for this case is $\phi = 2.3$ which is far away from the true state. This is because the number of copies measured in the experiment is too small to make a conclusive statement about the unknown phase ϕ. If the number of copies becomes larger, all the side peaks will diminish and become unnoticeable, and the global peak will stand out as the ML estimator near the true state. However, when the number of copies measured is small, fluctuations in the data could easily result in the scenario shown in Fig. 5.9, where the red line is outside the main interval. In such cases, one really is not equipped with enough confidence to report the ML point estimator, and rather reporting the MLR is the better way. Notice that even if the number of copies are large, the side peaks are still there and there would still be four disjoint regions (as long as one of the n_j is non-zero, which practically never

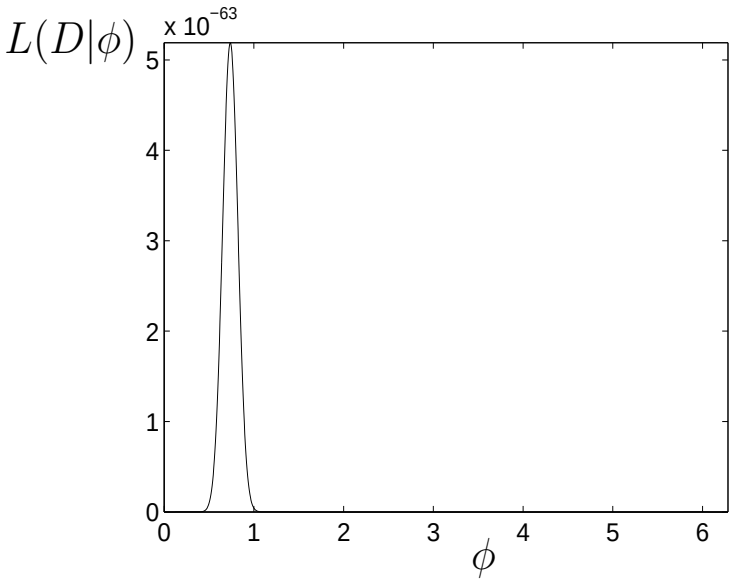

Figure 5.10: Likelihood as a function of ϕ when the number of copies used is large. The ML estimator now is very close to the true state $\phi = 0.75$. The other three peaks are so low that they are practically not there.

happens as N goes to infinity in an experiment). It is just they are so small to be even noticed.

Figure 5.10 shows the likelihood function when the number of copies used becomes larger. The global peak will stand out near the true state $\phi = 0.75$, whereas all the other peaks diminished so low that they are practically not there. However, one can see from Fig. 5.11, which is the logarithm of the likelihood function, that there are four maximua. Only the global peak will be the consistent estimator when N becomes larger.

Having settled the occurrence of multiple peaks in the likelihood function, we now return to the case of a realistic imperfection detection. The likelihood of obtaining the data D is then

$$L(D|\phi, n_0) = p_0^{n_0} \prod_{j=1}^{4} p_j^{n_j} \binom{N}{n_0}, \tag{5.43}$$

where $p_0 \equiv 1 - \sum_{j=1}^{4} p_j$ and $N = \sum_{j=0}^{4} n_j$, with n_0 denoting the unknown number of

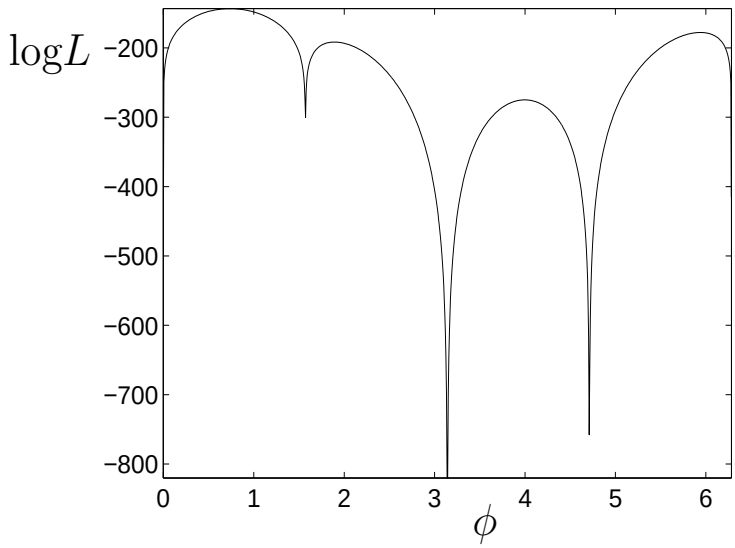

Figure 5.11: Logarithm of the likelihood function where one can see the appearance of four maxima.

undetected copies and $n \equiv \sum_{j=1}^{4} n_j$ is the number of detected copies.

We can estimate the phase ϕ and n_0 simultaneously in the same manner as Section 5.3. For ϕ, we have

$$\sqrt{A^2 + B^2} \sin\left(\phi + \tan^{-1}\frac{B}{A}\right) = 0, \tag{5.44}$$

where

$$A = \frac{n_0\eta_-}{4 - 2\eta_+ + \eta_-(\sin\phi - \cos\phi)} + \frac{n_2}{1 - \cos\phi} - \frac{n_1}{1 + \cos\phi}, \tag{5.45}$$

and

$$B = \frac{n_0\eta_-}{4 - 2\eta_+ + \eta_-(\sin\phi - \cos\phi)} + \frac{n_4}{1 + \sin\phi} - \frac{n_3}{1 - \sin\phi}, \tag{5.46}$$

with $\eta_+ \equiv \eta_1 + \eta_2$ and $\eta_- \equiv \eta_1 - \eta_2$.

For the undetected copies, we have

$$\hat{n}_0^{\mathrm{ML}} = \mathrm{floor}\left(\frac{np_0}{1 - p_0}\right). \tag{5.47}$$

In summary then, given a set of data, $\hat{\phi}^{\mathrm{ML}}$ and $\hat{n}_0^{\mathrm{ML}}$ are obtained from solving both Eqs. (5.44) and (5.47), simultaneously. One can then proceed with either reporting

combined optimal error regions for ϕ and n_0, or marginalizing over the nuisance parameters to report the optimal error interval for the parameter of interest, similar to the manner that we demonstrated in Section 5.3 above.

Instead of doing that, here we demonstrate how the detector efficiency could be simultaneously estimated with error intervals. For illustrative purpose, suppose η_1 is known and $\eta_2 = \eta$ is to be estimated. One first obtains a marginal likelihood on ϕ and η after marginalizing over n_0, so that

$$L(D|\phi,\eta) = \frac{(\nu T)^n}{n!} p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4} \mathrm{e}^{-\nu T p}, \tag{5.48}$$

where $p = p_1 + p_2 + p_3 + p_4$.

Now a suitable prior for η is the Beta distribution so that

$$p(\eta) = \frac{1}{B(\alpha,\beta)} \eta^{\alpha-1}(1-\eta)^{\beta-1}, \tag{5.49}$$

with

$$\alpha = \bar{\eta}\left(\frac{\bar{\eta}(1-\bar{\eta})}{(\delta_\eta)^2} - 1\right), \tag{5.50}$$

and

$$\beta = (1-\bar{\eta})\left(\frac{\bar{\eta}(1-\bar{\eta})}{(\delta_\eta)^2} - 1\right). \tag{5.51}$$

As usual, the calibration on the detector before the experiment provides us with the values of $\bar{\eta}$ and δ_η. If such calibration were not possible, the experimenter could assume a uniform prior for η over the interval zero to one. Which prior to use really depends on how much the experimenter knows about his or her apparatus.

We can then obtain the marginal likelihood on ϕ given by

$$\begin{aligned} L(D|\phi) =& \int_0^1 \mathrm{d}\eta L(D|\phi,\eta) p(\eta) \\ =& \mathrm{e}^{-\nu T(p_1+p_3)} \left(\frac{1-\cos\phi}{4}\right)^{n_2} \left(\frac{1+\sin\phi}{4}\right)^{n_4} \\ & \times \frac{p_1^{n_1} p_3^{n_3} (\nu T)^n}{n!} \frac{\Gamma(a)\Gamma(\beta)}{\Gamma(a+\beta)B(\alpha,\beta)} M(a,\beta,-c), \end{aligned} \tag{5.52}$$

where $\Gamma(\ )$ is the gamma function, $B(\ )$ is the beta function, and $M(\ )$ is the

confluent hypergeometric function, with

$$a = n_2 + n_4 + \alpha, \tag{5.53}$$

and

$$c = \nu T \left(\frac{2 + \sin\phi - \cos\phi}{4} \right). \tag{5.54}$$

If, however, the experimenter is interested in the value and credible interval of η only, then marginalizing over ϕ gives

$$L(D|\eta) = \frac{(\nu T)^n}{n!} \frac{\eta_1^{n_1+n_3} \eta^{n_2+n_4}}{4^n} \mathrm{e}^{-\nu T \frac{\eta_1+\eta}{2}} f(\eta), \tag{5.55}$$

where $f(\eta)$ is obtained from the integral

$$\begin{aligned} &\int_0^{2\pi} \mathrm{d}\phi \, \mathrm{e}^{-\frac{\nu T}{4}(\eta_1-\eta)(\cos\phi-\sin\phi)} (1+\cos\phi)^{n_1} (1-\cos\phi)^{n_2} (1-\sin\phi)^{n_3} (1+\sin\phi)^{n_4} \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{n_1} \left(\frac{\mathrm{d}}{\mathrm{d}u}\right)^{n_2} \left(\frac{\mathrm{d}}{\mathrm{d}v}\right)^{n_3} \left(\frac{\mathrm{d}}{\mathrm{d}w}\right)^{n_4} g(u,v,w,t)|_{t,u,v,w\to 0} \,, \end{aligned} \tag{5.56}$$

with

$$g(u,v,w,t) = \mathrm{e}^{t+u+v+w} I_0 \left(\sqrt{(t-u-d)^2 + (w-v+d)^2} \right), \tag{5.57}$$

where I_0 is the modified Bessel function of the first kind with order zero, and $d = \frac{\nu T}{4}(\eta_1 - \eta)$.

5.5 Discussions

In this chapter, we presented two typical examples illustrating the construction of error regions for quantum tomography experiment with additional unknown parameters. As there is a myriad of parameters that could be unknown in various experiment, it is not possible, nor the aim of this chapter, to study in details all such possible parameters. The two chosen examples are meant to illustrate the technique. However, as an outlook, we do mention two other examples that we plan to study in the future due to their unique features. One of them is quantum homodyne tomography [VR89]. It is a robust and versatile tool of quantum optics and has been applied in many different experimental settings for which the technique of maximum

likelihood has been used successfully. One of the many advantage of the ML technique is the possibility to incorporate the detector inefficiencies. A common model for a homodyne detector of non-unitary efficiency results in a generalized Bernoulli transformation for the transmitted density matrix. When aware of the homodyne detector efficiency η, one can run well-known iterative algorithm and reconstructs the original density matrix [Pv04].

To illustrate the idea of Bayesian credibility region in such experiments, it suffices to consider tomography of Gaussian states. Such states represent the wide class of coherent, squeezed and thermal states, all of them being characterized by a Gaussian Wigner function. Let's consider a thermal squeezed state subject to homodyne measurement. The thermal squeezed state is characterized by two state parameters, the squeezing parameter, ν, and the true mean photon number, n_{T}. The measurement device is characterized by an additional parameter, η, the efficiency of the detectors. The statistics of the quadrature distributions of such a state is given by

$$p(\theta, x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-x^2/2\sigma^2), \tag{5.58}$$

with

$$\sigma^2 = \eta \left[\frac{2n_{\mathrm{T}} + 1}{2Q} \left(Q^2 (\cos\theta)^2 + (\sin\theta)^2 \right) + \delta \right], \tag{5.59}$$

where $Q = \exp(2\nu)$, and $\delta = (1 - \eta)/2\eta$ [MŘH12]. Note that in the Ref. [MŘH12], such an example has already been studies, without however, attaching optimal error regions to such a scenario. Furthermore, the Figure. 3 presented in Ref. [MŘH12] is a joint property of the state parameters and additional parameters. The idea of marginalization can also been applied in this case. However, in this case, numerical integration is required for the marginalization.

Another example is an experiment involving dark counts of the detectors. At the moment, such experiments are always done with pre-calibration of the detector before the actual experiment. That is, one first switches on the detector alone and records the average number of counts per unit time due to the dark counts, N_{dark}, without running the experiment. One then runs the experiments and records the number of counts, N_{observed}. The number of actual counts is then corrected for the

dark count using

$$N_{\text{actual}} = N_{\text{observed}} - N_{\text{dark}}. \tag{5.60}$$

However, due to the statistical fluctuations in the data, this could sometimes results in a negative value of N_{actual}, which does not make sense. This situation is worse when the detector has a high dark count, or the state is nearly pure or rank-deficient, so that N_{actual} is very small for one of the POM element measured, or both.

A more sensible way to analyze such an experiment would be to estimate the dark counts simultaneously instead of doing a pre-calibration first and correction later. The techniques presented in this chapter then become relevant and we remark that such a study is still underway.

5.6 Conclusions

In conclusion, we note that, in a typical experiment, additional parameters apart from the state could also be unknown. Self-calibration, i.e. the simultaneous reconstruction of a quantum state and various additional parameters needed to characterize the measurement device, has been recently introduced. However, such a reconstruction should be attached with error regions to make them statistically meaningful. In this chapter, we illustrate that optimal error regions for such experiments can indeed be constructed efficiently using the bounded-likelihood region. The optimal error regions constructed is a joint property of the state parameters and the other parameters as a whole. We show how to separate the different estimators from their joint optimal regions and attaching error regions (intervals) to them separately by marginalizing over the nuisance parameters to obtain a marginal likelihood which only depends on the parameter of interest. The marginal likelihood characterizes one's evidence that he or she has about the parameter of interest. The bounded-likelihood intervals obtained from the marginal likelihood are the smallest credible intervals. Two typical experimental situations are studied in details as illustrations, and we offer two other examples to be studied as an outlook.

Chapter 6

Initial System-Bath State

In this chapter, we show how one uses ideas from quantum state tomography to deduce a reasonable and consistent initial system-bath state. In typical situations, such a state turns out to be uncorrelated or almost uncorrelated between the system and the bath. This has implications, in particular, on the subject of subsequent non-Markovian or non-completely-positive dynamics of the system, where the non-complete-positivity stems from initial correlations between the system and the bath.

6.1 Introduction

As mentioned previously, the first step in a quantum information processing or computation task always involves the preparation of a certain initial state, which would be subjected to some dynamical evolution later on, for example, by letting the state evolve freely, or sending it through a series of gates for manipulation aimed at a certain computation. The subsequent dynamics typically depends on the initial state prepared.

Such a dependence also appears in the study of open quantum systems [BP06], where the subject of interest includes not only the quantum system, but also the environment that it inevitably couples to. Strictly speaking, no quantum system is truly an isolated system and all quantum systems couple to some extent to the external environment. Hence, the initial condition in such a situation should not include the system-only state ρ_{S}, but rather the system-environment composite state ρ_{SE}.

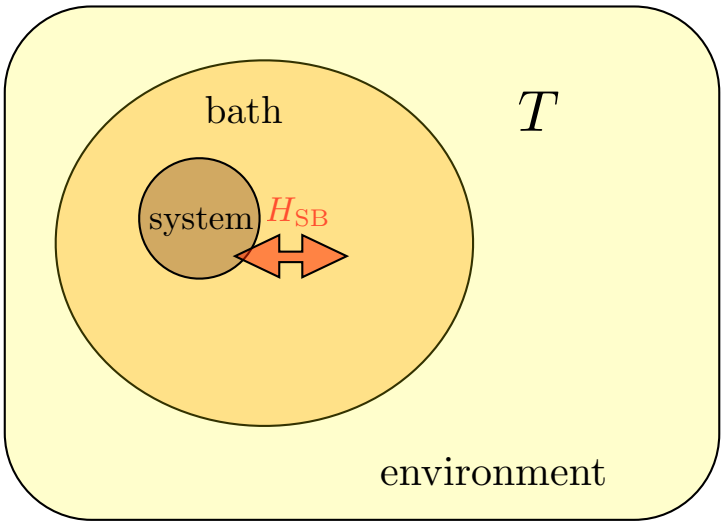

Figure 6.1: The system sitting in an immediate larger bath, which is further immersed in the environment. For example, in an ion-trap experiment, one typically has the system qubits, which are coupled to the bath qubits, and the system-bath composite is immersed in the external environment. Hence, the system-bath composite is not thought of as a closed quantum system here. However, the main coupling to the system comes from its immediate bath through the interaction Hamiltonian H_{SB}, and the coupling between the system and the larger external environment is negligible. The bath serves the purpose of the "memory-full" part of the environment that interacts with the system. The environment however provides a mechanism for the system-bath composite to be maintained at a certain temperature T.

Strictly speaking, in order to form a closed quantum system, the environment includes the "rest of the universe". However, the description involving the "rest of the universe" is normally neither feasible, nor is it necessary or sensible. Furthermore, in a typical experimental situation, such coupling to the "rest of the universe" is also not capturing the essence of the physics. Rather, one often has the system sitting in an immediate larger bath, which is further immersed in the environment. For example, in an ion-trap experiment, one typically has the system qubits, which are coupled to the bath qubits, and the system-bath composite is immersed in the external environment. Hence the situation we are considering here is slightly different from the one introduced in Chapter 2, and is a more realistic model. The typical scenario is depicted in Fig. 6.1. Here, the system-bath composite is not thought of as a closed quantum system, but immersed in an environment that acts

as the temperature reservoir. However, the main coupling to the system comes from its immediate bath through the interaction Hamiltonian H_{SB}, and the coupling between the system and the larger external environment is negligible. The bath in this case serves the purpose of the "memory-full" part of the environment that interacts directly with the system. Information flows between the system and the bath and very little goes into the larger external environment. The environment, however, provides a mechanism for the system-bath composite to be maintained at a certain temperature T. Hence, what we need here is an initial system-bath composite state ρ_{SB} acting on $\mathcal{H}_{\mathrm{S}} \otimes \mathcal{H}_{\mathrm{B}}$, for predictions of future behavior of the composite system. One may object by saying that what is important is the system-only state, and one does not need to predict any future behavior of the bath state. For that, one may resort to the master equation description of the system state, where the bath state is traced over. However, even in such a system-only description, the specification of an initial system-bath composite state is also required as an important step towards deriving the system-only master equation. Hence, one cannot avoid specifying an initial system-bath state even if the predictions later on only involve the system state.

What is this ρ_{SB} then? This, of course, depends on how we prepare the system-bath composite and is then verified by quantum state tomography. However, though the bath is a small bath compared with the "rest of the universe", it is still large compared with the system, and we do not have microscopic control on the bath, unlike what we have on the system. Hence, what we have is only the ability to prepare the system state, but we do not have the ability to prepare the bath state in the same state every time. As tomography is only done on the system, it verifies that we have accurately prepared the system in the state $\rho_{\mathrm{S}} = \mathrm{tr}_{\mathrm{B}}\{\rho_{\mathrm{SB}}\}$. With this ρ_{S} now at hand, one then has to infer a reasonable and consistent system-bath composite state ρ_{SB} that incorporates what we know and do not know about the bath.

The easiest way to accomplish this is probably to simply attach some reference bath state $\rho_{\mathrm{B}}^{(\mathrm{ref})}$ together with ρ_{S} and claim that $\rho_{\mathrm{SB}} = \rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{ref})}$, since a tensor-product state has a very straightforward interpretation in the classical sense. Indeed,

this state has been motivated traditionally by the argument that the system is well isolated from the bath—only when the system-bath interaction is weak does this split into system and bath make good physical sense—so that it can be described as having been prepared in some state independently of the bath. The preparation of the system is normally assumed to have little effect on the bath since the system is too insignificant to have any effect on the large bath, which is in its own reference state, as if the system were not there. This reference state is typically assumed to be the thermal state $\rho_{\mathrm{B}}^{(\mathrm{th})}$, since at thermal equilibrium with the environment, the bath is expected to follow the Boltzmann distribution in its eigenstates. This reasoning gives the $\rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})}$ as the initial system-bath state. One arrives at the same state by alternatively imagining that the system-bath interaction is originally zero and is "turned on" only at time $t = 0$.

The description using such an initial product state is one of the ingredients to what is called *Markovian* dynamics later on, where one neglects memory effects in the reduced system dynamics. However, practically, the isolation of the system from the bath can never be complete; for example, one can never decouple from the quantized electromagnetic field. One expects the initially perfectly uncorrelated or independent situation to be but an approximation, albeit often a good one. Recently, many articles [MRRAG12, DRS12] examined the effects of initial correlations with the bath on subsequent system dynamics, usually with the observation of deviation from purely Markovian behavior. Various measures of this deviation have been proposed [BLP09, RWC$^+$11, WECI08, RHP10]. Experiments have been performed, [LLH$^+$11, TLL$^+$12], demonstrating the possibility of artificially engineering initial correlation between the system and the (some controllable microscopic) bath. A more recent experiment [RWM$^+$15] studied how one could use these deviations to deduce the initial (engineered) correlations between a single qubit and its bath (also a single qubit); the method applies to larger systems and baths, as proposed in the theoretical paper [Mod12], although one questions its feasibility for a larger bath.

In such situations of initial system-bath correlations, the simple recipe of writing down an uncorrelated $\rho_{\mathrm{SB}} = \rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{ref})}$ is no longer fully justified. What is then a reasonable ρ_{SB}? In many of these works discussing the initial system-bath correla-

tions, since one generally lacks any recipe to write down a specific system-bath state, the system-bath state is taken as some arbitrary or given state, usually suggesting that it is some fixed state, just that we do not know what it is. Such a view of some unknown arbitrary state is hardly supported since the quantum mechanical state is *our* description of the system in question, taking into consideration all the available information we have about the state at hand. The system-bath state should thus capture what *we know* about the system and the bath. There are also papers which introduce a symbolic unknown system-bath state, as if the system-bath state can be anything and no concrete statement could be made [KMRRS07]. Such a point of view not only possesses little predictive power, but is also unfounded since we can and should always apply the standard methods of state estimation to deduce an initial system-bath state that represents our knowledge of the composite system. There have also been other attempts to characterize initial correlation based on initial preparation of the system state [Mod12, KMRRS07, MS10]. State preparation, however, can only be verified by the tools of tomography.

Hence, in this chapter, we examine two different ways of how one employs ideas from state estimation to write down a statistical operator for the initial system-bath state. One way is to apply the idea of maximum entropy [Jay57], and argue that the resulting state is the most reasonable initial state that one is allowed to write down, based on our ignorance. Any other guesses will introduce biases not supported by the evidence. Another way is to write down the Bayesian mean state in accordance with any prior information known (or with an uninformative prior if that is the case) to the experimenter. We show in this chapter that both approaches result in a system-bath state that is uncorrelated or almost uncorrelated.

6.2 Setting the stage

In a typical quantum experiment, the first step always involves the preparation of the system state at time $t = 0$. The success or accuracy of such a preparation is verified by quantum state tomography. The system is repeatedly prepared under the same experimental conditions and these independently and identically prepared copies of the system are measured using a POM $\{\Pi_i\}_{i=1}^{K}$, as discussed in Chapter 2.

For the purpose of this chapter, we assume that we use an IC POM to extract the information on the system. Our knowledge of the system can thus be phrased simply as the knowledge of the p_is for the tomography POM. The cases where the relative frequencies are not the true probabilities, due to, for example, small number of copies being measured, or the cases where there is contradictions in the data obtained, are mentioned before and well discussed in the quantum state tomography literature. We will refrain from discussing about these issues in this chapter and assume that one measures a sufficiently large number of copies so that the p_is are known from the data, enabling us to reconstruct a point estimator and it is not necessary to give a region of estimators. Note that, however, slightly different from the situation introduced in Chapter 2, here the input state is a system-bath state, that is ρ_{SB} from $\mathcal{H}_{\mathrm{S}} \otimes \mathcal{H}_{\mathrm{B}}$, instead of ρ_{S} from $\mathcal{H}_{\mathrm{S}}$. The POM elements are $\Pi_i \otimes \mathbb{1}_{\mathrm{B}}$, where the Π_is are the system operators and $\mathbb{1}_{\mathrm{B}}$ is the identity operator on the bath. In order not to overload the notation, we will not write $\mathbb{1}_{\mathrm{B}}$ explicitly when the meaning is clear from the context.

What about our knowledge of the bath? As discussed, by the "bath", we refer to the part of the relevant physical system that we do not have microscopic control over. However, we can have control over the macroscopic properties, such as temperature, pressure, volume, etc., and we can measure these macroscopic properties in the laboratory accurately. By "relevant", we mean that we include only those physical degrees of freedom that couple sufficiently, though weakly, to our controlled system to have any effects on its dynamics, for the duration of the experiment. For example, these degrees of freedom could be the cavity modes that are resonant or near-resonant to a few-level atom (the idealized system) inside the cavity; the surrounding environment of nuclear spins around a nitrogen-vacancy-centre spin site; etc. Based on the experimenter's knowledge and understanding of the actual physical situation in the experiment, this coupling can be identified as an interaction Hamiltonian H_{SB} between the system and the bath. Alongside this H_{SB}, there is also a system-only Hamiltonian H_{S}, and a bath-only Hamiltonian H_{B}. More general couplings, for example, evolution via a joint physical map, are possible, but for our current discussion, following also the usual scenario in the open-systems literature, we will restrict ourselves to this simplest joint unitary evolution description.

To summarize, in the system-bath model considered here, our knowledge of the system-bath composite at time $t = 0$ is as follows:

1. the microscopic system state from knowing the probabilities $p_i = \mathrm{tr}\{\rho_{\mathrm{S}}\Pi_i\}$, $i = 1, 2, \ldots, K$;

2. the bath degrees of freedom, for example the accessible states, the dimension of the relevant Hilbert space, etc.;

3. the overall Hamiltonian $H = H_{\mathrm{S}} + H_{\mathrm{B}} + H_{\mathrm{SB}}$;

4. the macroscopic properties of the bath and system, for example, temperature, pressure, volume, etc.

6.3 The maximum-entropy state

Without the ability to perform tomography on the bath for determining its microscopic state, we have only incomplete information about the overall system and bath composite at time $t = 0$. The initial system-bath state that we write down must hence reflect this incomplete knowledge. The use of maximum-entropy methods in such situations follows a long tradition dating back to the beginnings of statistical physics. More recently, Jaynes [Jay57, Jay82] argued lucidly for the viewpoint of maximum entropy as the most rational statistical inference approach in the face of incomplete information. We have already utilized this idea in another context; see Chapter 4.

The maximum-entropy state $\rho_{\mathrm{SB}}^{(\mathrm{ME})}$ for the system-bath composite is simply the state that attains

$$S_{\max} = S\left(\rho_{\mathrm{SB}}^{(\mathrm{ME})}\right) = \max_{\rho_{\mathrm{SB}}} S(\rho_{\mathrm{SB}}), \tag{6.1}$$

where the entropy, $S(\rho) = -\mathrm{tr}\{\rho \log \rho\}$, is the von Neumann entropy. Here, ρ_{SB} comes from the set of all system-bath states that are consistent with our partial knowledge of the system and the bath, that is $\mathrm{tr}_{\mathrm{B}}\{\rho_{\mathrm{SB}}\} = \rho_{\mathrm{S}}$, or in terms of the measured probabilities,

$$\mathrm{tr}\{\rho_{\mathrm{SB}}(\Pi_i \otimes \mathbb{1}_{\mathrm{B}})\} = p_i, \tag{6.2}$$

together with any additional macroscopic knowledge one may have about the bath. The constrained maximization can be incorporated using Lagrange multipliers, so that we instead maximize

$$\tilde{S}(\rho) \equiv S(\rho) - \sum_i \lambda_i(\mathrm{tr}\{\rho\Pi_i\} - p_i) - \nu(\mathrm{tr}\{\rho\} - 1) - \mu g(\rho) \tag{6.3}$$

over all system-bath ρ. Here, λ_is are the Lagrange multipliers for the system tomography information, the ν term enforces the unit-trace constraint for ρ, and $\mu g(\rho)$ is symbolic for the Lagrange multiplier term that accounts for any additional knowledge that one may have.

As an example, we consider the situation where the only relevant macroscopic property is the temperature, as is often the case in experiments. As a first check, suppose that the bath is initially in thermal equilibrium, and the system is "opened" and exposed to the bath at time $t = 0$. This describes many cold-atom experiments where the atoms are cooled in one part of the vacuum chamber, before being transported to a different part of the chamber where the experiment is performed. Before the introduction of the atoms into the final chamber, the bath degrees of freedom can be thought of as being in thermal equilibrium at some temperature T; the atoms, on the other hand, are prepared in some state ρ_S in the cooling chamber (assuming that it is unchanged under transport; if not, one has to replace ρ_S with the final system state after transport). Clearly, the initial system-bath state—upon introducing the system into the final chamber—should be the product $\rho_\mathrm{S} \otimes \rho_\mathrm{B}^{(\mathrm{th})}$, where $\rho_\mathrm{B}^{(\mathrm{th})}$ is the thermal state of the bath at temperature T. Indeed, maximizing entropy subjected to the tomography constraints on the system as well as the bath temperature yields immediately the product state above.

The situation becomes more complicated when the system is prepared *in situ*, that is, in constant contact with the bath, and the combined system-bath is held at temperature T. In this case, there is little support for an initially uncorrelated system-bath state $\rho_\mathrm{S} \otimes \rho_\mathrm{B}^{(\mathrm{ref})}$, although this is commonly used in the literature even for such cases. The preparation of the system state can sometimes be accomplished very rapidly, but even so, the coupling between system and bath is always on and can modify the outcome of the preparation procedure, albeit weakly. Many schemes

employ adiabatic gates, that is, slow processes that gradually move the system population into the desired state. And during this slow process, the interaction $H_{\rm SB}$ can be thought of as continuously keeping the system in thermal equilibrium with the bath. Even in the fast-gates situation, one cannot help but worry about the effects of $H_{\rm SB}$ when the preparation of a complicated system state involves a long sequence of gates. For example, in some ion-trap experiments, a sequence of over a hundred gates is applied to initialize the system into the desired state. In all such cases, the coupling to the bath has to be taken into account when stating the initial system-bath state.

An application of the maximum-entropy principle with the constraint of fixed overall temperature for the system and the bath—in thermal equilibrium with each other (and with the external environment)—with the tomographic constraints given in Eq. (6.2) yields the state

$$\rho_{\rm SB}^{\rm (ME)} = \exp\left(-\beta H - \Lambda\right)/{\rm tr}\{\ldots\}, \tag{6.4}$$

with the temperature parameter $\beta \equiv 1/kT$, k being the Boltzmann constant, and $\Lambda \equiv \sum_i \lambda_i \Pi_i$. The notation ${\rm tr}\{\ldots\}$ when appearing in the denominator of a fraction always means the trace of the numerator. This trace takes care of the unit trace constraint associated with the ν Lagrange multiplier.

As an illustration, we work out this maximum-entropy state for the Jaynes-Cummings model, where a single spin-half particle (the system) is coupled to a single mode of light (the bath) in a cavity. In this model, we have the overall Hamiltonian

$$H = \underbrace{\frac{1}{2}\hbar\omega_a\sigma_z}_{H_{\rm S}} + \underbrace{\hbar\omega a^\dagger a}_{H_{\rm B}} + \underbrace{J(\sigma_- a^\dagger + \sigma_+ a)}_{H_{\rm SB}}, \tag{6.5}$$

where $\sigma_\pm = \frac{1}{2}(\sigma_x \pm {\rm i}\sigma_y)$ are the atomic lowering (raising) operators, and $\sigma_i, i = x, y, z$ are the Pauli operators as usual; a and $a^\dagger$ are the photon ladder operators. By an appropriate choice of phases in the a, $a^\dagger$ operators, or in the spin states, J can be taken as real without loss of generality. It is a small quantity, corresponding to the condition that the system-bath coupling is weak. The $H_{\rm S}$ eigenstates are $|1\rangle$ (ground state) and $|0\rangle$ (excited state).

If one ignores the system-bath interaction, that is, putting $H_{\rm SB} = 0$, then we have

$$\rho_{\rm SB}^{\rm (ME)} = \frac{{\rm e}^{-\beta\hbar\omega a^\dagger a - \sum_i \lambda_i \Pi_i}}{{\rm tr}\{\cdots\}} = \rho_{\rm S} \otimes \rho_{\rm B}^{\rm (th)}, \tag{6.6}$$

where the system state is

$$\rho_{\rm S} = \frac{1}{2}(\mathbb{1} + \boldsymbol{s}\cdot\boldsymbol{\sigma}) = \frac{{\rm e}^{-\sum_i \lambda_i \Pi_i}}{{\rm tr}\{\cdots\}}. \tag{6.7}$$

Note here that one could have started using any legitimate POM specified by $\{\Pi_i\}$ (assumed to be IC), with constraint ${\rm tr}\{\rho_{\rm S}\Pi_i\} = p_i$. But one always has the freedom to switch to a different set of operator basis to express these constraints to make reference to the same $\rho_{\rm S}$. Here for convenience, we will choose the Pauli basis, with the constraints now appearing as ${\rm tr}\{\rho_{\rm S}\sigma_i\} = s_i$. Surely, $\sigma_i \neq \Pi_i$, but one can always find different Langrange multipliers, λ_is and $\tilde{\lambda}_i$s, such that

$$\rho_{\rm S} = \frac{{\rm e}^{-\sum_i \lambda_i \Pi_i}}{{\rm tr}\{\cdots\}} = \frac{{\rm e}^{-\sum_i \tilde{\lambda}_i \sigma_i}}{{\rm tr}\{\cdots\}}. \tag{6.8}$$

We will then put $\Lambda = \sum_i \lambda_i \sigma_i$ onwards for this example. Also note that one can always include the $H_{\rm S}$ term in the $\boldsymbol{\lambda}\cdot\boldsymbol{\sigma}$ part, so that for convenience, we will put $H_{\rm S} = 0$ subsequently.

However, if one takes into consideration the interaction due to a non-zero $H_{\rm SB}$, then,

$$\rho_{\rm SB}^{\rm (ME)} = \frac{{\rm e}^{-\beta(\hbar\omega a^\dagger a + J(\sigma a^\dagger + \sigma^\dagger a)) - \sum_i \lambda_i' \sigma_i}}{{\rm tr}\{\cdots\}}. \tag{6.9}$$

At this stage, it is important to realize that the Lagrange multipliers are different in these two expressions Eq. (6.6) and Eq. (6.9), due to the absence or presence of the term $H_{\rm SB}$. The difference $\delta\lambda_i = \lambda_i' - \lambda_i$ is small, as $H_{\rm SB}$ is assumed to be a weak perturbation. The λ_i and λ_i' are, of course, related via the constraint,

$$\frac{{\rm tr}_{\rm B}\{{\rm e}^{-\beta(H_{\rm B}+H_{\rm SB})-\Lambda'}\}}{{\rm tr}\{\cdots\}} = \frac{{\rm tr}_{\rm B}\{{\rm e}^{-\beta H_{\rm B}-\Lambda}\}}{{\rm tr}\{\cdots\}} = \rho_{\rm S}, \tag{6.10}$$

since both cases must result in the same system-only operator due to the tomographic constraint. Now, since J is a small quantity that we will keep to first order

(higher-order terms can be systematically worked out if necessary), the term $H_{\rm SB}$ and $\delta\Lambda = \Lambda' - \Lambda$ can be treated as small perturbations. We apply the variation of the exponential of an operator, given by

$$\delta\left(\mathrm{e}^{-\alpha A}\right) = -\alpha \int_0^1 \mathrm{d}\gamma\ \mathrm{e}^{-\gamma\alpha A}\,\delta A\,\mathrm{e}^{-(1-\gamma)\alpha A}. \tag{6.11}$$

on Eq. (6.10). Given the physical context, we expect $\|H_{\rm SB}\| \ll \|H_{\rm B}\|$, where $\|A\|$ denotes some operator norm of A. Let $\alpha \equiv \|H_{\rm B}\|$, $A \equiv \frac{1}{\alpha}(\beta H_{\rm B} + \Lambda)$, and $\delta A \equiv \frac{1}{\alpha}\left(\beta H_{\rm SB} + \delta\Lambda\right)$. If we now denote the function

$$f(\cdot) = \frac{\mathrm{e}^{-\alpha(\cdot)}}{\mathrm{tr}\{\mathrm{e}^{-\alpha(\cdot)}\}}, \tag{6.12}$$

we have

$$\begin{aligned}\delta f =& -\alpha f(A)\left[\int_0^1 \mathrm{d}\gamma\ \mathrm{e}^{(1-\gamma)\alpha A}\,\delta A\,\mathrm{e}^{-(1-\gamma)\alpha A} - \frac{\mathrm{tr}\{\mathrm{e}^{-\alpha A}\,\delta A\}}{\mathrm{tr}\{\mathrm{e}^{-\alpha A}\}}\right]\\ =& -f(A)\left[\int_0^1 \mathrm{d}\gamma\ \mathrm{e}^{(1-\gamma)\alpha A}\ (\beta H_{\rm SB} + \delta\Lambda)\ \mathrm{e}^{-(1-\gamma)\alpha A} - \frac{\mathrm{tr}\{\mathrm{e}^{-\alpha A}\,(\beta H_{\rm SB} + \delta\Lambda)\}}{\mathrm{tr}\{\mathrm{e}^{-\alpha A}\}}\right].\end{aligned} \tag{6.13}$$

Now due to the constraint given by Eq. (6.10), we need to satisfy $\mathrm{tr}_{\rm B}\{\delta f\} = 0$, which is

$$\begin{aligned}&\mathrm{tr}_{\rm B}\left\{\left(\rho_{\rm B}^{(\rm th)} \otimes \rho_{\rm S}\right)\int_0^1 \mathrm{d}\gamma\,\mathrm{e}^{(1-\gamma)(\beta H_{\rm B}+\Lambda)}(\beta H_{\rm SB} + \delta\Lambda)\mathrm{e}^{-(1-\gamma)(\beta H_{\rm B}+\Lambda)}\right\}\\ =&\rho_{\rm S}\frac{\mathrm{tr}\{\mathrm{e}^{-\beta H_{\rm B}-\Lambda}\,(\beta H_{\rm SB} + \delta\Lambda)\}}{\mathrm{tr}\{\mathrm{e}^{-\beta H_{\rm B}-\Lambda}\}}\\ =&\rho_{\rm S}\left(\mathrm{tr}\left\{\left(\rho_{\rm B}^{(\rm th)} \otimes \rho_{\rm S}\right)\beta H_{\rm SB}\right\} + \mathrm{tr}_{\rm S}\left\{\rho_{\rm S}\delta\Lambda\right\}\right).\end{aligned} \tag{6.14}$$

On collecting terms with $H_{\rm SB}$ on one side and $\delta\Lambda$ on the other, we have

$$\begin{aligned}&\rho_{\rm S}\left(\int_0^1 \mathrm{d}\gamma\ \mathrm{e}^{(1-\gamma)\Lambda}\delta\Lambda\mathrm{e}^{-(1-\gamma)\Lambda} - \mathrm{tr}_{\rm S}\{\rho_{\rm S}\delta\Lambda\}\right)\\ =&\rho_{\rm S}\left(\mathrm{tr}\left\{\left(\rho_{\rm B}^{(\rm th)} \otimes \rho_{\rm S}\right)\beta H_{\rm SB}\right\} - \int_0^1 \mathrm{d}\gamma\ \mathrm{e}^{(1-\gamma)\Lambda}\beta\,\mathrm{tr}_{\rm B}\{\rho_{\rm B}H_{\rm SB}\}\mathrm{e}^{-(1-\gamma)\Lambda}\right).\end{aligned} \tag{6.15}$$

One sees that a solution to this equation is to require

$$\delta\Lambda = -\mathrm{tr}_{\mathrm{B}}\left\{\beta\rho_{\mathrm{B}}^{(\mathrm{th})}H_{\mathrm{SB}}\right\}, \tag{6.16}$$

which gives us a relation between $\delta\Lambda$ to H_{SB}, where

$$\delta\Lambda \equiv \sum_i \delta\lambda_i\sigma_i. \tag{6.17}$$

We now would like to work out the specific form of the maximum entropy state for Jaynes-Cummings model, that is, Eq. (6.4), up to first order in J. For that purpose, we use the variation of the exponential of an operator again, which enables us to arrive at

$$\rho_{\mathrm{SB}}^{(\mathrm{ME})} = \rho_S \otimes \rho_{\mathrm{B}}^{(\mathrm{th})}\left[1 - \left(\int_0^1 \mathrm{d}\gamma\ \mathrm{e}^{(1-\gamma)A}\delta A\mathrm{e}^{-(1-\gamma)A} - \frac{\mathrm{tr}\{\mathrm{e}^{-A}\delta A\}}{\mathrm{tr}\{\mathrm{e}^{-A}\}}\right)\right], \tag{6.18}$$

where (we absorb the α into the A here for notational simplicity)

$$A = \beta\hbar\omega a^\dagger a + \sum_i \lambda_i\sigma_i, \tag{6.19}$$

and

$$\delta A = \beta J(\sigma_- a^\dagger + \sigma_+ a) + \sum_i \delta\lambda_i\sigma_i. \tag{6.20}$$

Now the second term in the integral in Eq. (6.18) simplifies to

$$\begin{aligned}\frac{\mathrm{tr}\{\mathrm{e}^{-A}\delta A\}}{\mathrm{tr}\{\mathrm{e}^{-A}\}} &= \beta J\mathrm{tr}\{\rho_{\mathrm{S}}\sigma\}\underbrace{\mathrm{tr}\left\{\rho_{\mathrm{B}}^{(\mathrm{th})}a^\dagger\right\}}_{=0} + \beta J\mathrm{tr}\left\{\rho_{\mathrm{S}}\sigma^\dagger\right\}\underbrace{\mathrm{tr}\left\{\rho_{\mathrm{B}}^{(\mathrm{th})}a\right\}}_{=0} + \sum_i \delta\lambda_i s_i \\ &= \sum_i \delta\lambda_i s_i. \end{aligned} \tag{6.21}$$

Due to the constraint in Eq. (6.15), we also have

$$\int_0^1 \mathrm{d}\gamma\ \mathrm{e}^{(1-\gamma)\Lambda}\delta\Lambda\mathrm{e}^{-(1-\gamma)\Lambda} = \mathrm{tr}\{\rho_{\mathrm{S}}\delta\Lambda\} = \sum_i \delta\lambda_i s_i, \tag{6.22}$$

so that this term takes care of the term in Eq. (6.21) since they occur with opposite

signs in Eq. (6.18). Now for the remaining part, we have

$$\mathrm{e}^{(1-\gamma)A}\sigma_- a^\dagger \mathrm{e}^{-(1-\gamma)A} = a^\dagger \mathrm{e}^{(1-\gamma)\beta\hbar\omega}\mathrm{e}^{(1-\gamma)\sum\lambda_i\sigma_i}\sigma_-\mathrm{e}^{-(1-\gamma)\sum\lambda_i\sigma_i}, \tag{6.23}$$

where we used $f(a^\dagger a)a^\dagger = a^\dagger f(a^\dagger a + 1)$, and

$$\mathrm{e}^{(1-\gamma)A}\sigma_+ a \mathrm{e}^{-(1-\gamma)A} = a\mathrm{e}^{-(1-\gamma)\beta\hbar\omega}\mathrm{e}^{(1-\gamma)\sum\lambda_i\sigma_i}\sigma_+\mathrm{e}^{-(1-\gamma)\sum\lambda_i\sigma_i}, \tag{6.24}$$

where we used $f(a^\dagger a)a = af(a^\dagger a - 1)$.

Hence, the maximum entropy state for the Jaynes-Cummings model is given by

$$\begin{aligned}\rho_{\mathrm{SB}}^{(\mathrm{ME})} =& \rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})} - \left(\int_0^1 \mathrm{d}\gamma \mathrm{e}^{(1-\gamma)\beta\hbar\omega}\rho_{\mathrm{S}}\mathrm{e}^{(1-\gamma)\Lambda}\sigma_-\mathrm{e}^{-(1-\gamma)\Lambda}\right)\otimes\left(\rho_{\mathrm{B}}^{(\mathrm{th})}a^\dagger\right)\\ &- \left(\int_0^1 \mathrm{d}\gamma \mathrm{e}^{-(1-\gamma)\beta\hbar\omega}\rho_{\mathrm{S}}\mathrm{e}^{(1-\gamma)\Lambda}\sigma_+\mathrm{e}^{-(1-\gamma)\Lambda}\right)\otimes\left(\rho_{\mathrm{B}}^{(\mathrm{th})}a\right).\end{aligned} \tag{6.25}$$

In order to work out explicitly the terms in the integral, some relations that are useful in what follows are:

$$\rho_{\mathrm{S}} = \frac{1}{2}\left(\mathbb{1} + \boldsymbol{s}\cdot\boldsymbol{\sigma}\right) = \frac{\mathrm{e}^{-\boldsymbol{\lambda}\cdot\boldsymbol{\sigma}}}{Z(\boldsymbol{\lambda})}, \tag{6.26}$$

where $\boldsymbol{s} = \langle\boldsymbol{\sigma}\rangle = s\boldsymbol{n}$ ($\boldsymbol{n}$ being a unit vector), the Lagrange multiplier given by

$$\boldsymbol{\lambda} = -(\tanh^{-1} s)\boldsymbol{n}, \tag{6.27}$$

and the partition function given by

$$Z(\boldsymbol{\lambda}) = 2\cosh|\boldsymbol{\lambda}| = \frac{2}{\sqrt{1-s^2}}. \tag{6.28}$$

Now, let $\mathrm{e}^{(1-\gamma)\Lambda} = \mu + \boldsymbol{\nu}\cdot\boldsymbol{\sigma}$. Then its inverse is given by

$$\mathrm{e}^{-(1-\gamma)\Lambda} = \frac{\mu - \boldsymbol{\nu}\cdot\boldsymbol{\sigma}}{\mu^2 - \boldsymbol{\nu}\cdot\boldsymbol{\nu}}. \tag{6.29}$$

With this, we can now work out

$$\begin{aligned} \mathrm{e}^{(1-\gamma)\Lambda}\sigma_{\pm}\mathrm{e}^{-(1-\gamma)\Lambda} &= \frac{1}{\mu^2-\boldsymbol{\nu}\cdot\boldsymbol{\nu}}(\mu+\boldsymbol{\nu}\cdot\boldsymbol{\sigma})\sigma_{\pm}(\mu-\boldsymbol{\nu}\cdot\sigma) \\ &= \frac{1}{\mu^2-\boldsymbol{\nu}\cdot\boldsymbol{\nu}}\left(\mu^2\sigma_{\pm}+\mu\boldsymbol{\nu}\cdot[\boldsymbol{\sigma},\sigma_{\pm}]-\boldsymbol{\nu}\cdot\boldsymbol{\sigma}\sigma_{\pm}\boldsymbol{\sigma}\cdot\boldsymbol{\nu}\right), \end{aligned}$$

where $[A,B]=AB-BA$ is the commutator of A and B. The γ-dependence in μ and in $\boldsymbol{\nu}$ are found to be

$$\mu=\frac{1}{2}\mathrm{tr}\left\{\mathrm{e}^{(1-\gamma)\boldsymbol{\lambda}\cdot\boldsymbol{\sigma}}\right\}=\cosh((1-\gamma)\lambda)=\frac{1}{\sqrt{1-\tanh^2\left((1-\gamma)\lambda\right)}}, \tag{6.30}$$

and

$$\boldsymbol{\nu}=\frac{\sinh\left((1-\gamma)\lambda\right)}{\lambda}\boldsymbol{\lambda}=-\sinh\left((1-\gamma)\lambda\right)\boldsymbol{n}. \tag{6.31}$$

In conclusion, the maximum entropy system-bath state for the Jaynes-Cummings model is given by

$$\begin{aligned} \rho_{\mathrm{SB}}^{(\mathrm{ME})} =& \rho_{\mathrm{S}}\otimes\rho_{\mathrm{B}}^{(\mathrm{th})}-\beta J\left(\rho_{\mathrm{S}}(A_+\sigma_- - B_+\boldsymbol{n}\cdot[\boldsymbol{\sigma},\sigma_-]-C_+\boldsymbol{n}\cdot\boldsymbol{\sigma}\sigma_-\boldsymbol{\sigma}\cdot\boldsymbol{n})\right)\otimes\left(\rho_{\mathrm{B}}^{(\mathrm{th})}a^{\dagger}\right) \\ & -\beta J\left(\rho_{\mathrm{S}}(A_-\sigma_+ - B_-\boldsymbol{n}\cdot[\boldsymbol{\sigma},\sigma_+]-C_-\boldsymbol{n}\cdot\boldsymbol{\sigma}\sigma_+\boldsymbol{\sigma}\cdot\boldsymbol{n})\right)\otimes\left(\rho_{\mathrm{B}}^{(\mathrm{th})}a\right), \end{aligned} \tag{6.32}$$

with the coefficient $A_{\pm}$, $B_{\pm}$, and $C_{\pm}$ given by Eq. (6.33) below.

$$\begin{aligned} A_{\pm} &= \int_0^1 \mathrm{d}\gamma \mathrm{e}^{\pm\gamma\beta\hbar\omega}\cosh^2(\gamma\lambda) \\ &= \pm\frac{-(\beta\hbar\omega)^2-2(\mathrm{e}^{\pm\beta\hbar\omega}-1)\lambda^2+\beta\hbar\omega\mathrm{e}^{\pm\beta\hbar\omega}\cosh\lambda(\beta\hbar\omega\cosh\lambda\mp2\lambda\sinh\lambda)}{(\beta\hbar\omega)^3-4\beta\hbar\omega\lambda^2}, \\ B_{\pm} &= \int_0^1 \mathrm{d}\gamma \mathrm{e}^{\pm\gamma\beta\hbar\omega}\frac{\sinh(2\gamma\lambda)}{2} \\ &= \frac{2\lambda+\mathrm{e}^{\pm\beta\hbar\omega}\left(-2\lambda\cosh(2\lambda)\pm\beta\hbar\omega\sinh(2\lambda)\right)}{2(\beta\hbar\omega)^2-4\lambda^2}, \\ C_{\pm} &= \int_0^1 \mathrm{d}\gamma \mathrm{e}^{\pm\gamma\beta\hbar\omega}\sinh^2(\gamma\lambda) \\ &= \pm\frac{2(\mathrm{e}^{\pm\beta\hbar\omega}-1)\lambda^2+\beta\hbar\omega\mathrm{e}^{\pm\beta\hbar\omega}\sinh\lambda(\beta\hbar\omega\sinh\lambda\mp2\lambda\cosh\lambda)}{(\beta\hbar\omega)^3-4\beta\hbar\omega\lambda^2}. \end{aligned} \tag{6.33}$$

Subsequently, the state as given in Eq. (6.32) should be used as the initial system-bath state in any quantum mechanical dynamical equations for predictions on the

future evolution of the system-bath state, rather than the naive uncorrelated product state given by $\rho_{\rm S}\otimes\rho_{\rm B}^{\rm (th)}$.

Now, one remark is in order: It is not so obvious that the above approach using $\delta(\mathrm{e}^A)$ gives a positive operator for $\rho_{\rm SB}$, i.e., positive up to $O(J^2)$. Here, we also give a different approach to this problem, starting from a different (exact) identity:

$$\exp\left(\mathrm{e}^{-B}A\mathrm{e}^{B}\right)=\mathrm{e}^{-B}\mathrm{e}^{A}\mathrm{e}^{B}, \tag{6.34}$$

easily proven using the Taylor expansion of the exponential, and the fact that $\mathrm{e}^{B}\mathrm{e}^{-B}=1$. For us (note that now with $A\equiv-\beta\hbar\omega a^\dagger a-\boldsymbol{\lambda}\cdot\boldsymbol{\sigma}$), we have

$$\frac{\mathrm{e}^{A}}{\mathrm{tr}\{\ldots\}}=\rho_{\rm S}\otimes\rho_{\rm B}^{\rm (th)}=\frac{\exp\left(-\beta\hbar\omega a^\dagger a-\boldsymbol{\lambda}\cdot\boldsymbol{\sigma}\right)}{\mathrm{tr}\{\ldots\}}, \tag{6.35}$$

and we need to find an operator B such that

$$\mathrm{e}^{-B}A\mathrm{e}^{B}\stackrel{!}{=}A-\beta J(\sigma_-a^\dagger+\sigma_+a). \tag{6.36}$$

Notice that in Eq. (6.36), there is no $\delta\Lambda$ term as in this case, $\delta\Lambda=0$, since $\mathrm{tr}_{\rm B}\left\{\rho_{\rm B}^{\rm (th)}H_{\rm SB}\right\}=0$ for this Jaynes-Cummings interaction Hamiltonian; see Eq. (6.16). With this operator B at hand, one can then write down the maximum entropy state as

$$\rho_{\rm SB}^{\rm (ME)}=\frac{\exp\left(\mathrm{e}^{-B}A\mathrm{e}^{B}\right)}{\mathrm{tr}\{\ldots\}}=\frac{\mathrm{e}^{-B}\mathrm{e}^{A}\mathrm{e}^{B}}{\mathrm{tr}\{\mathrm{e}^{A}\}}. \tag{6.37}$$

The last expression makes the positivity of $\rho_{\rm SB}$ manifest. Now the operator B should be something small, in particular, proportional to $J/\hbar\omega$. We can do everything to linear order in $J/\hbar\omega$, so that

$$\begin{aligned}\rho_{\rm SB}^{\rm (ME)}=\mathrm{e}^{-B}\rho_{\rm S}\otimes\rho_{\rm B}^{\rm (th)}\mathrm{e}^{B}=&(1-B)\rho_{\rm S}\otimes\rho_{\rm B}^{\rm (th)}(1+B)+O((J/\hbar\omega)^2)\\ =&\rho_{\rm S}\otimes\rho_{\rm B}^{\rm (th)}-\left[B,\rho_{\rm S}\otimes\rho_{\rm B}^{\rm (th)}\right]+O((J/\hbar\omega)^2),\end{aligned} \tag{6.38}$$

where the operator B satisfies

$$\mathrm{e}^{-B}A\mathrm{e}^{B}=A-[B,A]+O((J/\hbar\omega)^2)=A-\beta J(\sigma_-a^\dagger+\sigma_+a), \tag{6.39}$$

which is to say that

$$[B, A] = \beta J(\sigma_- a^\dagger + \sigma_+ a) + O((J/\hbar\omega)^2). \tag{6.40}$$

This last equation determines B, at least to linear order in $J/\hbar\omega$. And our task is to find the explicit form of B for this case.

Given the form of Eq. (6.40), one can try the following form for B,

$$B = \boldsymbol{\alpha}_+ \cdot \boldsymbol{\sigma} a^\dagger + \boldsymbol{\alpha}_- \cdot \boldsymbol{\sigma} a + u_+ \sigma_- a^\dagger + u_- \sigma_+ a, \tag{6.41}$$

for $\boldsymbol{\alpha}_\pm$ and $u_\pm$ to be determined by Eq. (6.40). It is now convenient to write $\sigma_\pm$ as

$$\sigma_\pm = \boldsymbol{\mu}_\pm \cdot \boldsymbol{\sigma}, \qquad \text{with} \quad \boldsymbol{\mu}_\pm \widehat{=} \frac{1}{2} \begin{pmatrix} 1 \\ \pm \mathrm{i} \\ 0 \end{pmatrix}. \tag{6.42}$$

Also, let $a_+ \equiv a^\dagger$ and $a_- \equiv a$, much like the raising and lowering operators for the spin. Some useful identities are (with $A \equiv -\beta\hbar\omega a^\dagger a - \boldsymbol{\lambda} \cdot \boldsymbol{\sigma}$)

$$\begin{aligned} [\boldsymbol{\alpha}_\pm \cdot \boldsymbol{\sigma} a_\pm, A] &= a_\pm [\pm\beta\hbar\omega \boldsymbol{\alpha}_\pm + \mathrm{i}2(\boldsymbol{\lambda} \times \boldsymbol{\alpha}_\pm)] \cdot \boldsymbol{\sigma} \ ; \\ [\sigma_\mp a_\pm, A] &= a_\pm (\pm\beta\hbar\omega \sigma_\mp + \mathrm{i}2(\boldsymbol{\lambda} \times \boldsymbol{\mu}_\mp) \cdot \boldsymbol{\sigma}) \ ; \\ \text{and} \ \ 2\boldsymbol{\lambda} \times \boldsymbol{\mu}_\mp &\widehat{=} \begin{pmatrix} \pm \mathrm{i}\lambda_z \\ \lambda_z \\ \mp \mathrm{i}\lambda_x - \lambda_y \end{pmatrix}. \end{aligned}$$

Putting all these into Eq. (6.40), we see that a simple solution is to have

$$u_\pm = \pm J/\hbar\omega,$$

and

$$\pm\beta\hbar\omega \boldsymbol{\alpha}_\pm + \mathrm{i}2(\boldsymbol{\lambda} \times \boldsymbol{\alpha}_\pm) \pm \mathrm{i}2(\boldsymbol{\lambda} \times \boldsymbol{\mu}_\mp)(J/\hbar\omega) = 0.$$

The set of equations involving $\boldsymbol{\alpha}_\pm$ can be rewritten as matrix equations

$$M_\pm \boldsymbol{\alpha}_\pm = \mp \mathrm{i}2(\boldsymbol{\lambda} \times \boldsymbol{\mu}_\mp)(J/\hbar\omega), \tag{6.43}$$

with $M_\pm$ given by

$$M_\pm \widehat{=} \begin{pmatrix} \pm\beta\hbar\omega & -\mathrm{i}2\lambda_z & \mathrm{i}2\lambda_y \\ \mathrm{i}2\lambda_z & \pm\beta\hbar\omega & -\mathrm{i}2\lambda_x \\ -\mathrm{i}2\lambda_y & \mathrm{i}2\lambda_x & \pm\beta\hbar\omega \end{pmatrix}. \tag{6.44}$$

M is invertible except in accidental situations [$\det(M) \sim (\beta\hbar\omega)^3 + (\lambda$ terms), which is never zero unless $\lambda \sim \beta\hbar\omega$], so

$$\boldsymbol{\alpha}_\pm = (M_\pm)^{-1}\big(\mp\mathrm{i}2(\boldsymbol{\lambda} \times \boldsymbol{\mu}_\mp)(J/\hbar\omega)\big); \tag{6.45}$$

one can then work out the explicit expressions for $\boldsymbol{\alpha}_\pm$ in terms of $\boldsymbol{\lambda}$. Once this is done, one can then put B into Eq. (6.38) to find out the maximum entropy state for the system-bath state.

Here, we mention some special cases to check whether both approaches give the same results. If ρ_S is a pure state, then the general form using either approach should simplify to $\rho_\mathrm{S} \otimes \rho_\mathrm{B}^{(\mathrm{th})}$, a product form as demanded by the unit-purity nature of ρ_S. Indeed for $\rho_\mathrm{S} = (1/2)(\mathbb{1} + \sigma_z)$, we have

$$\boldsymbol{\lambda} = \lambda\hat{e}_z, \quad \text{with} \quad \lambda \to \infty. \tag{6.46}$$

We can solve the equations for $\boldsymbol{\alpha}_\pm$ with λ as a parameter, and then take $\lambda \to \infty$ at the end. This gives

$$\boldsymbol{\alpha}_\pm \widehat{=} -\frac{1}{2}\frac{J}{\hbar\omega}\begin{pmatrix} \pm 1 \\ \mathrm{i} \\ 0 \end{pmatrix},$$

and with $u_\pm$, this gives $B = 0$. This means that

$$\rho_\mathrm{SB}^{(\mathrm{ME})} = \rho_\mathrm{S} \otimes \rho_\mathrm{B}^{(\mathrm{th})}, \tag{6.47}$$

as one is expecting.

For the previous approach, since $\rho_\mathrm{S}\sigma_- = 0$, we have

$$\rho_\mathrm{SB}^{(\mathrm{ME})} = \rho_\mathrm{S} \otimes \rho_\mathrm{B}^{(\mathrm{th})} - \left(\int_0^1 \mathrm{d}\gamma \mathrm{e}^{-(1-\gamma)\beta\hbar\omega}\rho_\mathrm{S}\mathrm{e}^{(1-\gamma)\Lambda}\sigma_+\mathrm{e}^{-(1-\gamma)\Lambda}\right) \otimes \left(\rho_\mathrm{B}^{(\mathrm{th})}a\right).$$

In this case,

$$\begin{aligned} \mathrm{e}^{(1-\gamma)\Lambda}\sigma_+\mathrm{e}^{-(1-\gamma)\Lambda} &= \frac{1}{\mu^2-\boldsymbol{\nu}\cdot\boldsymbol{\nu}}(\mu+\boldsymbol{\nu}\cdot\boldsymbol{\sigma})\sigma_+(\mu-\boldsymbol{\nu}\cdot\sigma) \\ &= \frac{1}{\mu^2-\boldsymbol{\nu}\cdot\boldsymbol{\nu}}\left(\mu^2\sigma_+ + \mu\nu_z[\sigma_z,\sigma_+] - \nu_z^2\sigma_z\sigma_+\sigma_z\right) \\ &= \frac{\mu+\nu_z}{\mu-\nu_z}\sigma_+. \end{aligned}$$

Then, we have,

$$\begin{aligned} &\int_0^1 \mathrm{d}\gamma \mathrm{e}^{-(1-\gamma)\beta\hbar\omega}\rho_{\mathrm{S}}\mathrm{e}^{(1-\gamma)\Lambda}\sigma_+\mathrm{e}^{-(1-\gamma)\Lambda} \\ =&\sigma_+\int_0^1 \mathrm{d}\gamma \mathrm{e}^{-(1-\gamma)\beta\hbar\omega}\mathrm{e}^{-2(1-\gamma)\lambda} = 0 \end{aligned} \tag{6.48}$$

as λ approaches infinity, and we similarly recover the expression

$$\rho_{\mathrm{SB}}^{(\mathrm{ME})} = \rho_{\mathrm{S}}\otimes\rho_{\mathrm{B}}^{(\mathrm{th})}.$$

The same product state is obtained if $\rho_{\mathrm{S}} = (1/2)(\mathbb{1}-\sigma_z) = |1\rangle\langle 1|$. This is but just a consistency check, since a pure ρ_{S} can never result from a rational state reconstruction from finite data.

We also work out the example of a completely mixed state on the system below. For $\rho_{\mathrm{S}} = \mathbb{1}/2$, we have $\boldsymbol{\lambda} = 0$ so that $\boldsymbol{\alpha}_\pm = 0$ and $B = (J/\hbar\omega)(\sigma_- a^\dagger - \sigma_+ a)$. This gives

$$\rho_{\mathrm{SB}}^{(\mathrm{ME})} = \mathrm{e}^A - \left[B,\mathrm{e}^A\right] = \rho_{\mathrm{S}}\otimes\rho_{\mathrm{B}}^{(\mathrm{th})}\Big[1-(J/\hbar\omega)(\sigma_- a^\dagger(e^{\beta\hbar\omega}-1) - \sigma_+ a(e^{-\beta\hbar\omega}-1)\Big]. \tag{6.49}$$

For the first approach, if $s = 0$, then the terms involving $B_\pm$ and $C_\pm$ vanish, and the integral involving $A_\pm$ is trivial. We thus have,

$$\rho_{\mathrm{SB}}^{(\mathrm{ME})} = \rho_{\mathrm{S}}\otimes\rho_{\mathrm{B}}^{(\mathrm{th})}\Big[1-(J/\hbar\omega)(\sigma_- a^\dagger(e^{\beta\hbar\omega}-1) - \sigma_+ a(e^{-\beta\hbar\omega}-1)\Big], \tag{6.50}$$

the same expression as Eq. (6.49).

Now for one last check, take $\rho_{\mathrm{S}} = \frac{1}{2}(\mathbb{1}+s_z\sigma_z)$, with $-1 < s_z < 1$, then, for both

approaches, we have

$$\rho_{\mathrm{SB}}^{(\mathrm{ME})} = \rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})} - \frac{\beta J}{(\beta\hbar\omega - 2\lambda)(\mathrm{e}^{\lambda} + \mathrm{e}^{-\lambda})} \times \left(\mathrm{e}^{\lambda} \left(\mathrm{e}^{-\beta\hbar\omega - 2\lambda} - 1 \right) \sigma_{-} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})} a^{\dagger} + \mathrm{e}^{-\lambda} \left(1 - \mathrm{e}^{-\beta\hbar\omega + 2\lambda} - 1 \right) \sigma_{+} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})} a \right), \tag{6.51}$$

with $\lambda = \tanh^{-1} s_z$.

From this expression, one observes that the maximum-entropy state differs from the uncorrelated product state by a term linear in βJ. Indeed the correction to the uncorrelated state can always be bounded by a term proportional to J. If the experiment is sensitive to effects of this size, the use of an initially uncorrelated state becomes inappropriate. Note that when $J = 0$, we have

$$\rho_{\mathrm{SB}}^{(\mathrm{ME})} = \exp\left(-\beta\hbar\omega a^{\dagger} a - \boldsymbol{\lambda} \cdot \boldsymbol{\sigma} \right) / \mathrm{tr}\{\ldots\} = \rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})}. \tag{6.52}$$

This recovers the case where the system and the bath are put into contact only at $t = 0$, prior to which J can be treated as 0.

After this example of Jaynes-Cummings model, we now turn back to a general scenario, where one might lack macroscopic information about the system and bath. In this case, there is no $\mu g(\rho)$ term, and we maximize

$$\tilde{S}(\rho) \equiv S(\rho) - \sum_i \lambda_i (\mathrm{tr}\{\rho \Pi_i\} - p_i) - \nu(\mathrm{tr}\{\rho\} - 1). \tag{6.53}$$

Stationarity requires

$$0 = \delta\tilde{S} = -\mathrm{tr}\left\{ \delta\rho \left(\log\rho + 1 + \sum_i \lambda_i \Pi_i + \nu \right) \right\} \tag{6.54}$$

for arbitrary $\delta\rho$, which is satisfied for a ρ given by

$$\rho_{\mathrm{SB}}^{(\mathrm{ME})} = \exp\left(-\sum_i \lambda_i \Pi_i \right) / \mathrm{tr}\{\ldots\}. \tag{6.55}$$

Noting that the POM outcomes Π_i are system-only operators, this simplifies to

$$\rho_{\mathrm{SB}}^{(\mathrm{ME})} = \rho_{\mathrm{S}} \otimes \frac{1}{d_{\mathrm{B}}} \mathbb{1}_B, \tag{6.56}$$

where d_{B} is the dimension of the bath, and

$$\rho_{\mathrm{S}} = \exp\left(- \sum_i \lambda_i \Pi_i \right) / \mathrm{tr}\{\ldots\} \tag{6.57}$$

denotes the system-only state. Note that it is also the maximum likelihood state. ρ_{S} is the usual state one would have deduced from the tomography data, with λ_is determined only by the constraint Eq. (6.2) [BD00]. The maximally mixed state on the bath appears naturally from this approach, corresponding to our complete ignorance of the bath state. This complete ignorance also motivate the Bayesian mean state as we will see later in the chapter. Partial information about the available energy in the bath, for example, extremely high-energy bath states are unlikely given the typical energy available in the experiment, can be incorporated by modifying the available Hilbert space of the bath so that the identity I_{B} has support only on states below some energy upper-bound. A smooth weighting function, rather than a sharp cutoff, can also be put in by hand. The precise way this cutoff is accomplished cannot matter—at least on the level of precision available in the experiment—to the dynamics of the system itself.

For a general H_{SB}, provided that the interaction is weak, the maximum entropy state is always close to $\rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})}$. When the interaction is strong, the maximum entropy state will possess nontrivial correlations between the system and the bath. Even in the case of weak interaction, as we have seen, there will be deviations of linear order in H_{SB}, like the $J/\hbar\omega$ in the Jaynes-Cummings model case. In a general situation, the deviation is of the order of the norm of H_{SB}. In some contexts, this can become important if the effects explored are also small, comparable to this order of magnitude.

In a general situation, one can show that

$$\left|\left| \rho_{\mathrm{SB}}^{(\mathrm{ME})} - \rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})} \right|\right| \leq 4\beta ||H_{\mathrm{SB}}||, \tag{6.58}$$

where $\|A\|$ is taken to denote the trace norm of the operator A. To prove this, one recalls that the maximum entropy state is given by

$$\rho^{(\mathrm{ME})} = \exp\Big(-\beta H - \Lambda\Big)/\mathrm{tr}\{\ldots\}, \tag{6.59}$$

and the uncorrelated state given by

$$\rho^{(\mathrm{sep})} = \rho_{\mathrm{S}} \otimes \frac{\exp(-\beta H_{\mathrm{B}})}{\mathrm{tr}\{\ldots\}} = \rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})}. \tag{6.60}$$

Again, this is very deceptive as the Lagrange multipliers in the two states are really different quantities, i.e., governed by different constraint equations, because of the presence or absence of H_{SB}; see Eq. (6.10). To be careful then, let us write down three states that are potentially useful to consider:

$$\begin{aligned}
\rho_1 &= \frac{\exp\Big(-\beta(H_{\mathrm{B}} + H_{\mathrm{SB}}) - \Lambda'\Big)}{\mathrm{tr}\{\ldots\}}, \quad \text{where} \quad \frac{\mathrm{tr}\Big\{\mathrm{e}^{-\beta(H_{\mathrm{B}}+H_{\mathrm{SB}})-\Lambda'}\Pi_i\Big\}}{\mathrm{tr}\Big\{\mathrm{e}^{-\beta(H_{\mathrm{B}}+H_{\mathrm{SB}})-\Lambda'}\Big\}} = p_i \quad \forall i. \\
\rho_2 &= \frac{\exp\Big(-\beta H_{\mathrm{B}} - \Lambda'\Big)}{\mathrm{tr}\{\ldots\}}, \quad \text{where} \quad \frac{\mathrm{tr}\Big\{\mathrm{e}^{-\beta(H_{\mathrm{B}}+H_{\mathrm{SB}})-\Lambda'}\Pi_i\Big\}}{\mathrm{tr}\Big\{\mathrm{e}^{-\beta(H_{\mathrm{B}}+H_{\mathrm{SB}})-\Lambda'}\Big\}} = p_i \quad \forall i. \\
\rho_3 &= \frac{\rho_{\mathrm{S}} \otimes \exp(-\beta H_{\mathrm{B}})}{\mathrm{tr}\{\ldots\}}, \quad \text{where} \quad \frac{\mathrm{tr}\Big\{\mathrm{e}^{-\Lambda}\Pi_i\Big\}}{\mathrm{tr}\Big\{\mathrm{e}^{-\Lambda}\Big\}} = p_i \quad \forall i.
\end{aligned}$$

Here, ρ_1 and ρ_2 have the same Lagrange-multipler values, despite H_{SB} not entering ρ_2 itself, which means that ρ_2 itself in fact does not satisfy the specified constraints, but instead have its parameters satisfying the same equations as those for ρ_1. ρ_3 has the actual Lagrange-multipler values when $H_{\mathrm{SB}} = 0$. ρ_1 (the maximum entropy state for $H_{\mathrm{SB}} \neq 0$) and ρ_3 (the product state when $H_{\mathrm{SB}} = 0$) are what we want to compare, while ρ_2 plays a convenient intermediate role. Note that, however, all these three states have the same temperature parameter β, since it is fixed by the external temperature bath.

Now we are interested in $\|\rho_1 - \rho_3\|$ in the trace norm sense, and

$$\|\rho_1 - \rho_3\| = \|\rho_1 - \rho_2 + \rho_2 - \rho_3\| \leq \|\rho_1 - \rho_2\| + \|\rho_2 - \rho_3\| \tag{6.61}$$

by the triangle inequality. For the first term, again, given the physical context, we

expect $\|H_{\rm SB}\| \ll \|H_{\rm B}\|$. Let $\alpha \equiv \|H_{\rm B}\|$, $A \equiv \frac{1}{\alpha}(\beta H_{\rm B} + \Lambda')$, and $a \equiv \frac{1}{\alpha}\beta H_{\rm SB}$. Then, we can write

$$\rho_1 - \rho_2 = f(A+a) - f(A), \quad \text{where} \quad f(\cdot) = \frac{\mathrm{e}^{-\alpha(\cdot)}}{\mathrm{tr}\{\mathrm{e}^{-\alpha(\cdot)}\}}. \tag{6.62}$$

Here, β and λ_i's inside Λ' are treated as constants determined by the constraint equations above. The term a is small compared to A, and $A \sim O(1)$. Similarly, we look for the variation of the exponential of an operator with

$$\delta f = -\beta f(A)\left[\int_0^1 \mathrm{d}\gamma\ \mathrm{e}^{(1-\gamma)\alpha A}\, H_{\rm SB}\, \mathrm{e}^{-(1-\gamma)\alpha A} - \frac{\mathrm{tr}\{\mathrm{e}^{-\alpha A}\, H_{\rm SB}\}}{\mathrm{tr}\{\mathrm{e}^{-\alpha A}\}}\right]. \tag{6.63}$$

It then follows that

$$\begin{aligned}
\|\delta f\| &= \left\| \beta f(A)\left[\int_0^1 \mathrm{d}\gamma\ \mathrm{e}^{(1-\gamma)\alpha A}\, H_{SB}\, \mathrm{e}^{-(1-\gamma)\alpha A} - \frac{\mathrm{tr}\{\mathrm{e}^{-\alpha A}\, H_{SB}\}}{\mathrm{tr}\{\mathrm{e}^{-\alpha A}\}}\right]\right\| \\
&\le \beta\, \|f(A)\| \left[\int_0^1 \mathrm{d}\gamma \left\|\mathrm{e}^{(1-\gamma)\alpha A} H_{SB} \mathrm{e}^{-(1-\gamma)\alpha A}\right\| + \frac{\left|\mathrm{tr}\{\mathrm{e}^{-\alpha A}\, H_{SB}\}\right|}{\mathrm{tr}\{\mathrm{e}^{-\alpha A}\}}\right] \\
&\le \beta\left[\int_0^1 \mathrm{d}\gamma \left\|H_{SB} \mathrm{e}^{-(1-\gamma)\alpha A} \mathrm{e}^{(1-\gamma)\alpha A}\right\| + \|H_{SB}\|\right] \\
&= 2\beta\, \|H_{SB}\|\,.
\end{aligned}$$

Similarly, one can show that

$$||\rho_2 - \rho_3|| \le 2\, ||\delta\Lambda||\,, \tag{6.64}$$

where we remind ourselves that

$$\delta\Lambda = -\mathrm{tr}_{\rm B}\left\{\beta \rho_{\rm B}^{(\rm th)} H_{\rm SB}\right\}. \tag{6.65}$$

Hence, one sees that in general, Eq. (6.58) is true. If the interaction is rather weak, which is usually the case, then approximating the initial state by $\rho_{\rm S} \otimes \rho_{\rm B}^{(\rm th)}$ could be a good starting point. However, one should be very careful of this nontrivial difference. If the effects explored are also small, compared with the norm of $H_{\rm SB}$, then using $\rho_{\rm S} \otimes \rho_{\rm B}^{(\rm th)}$ is hardly justifiable.

6.4 The Bayesian mean state

A different viewpoint from the maximum entropy approach is the following: In the set of ρ_{SB} consistent with the tomographic constraint, motivated by our complete ignorance on the bath as well as our inability to control it in any way, we should average over all such ρ_{SB}s. In any particular round of one realization, the bath state would be slightly different due to our inability to control it microscopically in any way. The average value of any measurement results that is produced over many such realizations should then be consistent with some form of averaged state over the bath. In this context, the Bayesian mean state appears naturally, that is we add all such ρ_{SB}s in the set consistent with the tomographic constraint on the system according to some reasonable weight, reporting

$$\overline{\rho_{\mathrm{SB}}} = \frac{\int (\mathrm{d}\rho_{\mathrm{B}}) L(D|\rho_{\mathrm{SB}}) \rho_{\mathrm{SB}}}{\int (\mathrm{d}\rho_{\mathrm{B}}) L(D|\rho_{\mathrm{SB}})} \tag{6.66}$$

as our estimator for the system-bath state. The integration over $(\mathrm{d}\rho_{\mathrm{B}})$ means the averaging is done on the bath only, and $L(D|\rho_{\mathrm{SB}})$ is the likelihood of the particular ρ_{SB}, given that data D have been obtained for the system.

For the case where we have no data on the bath, $L(D|\rho_{\mathrm{SB}})$ is a constant for all the ρ_{SB} in this set, since all such ρ_{SB} satisfies Eq. (6.2). More generally, if some partial measurement could be done on the bath, then we can weigh each ρ_{SB} with its likelihood. For now, we restrict ourselves to the former case, in which $\overline{\rho_{\mathrm{SB}}} = \int (\mathrm{d}\rho_{\mathrm{B}}) |\psi\rangle_{\mathrm{SB}}\langle\psi| / \mathrm{tr}\{\cdots\}$, where $|\psi\rangle_{\mathrm{SB}}$ is some pure system-bath state (it suffices to consider pure states on the system and bath).

However, in order to do this integration, one needs to specify a prior in $(\mathrm{d}\rho_{\mathrm{B}})$, which is the difficulty in Bayesian mean estimation. We will not discuss this controversy here, but just refer the readers to a vast amount of literature on how well motivated and accurate Bayesian mean estimators can be if used with discretion, how the choice of prior should be made, and how insignificant the choice of prior turns out to be on the final results when the data dominate the inference[BK10b, KW96, SNS+13]. Now, since the bath is not amenable to microscopic control, states

of the bath related to each other by some unitary transformation would not be distinguishable to us. Hence, one natural way to do this averaging is to write

$$\overline{\rho_{\mathrm{SB}}} = \int (\mathrm{d}U_{\mathrm{B}})(I_{\mathrm{S}} \otimes U_{\mathrm{B}})|\psi\rangle_{\mathrm{SB}}\langle\psi|(I_{\mathrm{S}} \otimes U_{\mathrm{B}}^{\dagger}), \tag{6.67}$$

where $(\mathrm{d}U_{\mathrm{B}})$ is the averaging over unitary operators on the bath according to the *Haar measure* [BŻ06]. This immediately gives the state

$$\overline{\rho_{\mathrm{SB}}} = \rho_{\mathrm{S}} \otimes \frac{\mathbb{1}_{\mathrm{B}}}{d_{\mathrm{B}}}, \tag{6.68}$$

which happens to be the same as what we obtained using the maximum-entropy reasoning in the absence of any knowledge of the bath. If one now imposes a temperature constraint on the bath, and uses the Boltzmann factor as a weight factor on top of the Haar averaging, one obtains, for the case of $H_{\mathrm{SB}} = 0$,

$$\overline{\rho_{\mathrm{SB}}} = \rho_{\mathrm{S}} \otimes \rho_{\mathrm{B}}^{(\mathrm{th})}, \tag{6.69}$$

which is another consistency check. However, for an arbitrary H_{SB}, it is not easy to work out the explicit form of the Bayesian mean estimator due to the difficulty in the integral over the state space as mentioned in Chapter 2, unless the H_{SB} is simple enough and has some nice properties. Below, we discuss a simple illustrative example.

We consider the Ising model in which a system qubit is weakly coupled to N bath qubits. The Hamiltonian is given by

$$H_{\mathrm{B}} = J_{ij} \sum_{i<j} \sigma_z^{(i)} \otimes \sigma_z^{(j)}, \text{ and } H_{\mathrm{SB}} = J_i \sum_i \sigma_z^{\mathrm{S}} \otimes \sigma_z^{(i)}, \tag{6.70}$$

for the bath and the system-bath interaction respectively. If N is small, there is no reason to expect the Bayesian mean state to be almost a separable state. However, if $N \gg 1$, then one expects the Bayesian mean state to be very close in trace distance to a product state. Indeed, for this particular form of the Hamiltonian, one sees that the off-diagonal entries of $\overline{\rho_{\mathrm{SB}}}$ average to zero with the Haar measure, and we recover a product state due to the averaging.

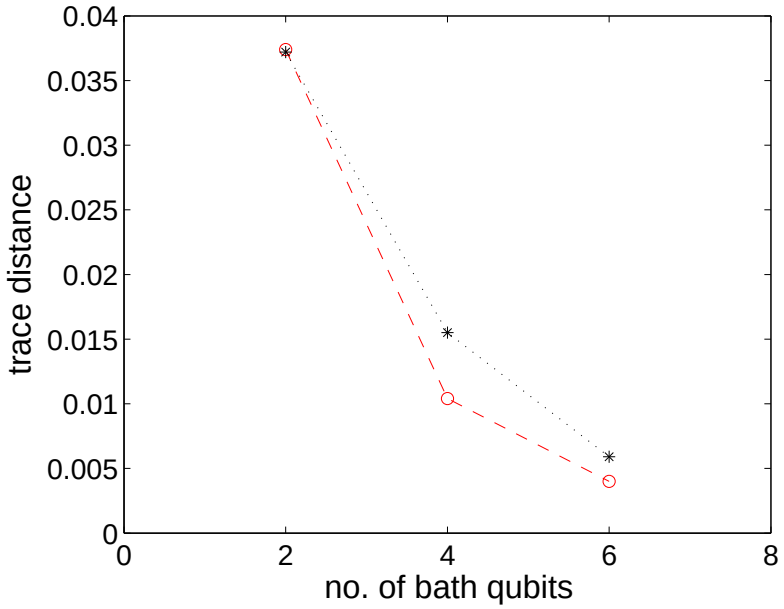

Figure 6.2: The trace distance between the Bayesian mean state and the separable state as a function of the number of bath qubits N in the linear Ising model. Case 1 with black dotted line is done with all J_{ij} and J_i terms being equal in Eq. (6.70), and Case 2 with red dashed line is a more realistic model in which the force is inversely proportional to the distance square. In both cases, one observes a decrease of trace distance as a function of N. In the more realistic model, the decrease is slightly faster.

To see how effective such an averaging is in removing any correlation that is there between the system and the bath, we carried out numerical simulations to model the qubits in the Ising model. We used a linear Ising model in which the qubits are put in a straight line (just like how ions are placed in some ion trap experiment using a linear Paul trap), for which the system qubit is in the center of the chain. We then put different numbers of bath qubits on both sides of it ($N = 2$, 4 and 6). We calculate the Bayesian mean estimator by averaging over 10^6 randomly generated unitary operators [PŻK98], and add these states with the Boltzmann weight. We then calculate the trace distance between the Bayesian mean estimator and the separable state that would be obtained if $H_{\rm SB} = 0$. Figure 6.2 shows the result of such simulation. It is observed that, as N increases, the trace distance between the Bayesian mean state and the product state decreases, justifying the intuitive claim we made above. In the simulation, we considered two cases: Case 1 (black dotted line) with all J_{ij} and J_i terms being equal in Eq. (6.70); and Case 2 (red dashed

line), which is a more realistic model, such that

$$J_{ij} \propto \frac{1}{r_{ij}}, \text{ and } J_i \propto \frac{1}{r_i}, \tag{6.71}$$

where r_{ij} is the distance between the i-th bath qubit and the j-th bath qubit, and r_i is the distance between the system qubit and the i-th bath qubit. In both cases, one observes a decrease of trace distance as a function of N. In Case 2, the decrease is slightly faster. It is also quite remarkable that even for $N = 2$, which is quite small, the trace distance between the two states is already rather small. The averaging in the Bayesian mean estimator process is quite efficient in removing the entanglement between the system and bath. As a result of the ignorance about the bath, the total system-bath state is almost not entangled.

More generally, one could allow for an arbitrary form of the Hamiltonian rather than the tensor product of Pauli operators above, and instead model the qubit interactions by a randomly generated Hermitian operator (possibly entangling) between each qubit pair. To see whether such general Hamiltonian still gives rise to a separable Bayesian mean estimator, we randomly generate a Hermitian matrix and use it to model the pairwise interactions between the qubits in the Ising model. We then do the same calculations as above for the trace distance between the Bayesian mean estimator and the separable state that would be obtained if $H_{\text{SB}} = 0$. The steps are then repeated for another randomly generated hermitian matrix. Preliminary simulation results show that, in this case, the trace distance between the Bayesian mean state and the separable state is greater than that using the tensor Pauli interaction above, likely due to entangling properties in the Hamiltonian. However, as N increases, the trace distance also decreases in a similar manner as in Fig. 6.2. In the case of the macroscopic environment that we had in mind, since $N \gg 1$, the resulting state is still expected to possess little entanglement. The computation becomes more numerically demanding when N is larger than 8. We are still in the process of working along this direction to obtain more detailed results. If on the other hand $N \approx O(1)$, then one does not expect to get a separable state between the system and the bath. However, in this case, the assumption that the bath is macroscopic, with many quantum degrees of freedom that one cannot control, is

perhaps difficult to justify.

6.5 Conclusion

The specification of an initial system-bath composite state is required for the prediction of future dynamics of the composite state, or even just the system-only state. That we usually do not have the ability to do complete tomography on the bath prevents us from determining the microscopic state of the bath. Traditionally, the system-bath is often taken to be in a product state with the bath in some reference state, or the composite system is in some unknown mysterious state. However, our state is our best estimate of the preparation given all the prior knowledge and measurement information we have at hand. In this chapter, we show how to take proper account of all the knowledge, together with the very definition of bath being microscopically uncontrollable, to arrive at a reasonable and consistent initial system-bath state.

Such a system-bath composite state turns out to be uncorrelated if there is no coupling between the system and bath, or if there is coupling, but the preparation of the system is done in the absence of the bath, as one expects. However, if *in-situ* preparation of the system is actually the case, as in some experiment, then one should be careful when using an initially uncorrelated state. It can at best be an approximation as the initial state is almost uncorrelated. The deviation from the uncorrelated state is however bounded by the norm of the interaction Hamiltonian. If the level of accuracy required in the prediction is comparable to the norm of the interaction Hamiltonian, then using an uncorrelated initial state becomes unjustified.

Having deduced a reasonable and consistent initial state, we must mention that deviations from predictions for evolutions of the system-bath state using this deduced initial state may arise in the actual experiment. When such deviations do arise, it is an indication that the initial state one wrote down is not consistent with the prior knowledge that one has about the system-bath composite. Such prior-data conflict should be carefully examined and is often due to a microscopically

controlled, repeatedly and identically prepared degree of freedom in the bath that one failed to identify and account for. Our description of the macroscopic bath should then be adjusted accordingly to properly account for this degree of freedom that one is actually controlling in the experiment.

Chapter 7

Conclusion and Outlook

As mentioned in the Introduction, quantum information and computation is a rich and rapidly progressing field that is currently under intense research and development. We can foresee that any future breakthrough in this field is most likely to come about due to a concerted effort in better understanding of the theory as well as advances in the experimental control and manipulations on the quantum systems. I must say that I am greatly thankful that due to fortunate circumstances, during my PhD studies, I was involved in both theoretical and experimental work. And it is the aim of this thesis to report those results obtained.

In terms of the experimental work, in Chapter 3, we introduced a novel, yet very practical method of generating mixed states by using the VPR. Such setups are not cumbersome to build up, and yet are versatile enough to realize a variety of two-photon states that one may need in a quantum information experiment. We have tested its robustness and accuracy, and also utilized it in an experiment on entanglement witness. Due to the limitations of our lab, we only have two pieces of VPR and a single controller that cannot independently control each VPR. As discussed in Chapter 2, if one has more VPRs, which can be independently controlled, the source is capable of producing more complex states, such as Collins-Gisin states, which have been used in tests of Bell's inequality. Furthermore, we are currently using this source to perform another experiment on quantitative aspects of wave-particle duality and will report the results in a short while.

In Chapter 4, we described the experiment performed, with our new source, on

entanglement detection using optimal-witness families. In this work, in contrast to conventional witness experiments, we are measuring a family of optimal witnesses in one go. The witness family is realized elegantly using a HOM interferometer. By projecting to the eigenbasis of the witness family, we realize the most direct measurement of these witnesses. We have further introduced adaptive measurement schemes that allow us to reduce the mean number of witness families that are needed to detect entanglement. Both simulation and experimental results give an affirmative answer that our scheme verifies the presence of entanglement with efficient use of the resources.

Further work in this area could involve an extension of such witness family measurement that enables IC tomography to higher dimensional quantum systems. For example, preliminary results have shown that in the case of a two qutrit system, 11 witness families are enough for tomographic completeness, rather than 81 witnesses. The minimal number of families of this kind that one would need is currently unknown for two qu-dit systems, with $d > 4$. Moreover, the adaptive methods introduced in Chapter 4 could also be applied to higher-dimensional quantum systems so that one is expecting a further reduction in the number of witness families that needs to be measured. As commented in Section 4.8, it is also interesting to apply such an idea to nonlinear witnesses where the more stringent bounds could be calculated after only partial tomography. This is currently unexplored.

As we were carrying out the experiments described in these chapters, we realized that in order to reconstruct the unknown quantum state, many additional parameters in the experimental setup need to be determined or pre-calibrated, for example, the quantum efficiency or the dark counts of the detectors, the splitting ratio of the beam splitters for various polarizations, etc. However, such pre-calibration may not always be feasible or easy to be done in the actual experimental setup. In Chapter 5, we looked into such cases of quantum state tomography with additional unknown parameters. We illustrated, with two typical examples, the construction of optimal error regions under such situations. The construction using bounded-likelihood regions are applied in this case, as the natural extension of the maximum-likelihood estimator. The optimal error regions constructed are the maximum likelihood re-

gions. That is, of all the regions of the same pre-chosen size, it is the one for which the data is the most likely. They also the smallest credible regions.

The optimal error regions constructed are a joint property of the quantum state parameters and the additional parameters. If the experimenter is interested in one or more of these parameters while uninterested in the nuisance parameters, he or she could marginalize over the nuisance parameters with a prior that incorporates all that is known for each of these nuisance parameters to obtain a marginal likelihood function for the parameters of interest. This marginalization encompasses all appropriate knowledge that one may have on these nuisance parameters and the resulting marginal likelihood function is the averaged likelihood function on the parameters on interest, representing one's best knowledge about the parameters of interest after taking the nuisance parameters into consideration. With this marginal likelihood function at hand, one can then proceed with the construction of bounded likelihood regions as before.

As there is a myriad of parameters that could be unknown in various experiments, it is not possible, nor the aim of this chapter, to study in details all such possible parameters. The two chosen examples are meant to illustrate the technique. However, as an outlook, we do mention two other examples that we plan to study in the future due to their unique features. One of them is quantum homodyne tomography [VR89]. It is a robust and versatile tool of quantum optics and has been applied in many different experimental settings for which the technique of maximum likelihood has been applied successfully. Another example is experiments involving dark count of the detectors. At the moment, such experiments are always done with pre-calibration of the detectors before the actual experiment. That is, one first switches on the detectors alone and records the average number of counts per unit time due to the dark counts, N_{dark}, without running the experiment. One then runs the experiments and records the number of counts, N_{observed}. The number of actual counts is then corrected for the dark count. However, due to the statistical fluctuations in the data, this could sometimes results in a negative value of N_{actual}. This situation is worse when the detector has a high dark count rate, or the state is nearly pure or rank-deficient, so that N_{actual} is very small for one of the POM ele-

ment measured, or both. A more sensible way to analyze such an experiment would be to estimate the dark counts simultaneously instead of doing a pre-calibration first and correction later. The techniques presented in this chapter then becomes relevant.

In Chapter 6, we studied the effect of the presence of coupling between the quantum system and the bath in a tomography experiment, so as to answer the question: What is the initial system-bath state? Such a question is very pertinent in studies of open quantum systems where the initial state of the system-bath composite is required for predictions of its future behavior. We disagree that such a state is somewhat arbitrary, as some authors assert, by introducing a symbolic system-bath state, as if no concrete statement could be made. We showed in this chapter that employing the standard tools of quantum state tomography, one can write down a reasonable and consistent initial system-bath state.

We take proper account of our prior knowledge of the system-bath composite, encompassing all measurements already taken to characterize the experimental circumstances. This knowledge, together with the very definition of the bath being microscopically uncontrollable, gives us a good handle on the initial system-bath state. Under typical conditions, the initial system-bath state is uncorrelated or almost uncorrelated. However, one should be very careful if *in-situ* preparation of the system state is the case, where the coupling between the system and the bath may result in correlations between them. The uncorrelated state can at best be an approximation, the deviation from which cannot be neglected, if the level of accuracy required in the prediction is comparable to the norm of the system-bath interaction Hamiltonian.

Having written down this initial system-bath state, the next step would be to study the effect of using this initial state in the various dynamical equations such as the master equation. Deviations from the predictions could be produced from the evolution of this initial system-bath state could be indicative of the presence of prior-data conflict. Such deviation could be due to a microscopically controlled, repeatedly identically prepared degree of freedom in the bath that one failed to identify and is an alert to call for a more thorough analysis of the bath.

Bibliography

[AGR81] A. Aspect, P. Grangier, and G. Roger. Experimental tests of realistic local theories via Bell's theorem. *Physical Review Letters*, 47:460–463, 1981.

[BD00] V. Bužek and G. Drobný. Quantum tomography via the maxent principle. *Journal of Modern Optics*, 47:2823, 2000.

[BDMDN$^+$03] M. Barbieri, F. De Martini, G. Di Nepi, P. Mataloni, G. M. D'Ariano, and C. Macchiavello. Detection of entanglement with polarized photons: Experimental realization of an entanglement witness. *Physical Review Letters*, 91:227901, 2003.

[BDPS00] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi. Maximum-likelihood estimation of the density matrix. *Physical Review A*, 61:010304(R), 2000.

[BK10a] R. Blume-Kohout. Hedged maximum likelihood quantum state estimation. *Physical Review Letters*, 105:200504, 2010.

[BK10b] R. Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12:043034, 2010.

[BLP09] H.-P. Breuer, E.-M. Laine, and J. Piilo. Measure for the degree of non-markovian behavior of quantum processes in open systems. *Physical Review Letters*, 103:210401, 2009.

[BP06] H.-P. Breuer and F. Petruccione. *The Theory of Open Quantum Systems*. Clarendon Press, Oxford, 2006.

[BPM+97] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger. Experimental quantum teleportation. *Nature*, 390(6660):575–579, 1997.

[BŻ06] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2006.

[CG04] D. Collins and N. Gisin. A relevant two qubit Bell inequality inequivalent to the CHSH inequality. *Journal of Physics A: Mathematical and General*, 37(5):1775, 2004.

[CGK98] I. L. Chuang, N. Gershenfeld, and M. Kubinec. Experimental implementation of fast quantum searching. *Physical Review Letters*, 80:3408, 1998.

[CKW00] V. Coffman, J. Kundu, and W. K. Wootters. Distributed entanglement. *Physical Review A*, 61:052306, 2000.

[CL08] B. P. Carlin and T. A. Louis. *Bayesian Methods for Data Analysis (Texts in Statistical Science)*. 3rd ed., Chapman & Hall/CRC, New York, 2008.

[CR12] M. Christandl and R Renner. Reliable quantum state tomography. *Physical Review Letters*, 109:120403, 2012.

[CS78] J. F. Clauser and A. Shimony. Bell's theorem. Experimental tests and implications. *Reports on Progress in Physics*, 41(12):1881, 1978.

[DLT+13] J. Dai, Y. L. Len, Y. S. Teo, L. A. Krivitsky, and B.-G. Englert. Controllable generation of mixed two-photon states. *New Journal of Physics*, 15(6):063011, 2013.

[DLT+14] J. Dai, Y. L. Len, Y. S. Teo, B.-G. Englert, and L. A. Krivitsky. Experimental detection of entanglement with optimal-witness families. *Physical Review Letters*, 113:170402, 2014.

[DRS12] A. R. U. Devi, A. K. Rajagopal, and Sudha. Open-system quantum dynamics with correlated initial states, not completely positive maps, and non-markovianity. *Physical Review A*, 83:022109, 2012.

[EGS06] M. K. Evans, I. Guttman, and T. Swartz. Optimality and computations for relative surprise inferences. *Canadian Journal of Statistics*, 34:113, 2006.

[Eke91] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67:661–663, 1991.

[EKW01] B.-G. Englert, C. Kurtsiefer, and H. Weinfurter. Universal unitary gate for single-photon two-qubit states. *Physical Review A*, 63(3):032303, 2001.

[Eng06] B.-G. Englert. *Lectures on Quantum Mechanics.* World Scientific, Singapore, 2006.

[Eng13] B.-G. Englert. On quantum theory. *The European Physical Journal D*, 67:238, 2013.

[Fan57] U. Fano. Description of states in quantum mechanics by density matrix and operator techniques. *Review of Modern Physics*, 29:74–93, 1957.

[Fis22] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368, 1922.

[GL06] O. Gühne and N. Lütkenhaus. Nonlinear entanglement witnesses. *Physical Review Letters*, 96:170502, 2006.

[GS22] W. Gerlach and O. Stern. Das magnetische Moment des Silberatoms. *Zeitschrift für Physik*, 9:353–355, 1922.

[GSB05] A. Gogo, W. D. Snyder, and M. Beck. Comparing quantum and

classical correlations in a quantum eraser. *Physical Review A*, 71(5):052103, 2005.

[GT09] O. Gühne and G. Tóth. Entanglement detection. *Physics Reports*, 474:1–75, 2009.

[Haa90] R. Haag. Fundamental irreversibility and the concept of events. *Communications in Mathematical Physics*, 132:245–251, 1990.

[Hec01] E. Hecht. *Optics, Fourth Edition.* Addison-Wesley, 2001.

[Hei27] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43:172–198, 1927.

[Hei30] W. Heisenberg. *The Physical Principles of Quantum Theory.* The University of Chicago Press, Chicago, IL, 1930.

[Hel76a] C. W. Helstrom. *Quantum Detection and Estimation Theory.* Academic Press, 1976.

[Hel76b] C. W. Helstrom. *Quantum Detection and Estimation Theory.* Academic Press, New York, 1976.

[HHH96] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996.

[HHHH09] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Review of Modern Physics*, 81:865, 2009.

[HMP+96] Z. Hradil, R. Myška, J. Peřina, M. Zawisky, Y. Hasegawa, and H. Rauch. Quantum phase in interferometry. *Physical Review Letters*, 76:4295, 1996.

[HOM87] C. K. Hong, Z. Y. Ou, and L. Mandel. Measurement of subpicosecond time intervals between two photons by interference. *Physical Review Letters*, 59:2044–2046, 1987.

[Jay57] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620, 1957.

[Jay82] E. T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70:939, 1982.

[JKMW01] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White. Measurement of qubits. *Physical Review A*, 64(5):052312, 2001.

[Joz94] R. Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41(12):2315–2323, 1994.

[KE04] P. G. Kwiat and B.-G. Englert. *Quantum erasing the nature of reality or, perhaps, the reality of nature? Science and Ultimate Reality: Quantum Theory, Cosmology, and Complexity.* Cambridge University Press, 2004.

[KKK10] M. Kotowski, M. Kotowski, and M. Kuś. Universal nonlinear entanglement witnesses. *Physical Review A*, 81:062318, 2010.

[KMRRS07] A. M. Kuah, K. Modi, C. A. Rodríguez-Rosario, and E. C. G. Sudarshan. How state preparation can affect a quantum experiment: Quantum process tomography for open systems. *Physical Review A*, 76:042113, 2007.

[Kra83] K. Kraus. *States, Effects, and Operations: Fundamental Notions of Quantum Theory.* Lecture Notes in Physics **190**, edited by A. Böhm, J. D. Dollard, W. K. Wootters, Springer, Heidelberg, 1983.

[KW96] R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91:1343–1370, 1996.

[KWW+99] P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P. H. Eberhard. Ultrabright source of polarization-entangled photons. *Physical Review A*, 60(2):773–776, 1999.

[Len14] Y. L. Len. *Entanglement Detection with Minimal Tomography Using Witness Bases Measurement*. Master Thesis, NUS, 2014.

[LHLLK06] A. Ling, P. Y. Han, A. Lamas-Linares, and C. Kurtsiefer. Preparation of Bell states with controlled white noise. *Laser Physics*, 16(7):1140–1144, 2006.

[LKCH00] M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki. Optimization of entanglement witnesses. *Physical Review A*, 62(5):052310, 2000.

[LLH+11] B.-H. Liu, L. Li, Y.-F. Huang, C.-F. Li, G.-C. Guo, E.-M. Laine, H.-P. Breuer, and J. Piilo. Experimental control of the transition from markovian to non-markovian dynamics of open quantum systems. *Nature Physics*, 7:931–934, 2011.

[Lou00] R. Loudon. *The Quantum Theory of Light, 3rd edition*. Oxford Science Publications, 2000.

[Max73] J. C. Maxwell. *A Treatise on Electricity and Magnetism*. Clarendon Press, Oxford, 1873.

[Mod12] K. Modi. Operational approach to open dynamics and quantifying initial correlations. *Scientific Reports*, 2:581, 2012.

[MŘH12] D. Mogilevtsev, J. Řeháček, and Z. Hradil. Self-calibration for self-consistent tomography. *New Journal of Physics*, 14:095001, 2012.

[MRRAG12] K. Modi, C. A. Rodríguez-Rosario, and A. Aspuru-Guzik. Positivity in the presence of initial system-environment correlation. *Physical Review A*, 86:064102, 2012.

[MS10] K. Modi and E. C. G. Sudarshan. Role of preparation in quantum process tomography. *Physical Review A*, 81:052119, 2010.

[NC10] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.

[NE12] H. K. Ng and B.-G. Englert. A simple minimax estimator for quantum states. *International Journal of Quantum Information*, 10:1250038, 2012.

[New87] I. Newton. *Philosophiae Naturalis Principia Mathematica*. Streator, London, 1687.

[NFM91] J. W. Noh, A. Fougères, and L. Mandel. Measurement of the quantum phase by photon counting. *Physical Review Letters*, 67:1426, 1991.

[Per96] A. Peres. Separability criterion for density matrices. *Physical Review Letters*, 77:1413, 1996.

[Pv04] M. G. A. Paris and J. Řeháček. *Quantum State Estimation*. Lecture Notes in Physics **649**, edited by Paris, M. G. A. and Řeháček, J. Springer, Heidelberg, 2004.

[PŻK98] M. Pozniak, K. Życzkowski, and M. Kuś. Composed ensembles of random unitary matrices. *Journal of Physics A: Mathematical and General*, 31:1059, 1998.

[QBJ13] N. Quesada, A. M. Brańczyk, and D. F. V. James. Self-calibrating tomography for multidimensional systems. *Physical Review A*, 87:062118, 2013.

[ŘH03] J. Řeháček and Z. Hradil. Quantification of entanglement by means of convergent iterations. *Physical Review Letters*, 90:127904, 2003.

[ŘHKL07] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky. Diluted maximum-likelihood algorithm for quantum tomography. *Physical Review A*, 75(4):042108, 2007.

[RHP10] A. Rivas, S. F. Huelga, and M. B. Plenio. Entanglement and non-markovianity of quantum evolutions. *Physical Review Letters*, 105:050403, 2010.

[ŘHZ+99] J. Řeháček, Z. Hradil, M. Zawisky, S. Pascazio, H. Rauch, and J. Peřina. Testing of quantum phase in matter-wave optics. *Physical Review A*, 60:473, 1999.

[RWC+11] D. Z. Rossatto, T. Werlang, L. K. Castelano, C. J. Villas-Boas, and F. F. Fanchini. Purity as a witness for initial system-environment correlations in open-system dynamics. *Physical Review A*, 84:042113, 2011.

[RWM+15] M. Ringbauer, C. J. Wood, K. Modi, A. Gilchrist, A. G. White, and A. Fedrizzi. Characterizing quantum dynamics with initial system-environment correlations. *Physical Review Letters*, 114:090402, 2015.

[SBC01] R. Schack, T. A. Brun, and C. M. Caves. Quantum Bayes rule. *Physcial Review A*, 64:014305, 2001.

[Sch01] J. Schwinger. *Quantum Mechanics. Symbolism of Atomic Measurements*. Berlin: Springer, 2001.

[SEW91] M. O. Scully, B.-G. Englert, and H. Walther. Quantum optical tests of complementarity. *Nature*, 351:111, 1991.

[SIK+13] S. S. Straupe, D. P. Ivanov, A. A. Kalinkin, I. B. Bobrov, S. P. Kulik, and D. Mogilevtsev. Self-calibrating tomography for angular schmidt modes in spontaneous parametric down-conversion. *Physical Review A*, 87:042109, 2013.

[SKR+15] C. Schwemmer, L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühne. Systematic errors in current quantum state tomography tools. *Physical Review Letters*, 114:080403, 2015.

[SNE14] J. Shang, H. K. Ng, and B.-G. Englert. Quantum state tomography: Mean squared error matters, bias does not. *Preprint at arXiv:1405.5350*, 2014.

[SNS+13] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert. Optimal error regions for quantum state estimation. *New Journal of Physics*, 15:123026, 2013.

[SSN+15a] Y.-L. Seah, J. Shang, H. K. Ng, D. J. Nott, and B.-G. Englert. Monte Carlo sampling from the quantum state space. II. *Preprint at arXiv:1407.7806*, 2015.

[SSN+15b] J. Shang, Y.-L. Seah, H. K. Ng, D. J. Nott, and B.-G. Englert. Monte Carlo sampling from the quantum state space. I. *Preprint at arXiv:1407.7805*, 2015.

[Ter00] B. M. Terhal. Bell inequalities and the separability criterion. *Physics Letters A*, 271(5):319–326, 2000.

[TLL+12] J.-S. Tang, C.-F. Li, Y.-L. Li, X-.B. Zou, G.-C. Guo, H.-P. Breuer, E.-M. Laine, and J. Piilo. Measuring non-markovianity of processes with controllable system-environment interaction. *Europhysics Letters*, 97:10002, 2012.

[TŘH14] Y. S. Teo, J. Řeháček, and Z. Hradil. Coarse-grained quantum state estimation for noisy measurements. *Physical Review A*, 88:022111, 2014.

[TSE+12] Y. S. Teo, B. Stoklasa, B.-G. Englert, J. Řeháček, and Z. Hradil. Incomplete quantum state estimation: A comprehensive study. *Physical Review A*, 85(4):42317, 2012.

[TZE+11] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil. Quantum-state reconstruction by maximizing likelihood and entropy. *Physical Review Letters*, 107(2):020404, 2011.

[VR89] K. Vogel and H. Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Physical Review A*, 40:2847, 1989.

[WAB+05] T.-C. Wei, J. B. Altepeter, D. Branning, P. M. Goldbart, D. F. V. James, E. Jeffrey, P. G. Kwiat, S. Mukhopadhyay, and N. A. Peters. Synthesizing arbitrary two-photon polarization mixed states. *Physical Review A*, 71:032329, 2005.

[WECI08] M. M. Wolf, J. Eisert, T. S. Cubitt, and Cirac J. I. Assessing non-markovian quantum dynamics. *Physical Review Letters*, 101:150402, 2008.

[Wer89] R. F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40:4277–4281, 1989.

[Woo98] W. K. Wootters. Entanglement of formation of an arbitrary state of two qubits. *Physical Review Letters*, 80:2245–2248, 1998.

[ZB05] M. Ziman and V. Bužek. Concurrence versus purity: Influence of local channels on Bell states of two qubits. *Physical Review A*, 72:052325, 2005.

[ZS01] K. Życzkowski and H.-J. Sommers. Induced measures in the space of mixed quantum states. *Journal of Physics A: Mathematical and General*, 34:7111, 2001.

[ZTE10] H. Zhu, Y. S. Teo, and B.-G. Englert. Minimal tomography with entanglement witnesses. *Physical Review A*, 81(5):052339, 2010.