

## Preface

This book is a comprehensive collection representing most of the theoretical and experimental developments of the last decade in the field of quantum estimation of states and operations. Though the field is fairly new, it has already been recognized as a necessary tool for researchers in quantum optics and quantum information. The subject has a fundamental interest of its own, since it concerns the experimental characterization of the quantum state, the basic object of the quantum description of physical systems. Moreover, quantum estimation techniques have been receiving attention for their crucial role in the characterization of registers at the quantum level, which itself is a basic tool in the development of quantum information technology.

The field is now mature and a stable part of many graduate curricula, but only a few review papers have been published in recent years, and no comprehensive volume with theoretical and experimental contributions has ever appeared. We anticipate readers in the areas of fundamental quantum mechanics, quantum and nonlinear optics, quantum information theory, communication engineering, imaging and pattern recognition.

As editors, we wish to thank Berge Englert for encouragement and support, and all the authors for their contributions, which will advance both the specific field and the general appreciation of it. Their efforts and the significant time they spent preparing the chapters are much appreciated. We are also grateful to Janine O'Guinn of the University of Oregon for her excellent work in copy-editing the volume. Finally, let us acknowledge support from EC project IST-2000-29681, and Czech Ministry of Education project LN00A015.

Milano and Olomouc,  
April 2004

*Matteo G A Paris  
Jaroslav Řeháček*



# Contents

## 1 Introduction

*Matteo G. A. Paris, Jaroslav Řeháček* ..... 1

---

## Part I Quantum Estimation

---

### 2 Quantum Tomographic Methods

*G. Mauro D'Ariano, Matteo G. A. Paris, Massimiliano F. Sacchi* ..... 9

### 3 Maximum-Likelihood Methods in Quantum Mechanics

*Zdeněk Hradil, Jaroslav Řeháček, Jaromír Fiurášek, Miroslav Ježek* .... 63

### 4 Qubit Quantum State Tomography

*Joseph. B. Altepeter, Daniel F. V. James, Paul G. Kwiat* ..... 117

### 5 Unknown Quantum States and Operations, a Bayesian View

*Christopher A. Fuchs and Rüdiger Schack* ..... 151

### 6 Quantum Tomography from Incomplete Data via *MaxEnt* Principle

*Vladimír Bužek* ..... 191

### 7 Experimental Quantum State Measurement of Optical Fields and Ultrafast Statistical Sampling

*Michael G. Raymer, Mark Beck* ..... 239

### 8 Characterization of Quantum Devices

*Giacomo Mauro D'Ariano, Paoloplacido Lo Presti* ..... 299

### 9 Quantum Operations on Qubits and their Characterization

*Francesco De Martini, Marco Ricci and Fabio Sciarrino* ..... 335

### 10 Maximum-Likelihood Estimation in Experimental Quantum Physics

VIII    Contents

<i>Gerald Badurek, Zdeněk Hradil, Alexander Lvovsky, Gabriel Molina-Teriza, Helmut Rauch, Jaroslav Řeháček, Alipasha Vaziri, Michael Zawisky</i> .....	375
--	-----

---

**Part II Quantum Decision**

---

<b>11 Discrimination of Quantum States</b>	
<i>János A. Bergou, Ulrike Herzog, Mark Hillery</i> .....	419
<b>12 Quantum States: Discrimination and Classical Information Transmission. A Review of Experimental Progress</b>	
<i>Anthony Chefles</i> .....	469
<b>Index</b> .....	515

# 1 Introduction

Matteo G. A. Paris<sup>1</sup> and Jaroslav Řeháček<sup>2</sup>

<sup>1</sup> Dipartimento di Fisica dell'Università di Milano, Italy

<sup>2</sup> Department of Optics, Palacky University, Olomouc, Czech Republic

The state of a physical system is the mathematical description of our knowledge of it, and provides information on its future and past. A state estimation technique is a method that provides the complete description of a system, *i.e.* achieves the maximum possible knowledge of the state, thus allowing one to make the best, at least the best probabilistic, predictions on the results of any measurement that may be performed on the system.

In classical physics the state of a system is a set of numbers, and it is always possible, at least in principle, to devise a procedure consisting of multiple measurements that fully recovers the state of the system. In Quantum Mechanics this is no longer possible, and this impossibility is inherently related to fundamental features of the theory, namely its linearity and the Heisenberg uncertainty principle. On one hand linearity implies the no-cloning theorem [1], which forbids us to create perfect copies of an arbitrary system in order to make multiple measurements on the same state. On the other hand, the uncertainty principle [2] says that one cannot perform an arbitrary sequence of measurements on a single system without disturbing it in some way, *i.e.* inducing a back-action which modifies the state itself. Therefore, it is not possible, even in principle, to determine the quantum state of a single system without having some prior knowledge on it [3]. This is consistent with the very definition of a quantum mechanical state, which in turn prescribes how to gain information about the state: many identical preparations taken from the same statistical ensemble are needed and different measurements should be performed on each of the copies.

Despite its fundamental interest the problem of inferring the state of a quantum system from measurements is *not* as old as quantum mechanics, and the first systematic approach was the work of U. Fano in the late fifties [4]. In the last decade a constantly increasing interest has been devoted to the subject. On one side, new developments in experimental techniques, especially in the fields of photodetection and nonlinear optical technology, resulted in a set of novel and beautiful experiments about quantum mechanics. On the other, increasing attention has been directed to quantum information technology, which is mostly motivated by the promising techniques of error correction and purification, which make possible fault tolerant quantum computing and long distance teleportation and cryptography. In particular, the development

of suitable purification protocols, and the possibility of a quantum characterization of communication channels, rely heavily on quantum estimation techniques.

This book aims to review all of the relevant quantum estimation techniques, and to assess the state of art in this novel field which has provoked renewed interest in fundamentals quantum mechanics. A number of leading experts have cooperated to describe the main features of the field. The rest of this introduction gives a brief description of their chapters. The volume is divided in two parts. The first is devoted to quantum estimation in the strict sense, both of quantum states and quantum operations, whereas the second (much shorter) part addresses the problem of state discrimination.

Part I of the book starts with Chapter 2 by G.M. D’Ariano *et al.*, which reviews quantum tomography, *i.e.* the determination of the expectation value of any operator (including nondiagonal projectors needed to construct a matrix representation of the density operator) for a generic quantum system from the measurement of a suitable set of observables (a quorum) on repeated preparations of the system. Topics include characterization of quora, determination of *pattern functions*, effect of instrumental noise, and examples of tomographic procedures for harmonic and spin systems.

Quantum estimation is in principle a deterministic problem, given that a quorum of observables is measured on the system of interest. However, often only partial information of the system can be achieved. Therefore, a question arises about what one can say about a quantum system given an arbitrary set of observations on repeated preparations of the system. In Chapter 3, Z. Hradil *et al.* give a statistical answer to this question using the maximum-likelihood principle. The formalism is applied to quantum-state estimation and discrimination as well as the estimation of quantum measurements and processes.

The polarization state of a photon is a natural experimental realization of a two-level quantum system – a qubit. For many experiments in quantum theory and quantum information it is very important to develop reliable sources of arbitrary polarization-entangled quantum states. Quantum estimation is important for the development of new quantum sources, since the quantum reconstruction techniques are natural means of calibration and tuning of experimental apparatuses. A detailed account of the production, characterization, and utilization of entangled states of light qubits is given by J.B. Altepeter *et al.* in Chapter 4.

Even in the realistic case of small ensembles, when the expectation values are not accessible, one can still infer the quantum state by means of the Bayesian principle of inference that provides a unique rule for updating the prior information about the quantum system after a measurement has been made. Although the principle itself is well justified, the notion of prior information is a highly subjective element of the theory. Therefore, in the Bayesian approach, the subjective interpretation of quantum states and operations is

stressed. The formulation of the quantum Bayesian inference is by Ch. Fuchs and R. Schack in Chapter 5, and they will then apply it to the reconstruction of quantum states and quantum operations.

Yet another principle of inference based on partial knowledge – Jaynes' principle of maximum entropy – comes from the information theory. Unlike the maximum likelihood estimation of Chapter 3 that always selects the most likely configuration, the principle of maximum entropy leads to the least biased estimate consistent with the given information. Its typical applications are momentum problems: the determination of the quantum state from the expectation values of a few, tomographically incomplete observations. An overview of the applications of Jaynes' principle to quantum reconstruction is reported by V. Bužek in Chapter 6.

The development of quantum estimation techniques started with the proposal by Vogel and Risken [5] and with the first experiments (which already showed reconstructions of coherent and squeezed states of a radiation field mode) performed in Mike Raymer's group at the University of Oregon. [6]. Chapter 7, by M. Raymer and M. Beck is a detailed review of the theoretical and experimental work on quantum state measurement based on homodyne detection, and discuss the determination of the quantum state of one or more modes of the radiation field.

Tomographic methods were initially employed only for measuring radiation states. However, they can profitably be used also to characterize devices through imprinting of quantum operations on quantum states. In Chapter 8 G.M. D'Ariano and P.L. Presti give a self-contained presentation of the theoretical bases of the method, together with examples of experimental setups based on homodyne tomography. As a contrast, Chapter 9 by F. De Martini *et al.* is devoted to reviewing the experimental realization of many unitary and non unitary operations on light qubit and their effective characterization by Pauli tomography of the polarization state.

The utility of the maximum-likelihood principle in experimental quantum estimation is demonstrated by Badurek *et al.* in Chapter 10, which closes the first part of the book. The ideas presented in Chapter 3 are systematically applied to experiments with quantum systems of increasing complexity starting with the quantum phase or simple two-dimensional systems and eventually coming to an infinite-dimensional mode of light.

The second part of the book consists of two chapters devoted to decisions among quantum hypotheses. Here we have a quantum system prepared in a state chosen from a discrete set, rather than from the whole set of possible states, and we want to discriminate among the set starting from the results of certain measurements performed on the system. To the extent that the quantum states to be discriminated are nonorthogonal, the problem is highly non-trivial, and of practical importance. Indeed, the increasing need for faster communication implies the steady decrease of the energy used for the transmission of a bit of information through the communication chan-

nel. When the carriers of information became truly microscopic systems the classical information they carry is encoded into their quantum state. A fundamental theorem of quantum theory tells us that it is not possible perfectly to distinguish between two non-orthogonal quantum states. This places a fundamental limit on the error rate of the communication because the orthogonality of the original alphabet is always degraded by the presence of the unavoidable noise during the transmission. Therefore, it is important to develop optimal discrimination techniques that keep the error rate as low as possible. Since orthogonality of quantum states cannot be restored by means of deterministic procedures, a novel discrimination technique, the so-called unambiguous quantum discrimination, has been suggested. In this approach inconclusive answers are accepted, and the compensation is an unambiguous answer when the operation succeeds. Since both ambiguous and unambiguous discriminations can be used for eavesdropping on quantum communication channels they are also crucial for the analysis of the security of quantum cryptography. In Chapter 11, J. Bergou *et al.* review various theoretical schemes that have been developed for discriminating among nonorthogonal quantum states, whereas a detailed account of experimental realizations is given by T. Chefles in Chapter 12.

The book contains several fairly self-contained groups of chapters, that could be employed for short courses. A course on theoretical estimation and detection in quantum theory might be based on Chapters 2, 3, 4, 5, 6, 8, and 11. Similarly, Chapters 4, 7, 9, 10, and 12 make up a course on experimental quantum estimation and detection. Standard deterministic methods of quantum estimation are covered by Chapters 2, 4, 7, 8, and 9, whereas methods of inference motivated by the statistical considerations or those coming from the information theory are treated in Chapters 3, 5, 6, and 10. In addition, second part of this volume, Chapters 11 and 12, contains a self-consistent exposition of quantum discrimination problems.

This book presents a young discipline that has grown vigorously in the last decade. We can expect further advances, most likely from applications to quantum information technology and implementations of (cryptographic or non-cryptographic) quantum communication schemes. The contents of the book suggests that progress, such as, for example, the use of entangled states and measures, or the extension to other physical systems, will come from quantum information, and will greatly benefit from an even closer collaboration among experimental and theoretical groups.

## References

1. W. K. Wootters and W. H. Zurek, *Nature* **299**, 802 (1982).
2. W. Heisenberg, *Zeit. für Physik*, **43**, 172 (1927); H. P. Robertson *Phys. Rev.* **34**, 163164 (1929)
3. G. M. D'Ariano and H. P. Yuen, *Phys. Rev. Lett.* **76**, 2832 (1996).

4. U. Fano, Rev. Mod. Phys. **29**, 74 (1957), Sec. 6.
5. K. Vogel and H. Risken, Phys. Rev. A, **40**, 2847 (1989).
6. D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993); M. G. Raymer, M. Beck, and D. F. McAlister, Phys. Rev. Lett. **72**, 1137 (1994).



## Part I

### Quantum Estimation



## 2 Quantum Tomographic Methods

G. Mauro D'Ariano<sup>1</sup>, Matteo G. A. Paris<sup>2</sup>, and Massimiliano F. Sacchi<sup>1</sup>

<sup>1</sup> Unità INFM and Dipartimento di Fisica “A. Volta”, Università di Pavia, Italia

<sup>2</sup> Unità INFM and Dipartimento di Fisica, Università di Milano, Italia.

### 2.1 Introduction

The state of a physical system is any mathematical object that provides a complete information on the system. The knowledge of the state is equivalent to know the result of any possible measurement on the system. In Classical Mechanics the state of a system, say a particle, is specified by the set of canonical coordinates and, at least in principle, it is always possible to devise a procedure made of multiple measurements which fully recovers the state. In Quantum Mechanics, on the contrary, this is not possible, due to the fundamental limitations related to the Heisenberg uncertainty principle [1, 2] and the no-cloning theorem [3]. In fact, on one hand one cannot perform an arbitrary sequence of measurements on a single system without inducing on it a back-action of some sort. On the other hand, it is not possible to create a perfect copy of the system without already knowing its state in advance. Thus, there is no way out, not even in principle, to infer the quantum state of a single system without having some prior knowledge on it [4]. For a quantum mechanical system it is possible to estimate the unknown state of a system when many identical copies are available in the same state, so that a different measurement can be performed on each copy. A procedure of this kind is called *quantum tomography*. The problem of finding a procedure to determine the state of a system from multiple copies was first addressed in 1957 by Fano [5], who called *quorum* a set of observables sufficient for a complete determination of the density matrix. However, since for a particle it is difficult to devise concretely measurable observables other than position, momentum and energy, the fundamental problem of measuring the quantum state has remained at the level of mere speculation up to almost ten years ago, when the issue finally entered the realm of experiments with the pioneering experiments by Raymer's group [6] in the domain of quantum optics. In quantum optics, in fact, using a balanced homodyne detector one has the unique opportunity of measuring all possible linear combinations of position and momentum of a harmonic oscillator representing a single mode of the electromagnetic field.

The first technique to estimate the elements of the density operator from homodyne measurements — so called *homodyne tomography* — originated from the observation by Vogel and Risken [7] that the collection of probabil-

ity distributions achieved by homodyne detection is just the Radon transform of the Wigner function  $W$ . Therefore, as in classical imaging, by Radon transform inversion one can obtain  $W$ , and then from  $W$  the matrix elements of the density operator. This first method, however, was affected by uncontrollable approximations, since arbitrary smoothing parameters are needed for the inverse Radon transform. In [8] the first exact technique was given for measuring experimentally the matrix elements of the density operator in the photon-number representation, by simply averaging functions of homodyne data. After that, the method was further simplified [9], and the feasibility for non-unit quantum efficiency of detectors—above some bounds—was established.

The exact homodyne method has been implemented experimentally to measure the photon statistics of a semiconductor laser [10], and the density matrix of a squeezed vacuum [11]. The success of optical homodyne tomography has then stimulated the development of state-reconstruction procedures for atomic beams [12], the experimental determination of the vibrational state of a molecule [13], of an ensemble of helium atoms [14], and of a single ion in a Paul trap [15].

Using quantum tomography the state is perfectly recovered in the limit of infinite number of measurements, while in the practical finite-measurements case, one can always estimate the statistical error that affects the reconstruction. For infinite dimensions the propagation of statistical errors of the density matrix elements make them useless for estimating the ensemble average of unbounded operators, and a method for estimating the ensemble average of arbitrary observable of the field without using the density matrix elements has been derived [16]. Further insights on the general method of state reconstruction has led to generalize homodyne tomography to any number of modes [17], and then to extend the tomographic method from the harmonic oscillator to an arbitrary quantum system using group theory [18–21]. A general data analysis method has been designed in order to unbias the estimation procedure from any known instrumental noise [20]. Moreover, algorithms have been engineered to improve the statistical errors on a given sample of experimental data—the so-called adaptive tomography [22]—and then max-likelihood strategies [23] have been used that improved dramatically statistical errors, however, at the expense of some bias in the infinite dimensional case, and of exponential complexity versus  $N$  for the joint tomography of  $N$  quantum systems. Quantum tomographic methods to perform fundamental tests of quantum mechanics have been proposed, as the measure of the non-classicality of radiation field of [24], and the test of the state reduction rule using light from parametric downconversion of [25].

The latest technical developments [26] derive the general tomographic method from spanning sets of operators, the previous group theoretical approaches [18–21] being just a particular case of this general method, where the group representation is just a device to find suitable operator “orthogonal-

ity” and “completeness” relations in the linear algebra of operators. Finally, very recently, a method for tomographic estimation of the unknown quantum operation of a quantum device has been derived [27], which uses a single fixed input entangled state, which plays the role of all possible input states in quantum parallel on the tested device, making finally the method a true “quantum radiography” of the functioning of a device.

This Chapter is structured to give a self-contained and unified derivation of the methods of quantum tomography. In Sect. 2 we introduce the generalized Wigner functions [28, 29] while in Sect. 3 we provide the basic elements of detection theory in quantum optics: photodetection, homodyne detection, and heterodyne detection. As we will see, heterodyne detection also provides a method for estimating the ensemble average of polynomials in the field operators, however, it is unsuitable for the density matrix elements in the photon-number representation. The effect of non unit quantum efficiency is taken into account for all such detection schemes. In Sect. 4 we give a brief history of quantum tomography, starting with the first proposal of Vogel and Risken [7] as the extension to the domain of quantum optics of the conventional tomographic imaging. As already mentioned, this method indirectly recovers the state of the system through the reconstruction of the Wigner function, and is affected by uncontrollable bias. The exact homodyne tomography method of [8] (successively simplified in [9]) is here presented on the basis of the general tomographic method of spanning sets of operators of [26]. As another application of the general method, the tomography of spin systems [30] is provided from the group theoretical method of [18–20]. In this section we also include further developments to improve the method, such as the deconvolution techniques of [20] to correct the effects of experimental noise by data processing, and the adaptive tomography [22] to reduce the statistical fluctuations of tomographic estimators. The generalization of [17] of homodyne tomography to many modes of radiation is reviewed in Sect. 5, where it is shown how tomography of a multimode field can be performed by using only a single local oscillator with a tunable field mode. Some results of Monte Carlo simulations from [17] are also shown for the state that describes light from parametric downconversion. Section 6 is devoted to reconstruction techniques [23] based on the maximum likelihood principle, which are suited to the estimation of a finite number of parameters, as proposed in [31], or to the state determination in the presence of very low number of experimental data [23]. Unfortunately, the algorithm of this method has exponential complexity versus the number of quantum systems for a joint tomography of many systems.

## 2.2 Wigner functions

Since Wigner’s pioneering work [28], generalized phase-space techniques have proved very useful in various branches of physics [33]. As a method to express

the density operator in terms of c-number functions, the Wigner functions often lead to considerable simplification of the quantum equations of motion, as for example, for transforming master equations in operator form into more amenable Fokker-Planck differential equations (see, for example, [34]). Using the Wigner function one can express quantum-mechanical expectation values in form of averages over the complex plane (the classical phase-space), the Wigner function playing the role of a c-number quasi-probability distribution, which generally can also have negative values. More precisely, the original Wigner function allows to easily evaluate expectations of symmetrically ordered products of the field operators, corresponding to the Weyl's quantization procedure [35]. However, with a slight change of the original definition, one defines generalized  $s$ -ordered Wigner function  $W_s(\alpha, \alpha^*)$ , as follows [29]

$$W_s(\alpha, \alpha^*) = \int_{\mathbb{C}} \frac{d^2\lambda}{\pi^2} e^{\alpha\lambda^* - \alpha^*\lambda + \frac{s}{2}|\lambda|^2} \text{Tr}[D(\lambda)\rho], \quad (2.1)$$

where  $\alpha^*$  denotes the complex conjugate of  $\alpha$ , the integral is performed on the complex plane with measure  $d^2\lambda = d\text{Re}\lambda d\text{Im}\lambda$ ,  $\rho$  represents the density operator, and

$$D(\alpha) \equiv \exp(\alpha a^\dagger - \alpha^* a) \quad (2.2)$$

denotes the displacement operator, where  $a$  and  $a^\dagger$  ( $[a, a^\dagger] = 1$ ) are the annihilation and creation operators of the field mode of interest. The Wigner function in (2.1) allows one to evaluate  $s$ -ordered expectation values of the field operators through the following relation

$$\text{Tr}[(a^\dagger)^n a^m :_s \rho] = \int_{\mathbb{C}} d^2\alpha W_s(\alpha, \alpha^*) \alpha^{*n} \alpha^m. \quad (2.3)$$

The particular cases  $s = -1, 0, 1$  correspond to *anti-normal*, *symmetrical*, and *normal* ordering, respectively. In these cases the generalized Wigner function  $W_s(\alpha, \alpha^*)$  are usually denoted by the following symbols and names

$\frac{1}{\pi}Q(\alpha, \alpha^*)$	for $s = -1$ “ $Q$ function”
$W(\alpha, \alpha^*)$	for $s = 0$ (usual Wigner function)
$P(\alpha, \alpha^*)$	for $s = 1$ “ $P$ function”

(2.4)

For the normal ( $s = 1$ ) and anti-normal ( $s = -1$ ) orderings, the following simple relations with the density matrix are well known

$$Q(\alpha, \alpha^*) \equiv \langle \alpha | \rho | \alpha \rangle, \quad (2.5)$$

$$\rho = \int_{\mathbb{C}} d^2\alpha P(\alpha, \alpha^*) |\alpha\rangle\langle\alpha|, \quad (2.6)$$

where  $|\alpha\rangle$  denotes the customary coherent state  $|\alpha\rangle = D(\alpha)|0\rangle$ ,  $|0\rangle$  being the vacuum state of the field. Among the three particular representations

(2.4), the  $Q$  function is positively definite and infinitely differentiable (it actually represents the probability distribution for ideal joint measurements of position and momentum of the harmonic oscillator: see Sect. 2.3.3). On the other hand, the  $P$  function is known to be possibly highly singular, and the only pure states for which it is positive are the coherent states [36]. Finally, the usual Wigner function has the remarkable property of providing the probability distribution of the quadratures of the field in the form of a marginal distribution, namely

$$\int_{-\infty}^{\infty} d\text{Im}\alpha W(\alpha e^{i\varphi}, \alpha^* e^{-i\varphi}) = {}_{\varphi} \langle \text{Re}\alpha | \rho | \text{Re}\alpha \rangle_{\varphi}, \quad (2.7)$$

where  $|x\rangle_{\varphi}$  denotes the (unnormalizable) eigenstate of the field quadrature

$$X_{\varphi} = \frac{a^{\dagger} e^{i\varphi} + a e^{-i\varphi}}{2} \quad (2.8)$$

with real eigenvalue  $x$ . Notice that any couple of quadratures  $X_{\varphi}$ ,  $X_{\varphi+\pi/2}$  is canonically conjugate, namely  $[X_{\varphi}, X_{\varphi+\pi/2}] = i/2$ , and it is equivalent to position and momentum of a harmonic oscillator. Usually, negative values of the Wigner function are viewed as signature of a non-classical state, the most eloquent example being the Schrödinger-cat state [37], whose Wigner function is characterized by rapid oscillations around the origin of the complex plane. From (2.1) one can notice that all  $s$ -ordered Wigner functions are related to each other through Gaussian convolution

$$W_s(\alpha, \alpha^*) = \int_{\mathbb{C}} d^2\beta W_{s'}(\beta, \beta^*) \frac{2}{\pi(s' - s)} \exp\left(-\frac{2}{s' - s} |\alpha - \beta|^2\right) \quad (2.9)$$

$$= \exp\left(\frac{s' - s}{2} \frac{\partial^2}{\partial \alpha \partial \alpha^*}\right) W_{s'}(\alpha, \alpha^*), \quad (s' > s). \quad (2.10)$$

Equation (2.9) shows the positivity of the generalized Wigner function for  $s < -1$ , as a consequence of the positivity of the  $Q$  function. The maximum value of  $s$  keeping the generalized Wigner functions as positive can be considered as an indication of the classical nature of the physical state [38].

An equivalent expression for  $W_s(\alpha, \alpha^*)$  can be derived as follows [32]. Equation (2.1) can be rewritten as

$$W_s(\alpha, \alpha^*) = \text{Tr}[\rho D(\alpha) \hat{W}_s D^{\dagger}(\alpha)], \quad (2.11)$$

where

$$\hat{W}_s = \int_{\mathbb{C}} \frac{d^2\lambda}{\pi^2} e^{\frac{s}{2}|\lambda|^2} D(\lambda). \quad (2.12)$$

Through the customary Baker-Campbell-Hausdorff (BCH) formula

$$\exp A \exp B = \exp\left(A + B + \frac{1}{2}[A, B]\right), \quad (2.13)$$

which holds when  $[A, [A, B]] = [B, [A, B]] = 0$ , one writes the displacement in normal order, and integrating on  $\arg(\lambda)$  and  $|\lambda|$  one obtains

$$\hat{W}_s = \frac{2}{\pi(1-s)} \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{2}{s-1} \right)^n a^{\dagger n} a^n = \frac{2}{\pi(1-s)} \left( \frac{s+1}{s-1} \right)^{a^{\dagger} a}, \quad (2.14)$$

where we used the normal-ordered forms

$$:(a^{\dagger} a)^n: = (a^{\dagger})^n a^n = a^{\dagger} a (a^{\dagger} a - 1) \dots (a^{\dagger} a - n + 1), \quad (2.15)$$

and the identity

$$:e^{-xa^{\dagger} a}: = \sum_{l=0}^{\infty} \frac{(-x)^l}{l!} (a^{\dagger})^l a^l = (1-x)^{a^{\dagger} a}. \quad (2.16)$$

The density matrix can be recovered from the generalized Wigner functions and, in particular, for  $s = 0$  one has the inverse of the Glauber formula

$$\rho = 2 \int_{\mathbb{C}} d^2 \alpha W(\alpha, \alpha^*) D(2\alpha) (-)^{a^{\dagger} a}, \quad (2.17)$$

whereas for  $s = 1$  one recovers (2.6) that defines the  $P$  function.

## 2.3 Elements of detection theory

Here we evaluate the probability distribution of the photocurrent of photodetectors, balanced homodyne detectors, and heterodyne detectors. We show that under suitable limits the respective photocurrents provide the measurement of the photon number distribution, of the quadrature, and of the complex amplitude of a single mode of the electromagnetic field. When the effect of non-unit quantum efficiency is taken into account an additional noise affects the measurement, giving a Bernoulli convolution for photo-detection, and a Gaussian convolution for homodyne and heterodyne detection. Extensive use of the results in this section will be made in the context of quantum homodyne tomography.

### 2.3.1 Photodetection

Light is revealed by exploiting its interaction with atoms/molecules or electrons in a solid, and, essentially, each photon ionizes a single atom or promotes an electron to a conduction band, and the resulting charge is then amplified to produce a measurable pulse. In practice, however, available photodetectors are not ideally counting all photons, and their performance is limited by a non-unit quantum efficiency  $\zeta$ . In fact, only a fraction  $\zeta$  of the incoming

photons lead to an electric signal, and ultimately to a *count*: some photons are either reflected from the surface of the detector, or are absorbed without being transformed into electric pulses.

Let us consider a light beam entering a photodetector of quantum efficiency  $\zeta$ , *i.e.* a detector that transforms just a fraction  $\zeta$  of the incoming light pulse into electric signal. If the detector is small with respect to the coherence length of radiation and its window is open for a time interval  $T$ , then the Poissonian process of counting gives a probability  $p(m; T)$  of revealing  $m$  photons that writes [39]

$$p(m; T) = \text{Tr} \left[ \rho : \frac{[\zeta I(T)T]^m}{m!} \exp[-\zeta I(T)T] : \right], \quad (2.18)$$

where  $\rho$  is the quantum state of light,  $::$  denotes the normal ordering of field operators, and  $I(T)$  is the beam intensity

$$I(T) = \frac{2\epsilon_0 c}{T} \int_0^T \mathbf{E}^{(-)}(\mathbf{r}, t) \cdot \mathbf{E}^{(+)}(\mathbf{r}, t) dt, \quad (2.19)$$

given in terms of the positive (negative) frequency part of the electric field operator  $\mathbf{E}^{(+)}(\mathbf{r}, t)$  ( $\mathbf{E}^{(-)}(\mathbf{r}, t)$ ). The quantity  $p(t) = \zeta \text{Tr} [\rho I(T)]$  equals the probability of a single count during the time interval  $(t, t + dt)$ . Let us now focus our attention to the case of the radiation field excited in a stationary state of a single mode at frequency  $\omega$ . Equation (2.18) can be rewritten as

$$p_\eta(m) = \text{Tr} \left[ \rho : \frac{(\eta a^\dagger a)^m}{m!} \exp(-\eta a^\dagger a) : \right], \quad (2.20)$$

where the parameter  $\eta = \zeta \hbar \omega / V$  denotes the overall *quantum efficiency* of the photodetector. Using (2.15) and (2.16) one obtains

$$p_\eta(m) = \sum_{n=m}^{\infty} \rho_{nn} \binom{n}{m} \eta^m (1-\eta)^{n-m}, \quad (2.21)$$

where  $\rho_{nn} \equiv \langle n | \rho | n \rangle = p_{\eta=1}(n)$ . Hence, for unit quantum efficiency a photodetector measures the photon number distribution of the state, whereas for non unit quantum efficiency the output distribution of counts is given by a Bernoulli convolution of the ideal distribution.

The effects of non unit quantum efficiency on the statistics of a photodetector, *i.e.* (2.21) for the output distribution, can be also described by means of a simple model in which the realistic photodetector is replaced with an ideal photodetector preceded by a beam splitter of transmissivity  $\tau \equiv \eta$ . The reflected mode is absorbed, whereas the transmitted mode is photo-detected with unit quantum efficiency. In order to obtain the probability of measuring  $m$  clicks, notice that, apart from trivial phase changes, a beam splitter of transmissivity  $\tau$  affects the unitary transformation of fields

$$\begin{pmatrix} c \\ d \end{pmatrix} \equiv U_\tau^\dagger \begin{pmatrix} a \\ b \end{pmatrix} U_\tau = \begin{pmatrix} \sqrt{\tau} & -\sqrt{1-\tau} \\ \sqrt{1-\tau} & \sqrt{\tau} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}, \quad (2.22)$$

where all field modes are considered at the same frequency. Hence, the output mode  $c$  hitting the detector is given by the linear combination

$$c = \sqrt{\tau}a - \sqrt{1-\tau}b, \quad (2.23)$$

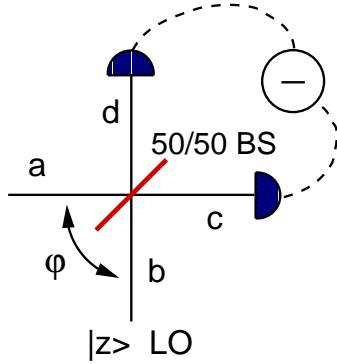
and the probability of counts reads

$$\begin{aligned} p_\tau(m) &= \text{Tr} [U_\tau \rho \otimes |0\rangle\langle 0| U_\tau^\dagger |m\rangle\langle m| \otimes 1] \\ &= \sum_{n=m}^{\infty} \rho_{nn} \binom{n}{m} (1-\tau)^{n-m} \tau^m. \end{aligned} \quad (2.24)$$

Equation (2.24) reproduces the probability distribution of (2.21) with  $\tau = \eta$ . We conclude that a photo-detector of quantum efficiency  $\eta$  is equivalent to a perfect photo-detector preceded by a beam splitter of transmissivity  $\eta$  which accounts for the overall losses of the detection process.

### 2.3.2 Balanced homodyne detection

The balanced homodyne detector provides the measurement of the quadrature of the field  $X_\varphi$  in (2.8). It was proposed by Yuen and Chan [40], and subsequently demonstrated by Abbas, Chan and Yee [41].



**Fig. 2.1.** Scheme of the balanced homodyne detector.

The scheme of a balanced homodyne detector is depicted in Fig. 2.1. The signal mode  $a$  interferes with a strong laser beam mode  $b$  in a balanced 50/50 beam splitter. The mode  $b$  is so-called *local oscillator* (LO) mode of the detector. It operates at the same frequency of  $a$ , and is excited by the laser

in a strong coherent state  $|z\rangle$ . Since in all experiments that use homodyne detectors the signal and the LO beams are generated by a common source, we assume that they have a fixed phase relation. In this case the LO phase provides a reference for the quadrature measurement, namely we identify the phase of the LO with the phase difference between the two modes. As we will see, by tuning  $\varphi = \arg z$  we can measure the quadrature  $X_\varphi$  at different phases.

After the beam splitter the two modes are detected by two identical photodetectors (usually linear avalanche photodiodes), and finally the difference of photocurrents at zero frequency is electronically processed and rescaled by  $2|z|$ . According to (2.22), the modes at the output of the 50/50 beam splitter ( $\tau = 1/2$ ) write

$$c = \frac{a - b}{\sqrt{2}}, \quad d = \frac{a + b}{\sqrt{2}}, \quad (2.25)$$

hence the difference of photocurrents is given by the following operator

$$I = \frac{d^\dagger d - c^\dagger c}{2|z|} = \frac{a^\dagger b + b^\dagger a}{2|z|}. \quad (2.26)$$

Let us now proceed to evaluate the probability distribution of the output photocurrent  $I$  for a generic state  $\rho$  of the signal mode  $a$ . In the following treatment we will follow [42, 43].

Let us consider the moments generating function of the photocurrent  $I$

$$\chi(\lambda) = \text{Tr} [\rho \otimes |z\rangle\langle z| e^{i\lambda I}], \quad (2.27)$$

which provides the probability distribution of  $I$  as the Fourier transform

$$P(I) = \int_{-\infty}^{+\infty} \frac{d\lambda}{2\pi} e^{-i\lambda I} \chi(\lambda). \quad (2.28)$$

Using the BCH formula [44, 45] for the  $SU(2)$  group, namely

$$\exp(\xi ab^\dagger - \xi^* a^\dagger b) = e^{\xi b^\dagger a} (1 + |\zeta|^2)^{\frac{1}{2}(b^\dagger b - a^\dagger a)} e^{-\zeta^* a^\dagger b}, \quad \zeta = \frac{\xi}{|\xi|} \tan|\xi|, \quad (2.29)$$

one can write the exponential in (2.27) in normal-ordered form with respect to mode  $b$  as follows

$$\chi(\lambda) = \left\langle e^{i \tan(\frac{\lambda}{2|z|}) b^\dagger a} \left[ \cos\left(\frac{\lambda}{2|z|}\right) \right]^{a^\dagger a - b^\dagger b} e^{i \tan(\frac{\lambda}{2|z|}) a^\dagger b} \right\rangle_{ab}. \quad (2.30)$$

Since mode  $b$  is in a coherent state  $|z\rangle$  the partial trace over  $b$  can be evaluated as follows

$$\begin{aligned}\chi(\lambda) &= \left\langle e^{i \tan(\frac{\lambda}{2|z|}) z^* a} \left[ \cos\left(\frac{\lambda}{2|z|}\right) \right]^{a^\dagger a} e^{i \tan(\frac{\lambda}{2|z|}) z a^\dagger} \right\rangle_a \\ &\times \left\langle z \left| \left[ \cos\left(\frac{\lambda}{2|z|}\right) \right]^{-b^\dagger b} \right| z \right\rangle.\end{aligned}\quad (2.31)$$

Using now (2.13), one can rewrite (2.31) in normal order with respect to  $a$ , namely

$$\chi(\lambda) = \left\langle e^{iz \sin(\frac{\lambda}{2|z|}) a^\dagger} \exp\left[-2 \sin^2\left(\frac{\lambda}{4|z|}\right) (a^\dagger a + |z|^2)\right] e^{iz^* \sin(\frac{\lambda}{2|z|}) a} \right\rangle_a \quad (2.32)$$

In the strong-LO limit  $z \rightarrow \infty$ , only the lowest order terms in  $\lambda/|z|$  are retained,  $a^\dagger a$  is neglected with respect to  $|z|^2$ , and (2.32) simplifies as follows

$$\lim_{z \rightarrow \infty} \chi(\lambda) = \left\langle e^{i\frac{\lambda}{2} e^{i\varphi} a^\dagger} \exp\left[-\frac{\lambda^2}{8}\right] e^{i\frac{\lambda}{2} e^{-i\varphi} a} \right\rangle_a = \langle \exp[i\lambda X_\varphi] \rangle_a, \quad (2.33)$$

where  $\varphi = \arg z$ . The generating function in (2.33) is then equivalent to the POVM

$$\Pi(x) = \int_{-\infty}^{+\infty} \frac{d\lambda}{2\pi} \exp[i\lambda(X_\varphi - x)] = \delta(X_\varphi - x) \equiv |x\rangle_{\varphi\varphi} \langle x|, \quad (2.34)$$

namely the projector on the eigenstate of the quadrature  $X_\varphi$  with eigenvalue  $x$ . In conclusion, the balanced homodyne detector achieves the ideal measurement of the quadrature  $X_\varphi$  in the strong LO limit. In this limit, the probability distribution of the output photocurrent  $I$  approaches exactly the probability distribution  $p(x, \varphi) = \varphi \langle x | \rho | x \rangle_\varphi$  of the quadrature  $X_\varphi$ , and this for any state  $\rho$  of the signal mode  $a$ .

It is easy to take into account non-unit quantum efficiency at detectors. According to (2.23) one has the replacements

$$c \implies \sqrt{\eta}c - \sqrt{1-\eta}u, \quad u, v \text{ vacuum modes} \quad (2.35)$$

$$d \implies \sqrt{\eta}d - \sqrt{1-\eta}v, \quad (2.36)$$

and now the output current is rescaled by  $2|z|\eta$ , namely

$$I_\eta \simeq \frac{1}{2|z|} \left\{ \left[ a + \sqrt{\frac{1-\eta}{2\eta}}(u+v) \right] b^\dagger + h.c \right\}, \quad (2.37)$$

where only terms containing the strong LO mode  $b$  are retained. The POVM is then obtained by replacing

$$X_\varphi \rightarrow X_\varphi + \sqrt{\frac{1-\eta}{2\eta}}(u_\varphi + v_\varphi) \quad (2.38)$$

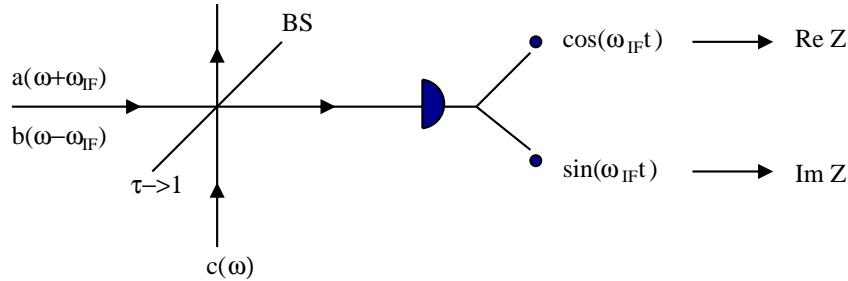
in (2.34), with  $w_\varphi = (w^\dagger e^{i\varphi} + w e^{-i\varphi})/2$ ,  $w = u, v$ , and tracing the vacuum modes  $u$  and  $v$ . One then obtains

$$\begin{aligned}\Pi_\eta(x) &= \int_{-\infty}^{+\infty} \frac{d\lambda}{2\pi} e^{i\lambda(X_\varphi - x)} |\langle 0 | e^{i\lambda\sqrt{\frac{1-\eta}{2\eta}} u_\varphi} | 0 \rangle|^2 = \int_{-\infty}^{+\infty} \frac{d\lambda}{2\pi} e^{i\lambda(X_\varphi - x)} e^{-\lambda^2 \frac{1-\eta}{8\eta}} \\ &= \frac{1}{\sqrt{2\pi\Delta_\eta^2}} \int_{-\infty}^{+\infty} dx' e^{-\frac{1}{2\Delta_\eta^2}(x-x')^2} |x'\rangle_\varphi \langle x'|,\end{aligned}\quad (2.39)$$

where  $\Delta_\eta^2 = \frac{1-\eta}{4\eta}$ . Thus the POVM, and in turn the probability distribution of the output photocurrent, are just the Gaussian convolution of the ideal ones with rms  $\Delta_\eta$ .

### 2.3.3 Heterodyne detection

Heterodyne detection allows one to perform the joint measurement of two conjugated quadratures of the field [46, 47]. The scheme of the heterodyne detector is depicted in Fig. 2.2.



**Fig. 2.2.** Scheme of the heterodyne detector.

A strong local oscillator at frequency  $\omega$  in a coherent state  $|\alpha\rangle$  hits a beam splitter with transmissivity  $\tau \rightarrow 1$ , and with the coherent amplitude  $\alpha$  such that  $\gamma \equiv |\alpha|\sqrt{\tau(1-\tau)}$  is kept constant. If the output photocurrent is sampled at the intermediate frequency  $\omega_{IF}$ , just the field modes  $a$  and  $b$  at frequency  $\omega \pm \omega_{IF}$  are selected by the detector. Modes  $a$  and  $b$  are usually referred to as signal band and image band modes, respectively. In the strong LO limit, upon tracing the LO mode, the output photocurrent  $I(\omega_{IF})$  rescaled by  $\gamma$  is equivalent to the complex operator

$$Z = \frac{I(\omega_{IF})}{\gamma} = a - b^\dagger,\quad (2.40)$$

where the arbitrary phases of modes have been suitably chosen. The heterodyne photocurrent  $Z$  is a normal operator, equivalent to a couple of commuting selfadjoint operators

$$Z = \text{Re}Z + i\text{Im}Z, \quad [Z, Z^\dagger] = [\text{Re}Z, \text{Im}Z] = 0. \quad (2.41)$$

The POVM of the detector is then given by the orthogonal eigenvectors of  $Z$ . It is convenient to introduce the notation of [48] for vectors in the tensor product of Hilbert spaces  $\mathcal{H} \otimes \mathcal{H}$

$$|A\rangle\langle\! \rangle = \sum_{nm} A_{nm} |n\rangle \otimes |m\rangle \equiv (A \otimes I) |I\rangle\langle\! \rangle \equiv (I \otimes A^\tau) |I\rangle\langle\! \rangle, \quad (2.42)$$

where  $A^\tau$  denotes the transposed operator with respect to some pre-chosen orthonormal basis. Equation (2.42) exploits the isomorphism between the Hilbert space of the Hilbert-Schmidt operators  $A, B \in \text{HS}(\mathcal{H})$  with scalar product  $\langle A, B \rangle = \text{Tr}[A^\dagger B]$ , and the Hilbert space of bipartite vectors  $|A\rangle\langle\! \rangle, |B\rangle\langle\! \rangle \in \mathcal{H} \otimes \mathcal{H}$ , where one has  $\langle\langle A|B \rangle\langle\! \rangle \equiv \langle A, B \rangle$ .

Using the above notation it is easy to write the eigenvectors of  $Z$  with eigenvalue  $z$  as  $\frac{1}{\sqrt{\pi}} |D(z)\rangle\langle\! \rangle$ . In fact one has [49]

$$\begin{aligned} Z |D(z)\rangle\langle\! \rangle &= (a - b^\dagger)(D_a(z) \otimes I_b) |I\rangle\langle\! \rangle = (D_a(z) \otimes I_b)(a - b^\dagger + z) \sum_{n=0}^{\infty} |n\rangle \otimes |n\rangle \\ &= z(D_a(z) \otimes I_b) |I\rangle\langle\! \rangle = z |D(z)\rangle\langle\! \rangle. \end{aligned} \quad (2.43)$$

The orthogonality of such eigenvectors can be verified through the relation

$$\langle\langle D(z)|D(z')\rangle\langle\! \rangle = \text{Tr}[D^\dagger(z)D(z')] = \pi\delta^{(2)}(z - z'), \quad (2.44)$$

where  $\delta^{(2)}(\alpha)$  denotes the Dirac delta function over the complex plane

$$\delta^{(2)}(\alpha) = \int_{\mathbb{C}} \frac{d^2\gamma}{\pi^2} \exp(\gamma\alpha^* - \gamma^*\alpha). \quad (2.45)$$

In conventional heterodyne detection the image band mode is in the vacuum state, and one is just interested in measuring the field mode  $a$ . In this case we can evaluate the POVM upon tracing on mode  $b$ . One has

$$\begin{aligned} \Pi(z, z^*) &= \frac{1}{\pi} \text{Tr}_b [|D(z)\rangle\langle\! \rangle |D(z)| I_a \otimes |0\rangle\langle 0|] \\ &= \frac{1}{\pi} D(z) |0\rangle\langle 0| D^\dagger(z) = \frac{1}{\pi} |z\rangle\langle z|. \end{aligned} \quad (2.46)$$

Namely, one obtains the projectors on coherent states. The coherent-state POVM provides the optimal joint measurement of conjugated quadratures of the field [50]. In fact, heterodyne detection allows one to measure the Q-function in (2.4). According to (2.3) then it provides the expectation value of anti-normal ordered field operator. For a state  $\rho$  the expectation value of any quadrature  $X_\varphi$  is obtained as

$$\langle X_\varphi \rangle = \text{Tr}[\rho X_\varphi] = \int_{\mathbb{C}} \frac{d^2\alpha}{\pi} \text{Re}(\alpha e^{-i\varphi}) Q(\alpha, \alpha^*). \quad (2.47)$$

The price to pay for jointly measuring a pair of non commuting observables is an additional noise. The rms fluctuation is evaluated as follows

$$\int_{\mathbb{C}} \frac{d^2\alpha}{\pi} [\text{Re}(\alpha e^{-i\varphi})]^2 Q(\alpha, \alpha^*) - \langle X_\varphi \rangle^2 = \langle \Delta X_\varphi^2 \rangle + \frac{1}{4}, \quad (2.48)$$

where  $\langle \Delta X_\varphi^2 \rangle$  is the intrinsic noise, and the additional term is usually referred to as “the additional 3dB noise due to the joint measure” [51–53].

The effect of non-unit quantum efficiency can be taken into account in analogous way as in Sect. 2.3.2 for homodyne detection. The heterodyne photocurrent is rescaled by an additional factor  $\eta^{1/2}$ , and vacuum modes  $u$  and  $v$  are introduced, thus giving [54]

$$Z_\eta = a - b^\dagger + \sqrt{\frac{1-\eta}{\eta}} u - \sqrt{\frac{1-\eta}{\eta}} v^\dagger. \quad (2.49)$$

Upon tracing over modes  $u$  and  $v$ , one obtain the POVM

$$\begin{aligned} \Pi_\eta(z, z^*) &= \int_{\mathbb{C}} \frac{d^2\gamma}{\pi^2} u \langle 0|_v \langle 0| e^{\gamma(Z_\eta^\dagger - z^*) - \gamma^*(Z_\eta - z)} |0\rangle_u |0\rangle_v \\ &= \int_{\mathbb{C}} \frac{d^2\gamma}{\pi^2} e^{\gamma(Z^\dagger - z^*) - \gamma^*(Z - z)} e^{-\frac{1-\eta}{\eta}|\gamma|^2} \\ &= \frac{\eta}{\pi(1-\eta)} e^{-\frac{\eta}{1-\eta}|Z-z|^2} = \int_{\mathbb{C}} \frac{d^2z'}{\pi\Delta_\eta^2} e^{-\frac{|z'-z|^2}{\Delta_\eta^2}} |D(z')\rangle\langle D(z')|. \end{aligned} \quad (2.50)$$

The probability distribution is then a Gaussian convolution on the complex plane of the ideal probability with rms  $\Delta_\eta^2 = (1-\eta)/\eta$ .

Analogously, the coherent-state POVM for conventional heterodyne detection with vacuum image band mode is replaced with

$$\Pi_\eta(z, z^*) = \int_{\mathbb{C}} \frac{d^2z'}{\pi\Delta_\eta^2} e^{-\frac{|z'-z|^2}{\Delta_\eta^2}} |z'\rangle\langle z'|. \quad (2.51)$$

From (2.9) we can equivalently say that the heterodyne detection probability density is given by the generalized Wigner function  $W_s(\alpha, \alpha^*)$ , with  $s = 1 - \frac{2}{\eta}$ . Notice that for  $\eta < 1$ , the average of functions  $\alpha^n \alpha^{*m}$  is related to the expectation value of a different ordering of field operators. However, one has the relevant identity [29, 55]

$$:(a^\dagger)^n a^m:_s = \sum_{k=0}^{(n,m)} k! \binom{n}{k} \binom{m}{k} \left(\frac{t-s}{2}\right)^k :(a^\dagger)^{n-k} a^{m-k}:_t, \quad (2.52)$$

where  $(n, m) = \min(n, m)$ , and then

$$\begin{aligned} &\int_{\mathbb{C}} d^2\alpha W_{1-\frac{2}{\eta}}(\alpha, \alpha^*) \alpha^m \alpha^{*n} \\ &= \sum_{k=0}^{(n,m)} k! \binom{n}{k} \binom{m}{k} \left(\frac{1-\eta}{\eta}\right)^k \langle a^{m-k} (a^\dagger)^{n-k} \rangle. \end{aligned} \quad (2.53)$$

Notice that the measure of the Q-function (or any smoothed version for  $\eta < 1$ ) does not allow one to recover the expectation value of *any* operator through an average over heterodyne outcomes. In fact, one needs the admissibility of anti-normal ordered expansion [56] and the convergence of the integral in (2.53). In particular, the matrix elements of the density operator cannot be recovered.

Finally, it is worth mentioning that the above results hold also for an image-band mode with the same frequency of the signal. In this case a measurement scheme based on multiport homodyne detection should be used [47, 55, 57–63].

## 2.4 General tomographic method

In the first part of the Section a brief history of tomography is presented. Then, we give a sketch of the conventional medical tomography, and we show its analogy with the optical homodyne tomography for the reconstruction of the Wigner function proposed by Vogel and Risken [7]. However the limits and the intrinsic unreliability of this method are explained.

The first exact method was given in [8], and successively refined in [9]. It allows the reconstruction of the density matrix  $\rho$ , bypassing the inversion of the Wigner function. Analogously, it provides the expectation values of arbitrary operators, directly as an average of “Kernel functions” evaluated on the experimental data collected by homodyne detection.

The general tomographic method is presented in Sect. 2.4.4. The concept of “quorum”, namely the complete set of observables whose measurement provides the expectation value of any desired operator, is introduced. We show that some “orthogonality” and “completeness” relations in the linear algebra of operators are sufficient to individuate a quorum [26].

In Sect. 2.4.9 some developments of the basic tomographic method are shown. First, the deconvolution of noise given by the imperfections of detectors and/or experimental apparatus. Such noise can be eliminated under the hypothesis that the pertaining CP-map is invertible [20]. Then, we show that also the statistical random noise can be reduced through the adaptive tomography technique [22].

The relevant topic of multimode tomography with a single oscillator is given separate treatment in the following Section.

### 2.4.1 Brief historical excursus

The problem of quantum state determination through repeated measurements on identically prepared systems was already stated in 1957 by Fano [5]. He was aware that more than two observables are needed for this purpose. It

was only with the proposal by Vogel and Risken [7] however, that quantum tomography was born. The first experiments, which already showed reconstructions of coherent and squeezed states were performed in Michael Raymer's group at the University of Oregon [6]. The main idea at the basis of the first proposal is that it is possible to extend to the quantum domain the algorithms that are conventionally used in medical imaging to recover two dimensional distributions (say of mass) from unidimensional projections in different directions. However, the first tomographic method is unreliable for the measurement of unknown quantum states, since some arbitrary smoothing parameters have to be introduced. The exact unbiased tomographic procedure was proposed in [8], and successively simplified in [9]. The exact homodyne method has been implemented experimentally to measure the photon statistics of a semiconductor laser [10], and the density matrix of a squeezed vacuum [11]. The success of optical homodyne tomography has then stimulated the development of state-reconstruction procedures for atomic beams [12], the experimental determination of the vibrational state of a molecule [13], of an ensemble of helium atoms [14], and of a single ion in a Paul trap [15].

More recently, quantum tomography has been generalized to the estimation of an arbitrary observable of the field [16], with any number of modes [17], and, finally, to arbitrary quantum systems via group theory [18, 20, 21]. Further developments such as noise deconvolution [20] and adaptive tomography [22] were found. The use of max-likelihood strategies [23] has made possible to reduce dramatically the number of experimental data (by a factor  $10^3 \div 10^5$ !) with negligible bias for most practical cases of interest. Finally, very recently, a method for tomographic estimation of the unknown quantum operation of a quantum device has been presented [27], where a fixed input entangled state is used. Similarly, one can also estimate the ensemble average of all operators by measuring only one fixed "universal" observable on an extended Hilbert space [64]. The latest development [26] deduces the general tomographic method from the property of spanning sets of operators. In fact, the group structure is not necessary to individuate a "quorum", but just some "orthogonality" and "completeness" relations in the linear algebra of operators are sufficient to that purpose. The general method will be presented in this context.

#### 2.4.2 Conventional tomographic imaging

In conventional medical tomography, one collects data in the form of marginal distributions of the mass function  $m(x, y)$ . In the complex plane the marginal  $r(x, \varphi)$  is a projection of the complex function  $m(x, y)$  on the direction indicated by the angle  $\varphi \in [0, \pi]$ , namely

$$r(x, \varphi) = \int_{-\infty}^{+\infty} \frac{dy}{\pi} m((x + iy)e^{i\varphi}, (x - iy)e^{-i\varphi}). \quad (2.54)$$

The collection of marginals for different  $\varphi$  is called ‘‘Radon transform’’. The tomography process essentially consists in the inversion of the Radon transform (2.54), in order to recover the mass function  $m(x, y)$  from the marginals  $r(x, \varphi)$ .

Here we derive inversion of (2.54). Consider the identity

$$m(\alpha, \alpha^*) = \int_{\mathbb{C}} d^2\beta \delta^{(2)}(\alpha - \beta) m(\beta, \beta^*) , \quad (2.55)$$

where  $\delta^{(2)}(\alpha)$  denotes the Dirac delta function of (2.45), and  $m(\alpha, \alpha^*) \equiv m(x, y)$ , with  $\alpha = x + iy$  and  $\alpha^* = x - iy$ . Equation (2.45) rewrites

$$\delta^{(2)}(\alpha) = \int_0^{+\infty} \frac{dk}{4} k \int_0^{2\pi} \frac{d\varphi}{\pi^2} e^{-ik\alpha_\varphi} = \int_{-\infty}^{+\infty} \frac{dk}{4} |k| \int_0^\pi \frac{d\varphi}{\pi^2} e^{-ik\alpha_\varphi}, \quad (2.56)$$

with  $\alpha_\varphi \equiv \text{Re}(\alpha e^{-i\varphi}) = -\alpha_{\varphi+\pi}$ . From (2.55) and (2.56) the inverse Radon transform is obtained as follows

$$m(x, y) = \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} dx' r(x', \varphi) \int_{-\infty}^{+\infty} \frac{dk}{4} |k| e^{ik(x' - \alpha_\varphi)} . \quad (2.57)$$

Equation (2.57) is conventionally written as

$$m(x, y) = \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} dx' r(x', \varphi) K(x' - \alpha_\varphi), \quad (2.58)$$

where  $K(x)$  is given by

$$K(x) \equiv \int_{-\infty}^{+\infty} \frac{dk}{4} |k| e^{ikx} = \frac{1}{2} \text{Re} \int_0^{+\infty} dk k e^{ikx} = -\frac{1}{2} \mathcal{P} \frac{1}{x^2}, \quad (2.59)$$

with  $\mathcal{P}$  denoting the Cauchy principal value. Integrating by parts (2.58) one obtains the tomographic formula that is usually found in medical imaging, *i.e.*

$$m(x, y) = \frac{1}{2\pi} \int_0^\pi d\varphi \mathcal{P} \int_{-\infty}^{+\infty} dx' \frac{1}{x' - \alpha_\varphi} \frac{\partial}{\partial x'} r(x', \varphi) , \quad (2.60)$$

which allows the reconstruction of the mass distribution  $m(x, y)$  from its projections along different directions  $r(x, \varphi)$ .

### 2.4.3 Extension to the quantum domain

In the quantum imaging process one would like to reconstruct a quantum state in the form of its Wigner function, by starting from its marginal probability distributions. As shown in Sect. 2.2, the Wigner function is a real normalized function that is in one-to-one correspondence with the state density

operator  $\rho$ . As noticed in (2.7), the probability distributions of the quadrature operators  $X_\varphi = (a^\dagger e^{i\varphi} + ae^{-i\varphi})/2$  are the marginal probabilities of the Wigner function for the state  $\rho$ . Thus, by applying the same procedure outlined in the previous subsection, Vogel and Risken [7] proposed a method to recover the Wigner function *via* an inverse Radon transform from the quadrature probability distributions  $p(x, \varphi)$ , namely

$$W(x, y) = \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} dx' p(x', \varphi) \int_{-\infty}^{+\infty} \frac{dk}{4} |k| e^{ik(x' - x \cos \varphi - y \sin \varphi)} \quad (2.61)$$

Notice that in the original paper [7] conventional tomographic imaging is never referred to. As shown in Sect. 2.3.2 the experimental data, distributed according to the quadrature probability density, can be obtained by using the homodyne detector which measures the quadrature of the field. The method proposed by Vogel and Risken, namely the inversion of the Radon transform, was the one which has been used in the first experiments [6].

However, this first method is unreliable for the reconstruction of unknown quantum states, since there is an intrinsic unavoidable systematic error. In fact the integral on  $k$  in (2.61) is unbounded. In order to use the inverse Radon transform, one would need the analytical form of the marginal distribution of the quadrature  $p(x, \varphi)$ . This can be obtained by collecting the experimental data into histograms and splining these histograms. This is not an unbiased procedure since the degree of splining, the width of the histogram bins and the number of different phases on which the experimental data should be collected are arbitrary parameters and introduce systematic errors whose effects cannot be easily controlled. For example, the effect of using high degrees of splining is the wash-out of the quantum features of the state, and, *vice-versa*, the effect of low degrees of splining is to create negative bias for the probabilities in the reconstruction (see [8] for details).

A new approach to optical tomography was proposed in [8]. This approach, that will be referred to as ‘quantum homodyne tomography’, allows one to recover the quantum state of the field  $\rho$  (and also the mean values of arbitrary operators) directly from the data, abolishing all the sources of systematic errors. Only statistical errors are present, and they can be reduced arbitrarily by collecting more experimental data. The correct method will be derived from the general tomographic theory in Sect. 2.4.5.

#### 2.4.4 General method of quantum tomography

In the following the general method of quantum tomography will be explained. First, we give the basics of Monte Carlo integral theory which are needed to implement the tomographic algorithms in actual experiments and in numerical simulations. Then, we derive the formulas on which all schemes of state reconstruction are based.

### Basic statistics

The aim of quantum tomography is to estimate, in arbitrary quantum systems, the mean value  $\langle O \rangle$  of a system operator  $O$  using only the results of the measurements on a set of observables  $\{Q_\lambda, \lambda \in \Lambda\}$ , called “quorum”. The procedure by which this can be obtained needs the so called “Kernel function”  $\mathcal{R}_\lambda[O](x_\lambda)$  which is a function of the eigenvalues  $x_\lambda$  of the quorum operators. Integrating the Kernel with the probability  $p_\lambda(x_\lambda)$  of having outcome  $x_\lambda$  when measuring  $Q_\lambda$ , the mean value of  $O$  is obtained as follows

$$\langle O \rangle = \int_{\Lambda} d\lambda \int_{-\infty}^{+\infty} dx_\lambda p_\lambda(x_\lambda) \mathcal{R}_\lambda[O](x_\lambda), \quad (2.62)$$

where the first integral is performed on the values of  $\lambda$  that designate all quorum observables, and the second on the eigenvalues of the quorum observable  $Q_\lambda$  determined by the  $\lambda$  variable of the outer integral. Both integrals in (2.62) can suitably replaced by sums.

The algorithm to estimate  $\langle O \rangle$  with (2.62) is the following. One chooses a quorum operator  $Q_\lambda$  by choosing  $\lambda$  with uniform probability in  $\Lambda$  and performs a measurement, obtaining the result  $x_i$ . By repeating the procedure  $N$  times, one collects the set of experimental data  $\{(\lambda_i, x_i), \text{ with } i = 1, \dots, N\}$ , where  $\lambda_i$  identifies the quorum observable used for the  $i$ th measurement, and  $x_i$  its result. From the same set of data the mean value of any operator  $O$  can be obtained. In fact, one evaluates the Kernel function for  $O$  and the quorum  $Q_\lambda$ , and then samples the double integral of (2.62) using the limit

$$\langle O \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathcal{R}_{\lambda_i}[O](x_i). \quad (2.63)$$

Of course the finite sum,

$$F_N = \frac{1}{N} \sum_{i=1}^N \mathcal{R}_{\lambda_i}[O](x_i), \quad (2.64)$$

gives an approximation of  $\langle O \rangle$ . To estimate the error in the approximation one applies the central limit theorem that we recall here.

**Central limit theorem.** Consider  $N$  statistically uncorrelated random variables  $\{z_i, i = 1, \dots, N\}$ , with mean values  $\mu(z_i)$ , variances  $\sigma^2(z_i)$  and bounded third order moments. If the variances  $\sigma^2(z_i)$  are all of the same order then the statistical variable “sum”  $y$  defined as

$$y = \sum_{i=1}^N z_i \quad (2.65)$$

has mean and variance

$$\mu(y) = \sum_{i=1}^N \mu(z_i), \quad \sigma^2(y) = \sum_{i=1}^N \sigma^2(z_i). \quad (2.66)$$

The distribution of  $y$  approaches asymptotically a Gaussian for  $N \rightarrow \infty$ . In practical cases, the distribution of  $y$  can be considered Gaussian already for  $N$  as low as  $N \sim 10$ .

For our needs the hypotheses are met if the Kernel function  $\mathcal{R}_{\lambda_i}[O](x_i)$  in (2.64) has limited moments up to the third order, since  $x_i$  come from the same probability density, and hence all  $z_i = \frac{1}{N}\mathcal{R}_{\lambda_i}[O](x_i)$  have the same mean  $\mu(z_i) = \frac{1}{N}\langle O \rangle$  and variance

$$\sigma^2(z_i) = \frac{1}{N^2} \left[ \lim_{M \rightarrow \infty} \sum_{j=1}^M \mathcal{R}_{\lambda_i}^2[O](x_j) - \langle O \rangle^2 \right]. \quad (2.67)$$

Using the central limit theorem, we can conclude that the experimental average  $y \equiv F_N$  in (2.64) is a statistical variable distributed as a Gaussian with mean value  $\mu(y) = \langle O \rangle$  and variance

$$\sigma^2(y) = \frac{1}{N} \left[ \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{j=1}^M \mathcal{R}_{\lambda_i}^2[O](x_j) - \langle O \rangle^2 \right]. \quad (2.68)$$

Then the tomographic estimation converges with statistical error that decreases as  $N^{-1/2}$ .

Since the statistical variable  $F_N$  converges to  $\langle O \rangle$  and is distributed as a Gaussian we can also evaluate the statistical error that affects the tomographic reconstruction. Upon dividing the experimental data into  $M$  statistical blocks of equal dimension one evaluates the average in (2.64) for each block. A set  $F_n$  ( $n = 1, 2, \dots, M$ ) is then obtained and it is Gaussian distributed with mean value

$$m = \frac{1}{M} \sum_{n=1}^M F_n \quad (2.69)$$

and variance

$$s^2 = \frac{1}{M-1} \sum_{n=1}^M (F_n - m)^2. \quad (2.70)$$

Notice that the factor  $1/(M-1)$  in the variance comes from the fact that we are using the “experimental” estimated mean value  $m$  in place of the “real” one  $\mu$ . The variance of the mean  $m$  is given by

$$\sigma^2(m) \equiv \sigma^2 \left( \frac{1}{M} \sum_{n=1}^M F_n \right) = \frac{1}{M^2} \sum_{n=1}^M \sigma^2(F_n) = \frac{1}{M} s^2, \quad (2.71)$$

and thus the error on the mean  $m$  estimated from the data is given by

$$\epsilon = \frac{s}{\sqrt{M}} = \sqrt{\sum_{n=1}^M \frac{(F_n - m)^2}{M(M-1)}}. \quad (2.72)$$

The usual statistical interpretation applies: the “real” value  $\mu = \langle O \rangle$  is to be found in the interval  $[m - \epsilon, m + \epsilon]$  with  $\sim 68\%$  probability, in  $[m - 2\epsilon, m + 2\epsilon]$  with  $\sim 95\%$  probability, and in  $[m - 3\epsilon, m + 3\epsilon]$  with  $\sim$  unit probability. In order to test that the confidence intervals are estimated correctly, one can check that the  $F_n$  distribution is actually Gaussian. This can be done by comparing the histogram of the block data with a Gaussian, or by using the  $\chi^2$  test.

### Characterization of the quorum

As we will see, different estimation techniques have been proposed tailored to different systems, such as the radiation field [9,17], trapped ions and molecular vibrational states [65], spin systems [66], etc. As a matter of fact, all these schemes can be embodied in the following approach.

The tomographic reconstruction of an operator  $O$  is possible when there exists a resolution of the form

$$O = \int_A d\lambda \operatorname{Tr} [OB^\dagger(\lambda)] C(\lambda), \quad (2.73)$$

where  $\lambda$  is a (possibly multidimensional) parameter living on a (discrete or continuous) manifold  $A$ . The only hypothesis in (2.73) is the existence of the trace. If, for example,  $O$  is a trace-class operator, then we do not need to require  $B(\lambda)$  to be of Hilbert-Schmidt class, since it is sufficient to require  $B(\lambda)$  bounded. The operators  $C(\lambda)$  are functions of the quorum of observables measured for the reconstruction, whereas the operators  $B(\lambda)$  form the *dual basis* of the set  $C(\lambda)$ . The term

$$\mathcal{E}[O](\lambda) = \operatorname{Tr} [OB^\dagger(\lambda)] C(\lambda) \quad (2.74)$$

represents the quantum estimator for the operator  $O$ . The expectation value of  $O$  is given by the ensemble average

$$\langle O \rangle \equiv \operatorname{Tr} [O\rho] = \int_A d\lambda \operatorname{Tr} [OB^\dagger(\lambda)] \operatorname{Tr} [C(\lambda)\rho] \equiv \int_A d\lambda \langle \mathcal{E}[O](\lambda) \rangle, \quad (2.75)$$

where  $\rho$  is the density matrix of the quantum system under investigation. Notice that the quantity  $\operatorname{Tr} [C(\lambda)\rho]$  depends only on the quantum state, and it is related to the probability distribution of the measurement outcomes, whereas the term  $\operatorname{Tr} [OB^\dagger(\lambda)]$  depends only on the quantity to be measured. In particular, the tomography of the quantum state of a system corresponds

to writing (2.73) for the operators  $O = |k\rangle\langle n|$ ,  $\{|n\rangle\}$  being a given Hilbert space basis. For a given system, the existence of a set of operators  $C(\lambda)$ , together with its dual basis  $B(\lambda)$  allows universal quantum estimation, i. e. the reconstruction of any operator.

We now give two characterizations of the sets  $B(\lambda)$  and  $C(\lambda)$  that are necessary and sufficient conditions for writing (2.73).

**Condition 1: bi-orthogonality**

Let us consider a complete orthonormal basis of vectors  $|n\rangle$  ( $n = 0, 1, \dots$ ). Formula (2.73) is equivalent to the bi-orthogonality condition

$$\int_A d\lambda \langle q|B^\dagger(\lambda)|p\rangle \langle m|C(\lambda)|l\rangle = \delta_{mp}\delta_{lq}, \quad (2.76)$$

where  $\delta_{ij}$  is the Kronecker delta. Equation (2.76) can be straightforwardly generalized to a continuous basis.

**Condition 2: completeness**

If the set of operators  $C(\lambda)$  is *irreducible*, namely if any operator can be written as a linear combination of the  $C(\lambda)$  as

$$O = \int_A d\lambda a(\lambda) C(\lambda), \quad (2.77)$$

then (2.73) is also equivalent to the trace condition

$$\text{Tr}[B^\dagger(\lambda) C(\mu)] = \delta(\lambda, \mu), \quad (2.78)$$

where  $\delta(\lambda, \mu)$  is a *reproducing kernel* for the set  $B(\lambda)$ , namely a function or a tempered distribution which satisfies

$$\int_A d\lambda B(\lambda) \delta(\lambda, \mu) = B(\mu). \quad (2.79)$$

An analogous identity holds for the set of  $C(\lambda)$

$$\int_A d\lambda C(\lambda) \delta(\lambda, \mu) = C(\mu). \quad (2.80)$$

The proofs are straightforward. The irreducibility condition on the operators  $C(\lambda)$  is essential for the equivalence of (2.73) and (2.78). A simple counterexample is provided by the set of projectors  $P(x) = |x\rangle\langle x|$  over the eigenstates of a selfadjoint operator  $X$ . In fact, (2.78) is satisfied by  $P(x)$ . However, since they do not form an irreducible set, it is not possible to express a generic operator as  $O \neq \int_X dx \langle x|O|x\rangle |x\rangle\langle x|$ .

If either the set  $B(\lambda)$  or the set  $C(\lambda)$  satisfy the additional trace condition

$$\text{Tr}[B^\dagger(\mu)B(\lambda)] = \delta(\lambda, \mu), \quad (2.81)$$

$$\text{Tr}[C^\dagger(\mu)C(\lambda)] = \delta(\lambda, \mu), \quad (2.82)$$

then we have  $C(\lambda) = B(\lambda)$  (notice that neither  $B(\lambda)$  nor  $C(\lambda)$  need to be unitary). In this case, (2.73) can be rewritten as

$$O = \int_A d\lambda \operatorname{Tr} [OC^\dagger(\lambda)] C(\lambda) . \quad (2.83)$$

A set of observables  $Q_\lambda$  constitute a quorum when there are functions  $f_\lambda(Q_\lambda) = C(\lambda)$  so that  $C(\lambda)$  form an irreducible set. The quantum estimator for  $O$  in (2.74) then writes as a function of the quorum operators

$$\mathcal{E}[O](\lambda) \equiv \mathcal{E}_\lambda[O](Q_\lambda) . \quad (2.84)$$

Notice that if a set of observables  $Q_\lambda$  constitutes a quorum, than the set of projectors  $|q\rangle_{\lambda\lambda}\langle q|$  over their eigenvectors provides a quorum too, with the measure  $d\lambda$  in (2.73) including the measure  $dq$ . Notice also that, even once the quorum has been fixed, the unbiased estimator for an operator  $O$  will not in general be unique, since there can exist functions  $\mathcal{N}(Q_\lambda)$  that satisfies

$$\int_A d\lambda \mathcal{N}(Q_\lambda) = 0 , \quad (2.85)$$

and that will be called ‘null estimators’. Two unbiased estimators that differ by a null estimator yield the same results when estimating the operator mean value. We will see in Sect. 2.4.9 how the null estimators can be used to reduce the statistical noise.

In terms of a quorum of observables  $Q_\lambda$  (2.75) rewrites

$$\begin{aligned} \langle O \rangle &= \int_A d\lambda \operatorname{Tr} [OB^\dagger(\lambda)] \operatorname{Tr} [f_\lambda(Q_\lambda)\rho] \\ &= \int_A d\lambda \int dq_\lambda p_\lambda(q_\lambda) \operatorname{Tr} [OB^\dagger(\lambda)] f_\lambda(q_\lambda) , \end{aligned} \quad (2.86)$$

where  $p_\lambda(q_\lambda) = \lambda \langle q|\rho|q \rangle_\lambda$  is the probability density of getting the outcome  $q_\lambda$  from the measurement of  $Q_\lambda$  on the state  $\rho$ . Equation (2.86) is equivalent to the expression (2.62), with the Kernel function

$$\mathcal{R}_\lambda[O](q_\lambda) = \operatorname{Tr} [OB^\dagger(\lambda)] f_\lambda(q_\lambda) . \quad (2.87)$$

Of course it is of interest to connect a quorum of observables to a resolution of the form (2.73), since only in this case can there be a feasible reconstruction scheme. If a resolution formula is written in terms of a set of selfadjoint operators, the set itself constitutes the desired quorum. However, in general a quorum of observables is functionally connected to the corresponding resolution formula. If the operators  $C(\lambda)$  are unitary, then they can always be considered as exponential of a set of selfadjoint operators, say  $Q_\lambda$ . The quantity  $\operatorname{Tr} [C(\lambda)\rho]$  is thus connected with the moment generating function of the set  $Q_\lambda$ , and hence to the probability density  $p_\lambda(q_\lambda)$  of the

measurement outcomes, which play the role of the Radon transform in the quantum tomography of the harmonic oscillator. In general, the operators  $C(\lambda)$  can be any function (neither self-adjoint nor unitary) of observables and, even more generally, they may be connected to POVMs rather than observables.

The dual set  $B(\lambda)$  can be obtained from the set  $C(\lambda)$  by inverting (2.78). For finite quorums, this resorts to a matrix inversion or, alternatively, to a Gram-Schmidt orthogonalization procedure [26]. No such a general procedure exists for a continuous spanning set. Many cases, however, satisfy conditions (2.81) and (2.82), and thus we can write  $B(\lambda) = C^\dagger(\lambda)$ .

#### 2.4.5 Quantum estimation for harmonic system

The harmonic oscillator models several systems of interest in quantum mechanics, as the vibrational states of molecules, the motion of an ion in a Paul trap, and a single mode radiation field. Different proposals have been suggested in order to reconstruct the quantum state of a harmonic system. They can be summarized using the framework of the previous subsection, which is also useful for devising novel estimation techniques. Here, the basic resolution formula involves the set of displacement operators  $D(\alpha) = \exp(\alpha a^\dagger - \alpha^* a)$ , which can be viewed as exponentials of the field-quadrature operators  $X_\varphi = (a^\dagger e^{i\varphi} + a e^{-i\varphi})/2$ . We have shown in Sect. 2.3.2 that for a single-mode radiation field  $X_\varphi$  is measured through homodyne detection. For the vibrational tomography of a molecule or a trapped ion  $X_\varphi$  corresponds to a time-evolved position or momentum. The set of displacement operators satisfies (2.78) and (2.82), since

$$\text{Tr}[D(\alpha)D^\dagger(\beta)] = \pi\delta^{(2)}(\alpha - \beta) , \quad (2.88)$$

whereas (2.83) reduces to the Glauber formula

$$O = \int_{\mathbb{C}} \frac{d^2\alpha}{\pi} \text{Tr}[OD^\dagger(\alpha)] D(\alpha) . \quad (2.89)$$

Changing to polar variables  $\alpha = (-i/2)ke^{i\varphi}$ , (2.89) becomes

$$O = \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} \frac{dk|k|}{4} \text{Tr}[O e^{ikX_\varphi}] e^{-ikX_\varphi} , \quad (2.90)$$

which shows explicitly the dependence on the quorum  $X_\varphi$ . Taking the ensemble average of both members and evaluating the trace over the set of eigenvectors of  $X_\varphi$ , one obtains

$$\langle O \rangle = \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} dx p(x, \varphi) \mathcal{R}[O](x, \varphi) , \quad (2.91)$$

where  $p(x; \varphi) = {}_\varphi\langle x|\rho|x\rangle_\varphi$  is the probability distribution of quadratures outcome. The Kernel function for the operator  $O$  is given by

$$\mathcal{R}[O](x, \varphi) = \text{Tr}[OK(X_\varphi - x)], \quad (2.92)$$

where  $K(x)$  is the same as in (2.59).

Equation (2.91) is the basis of quantum homodyne tomography. Notice that the operator  $K(x)$  is unbounded, however any matrix element  $\langle\psi|\rho|\phi\rangle$  such that  $\langle\psi|K(X_\varphi - x)|\phi\rangle$  is bounded can be estimated. According to Sect. 2.4.4, the matrix element  $\langle\psi|\rho|\phi\rangle$  can be directly sampled from the homodyne experimental values. In fact, for bounded  $\langle\psi|K(X_\varphi - x)|\phi\rangle$ , the central limit theorem guarantees that

$$\begin{aligned} \langle\psi|\rho|\phi\rangle &= \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} dx \ p(x, \varphi) \ \langle\psi|K(X_\varphi - x)|\phi\rangle \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^N \langle\psi|K(x_{\varphi_n} - x_n)|\phi\rangle, \end{aligned} \quad (2.93)$$

where  $x_n$  is the homodyne outcome measured at phase  $\varphi_n$  and distributed with probability  $p(x, \varphi)$ . Systematic errors are eliminated by choosing randomly each phase  $\varphi_n$  at which homodyne measurement is performed. As shown in Sect. 2.4.4, for finite number of measurements  $N$ , the estimate (2.93) of the integral is Gaussian distributed around the true value  $\langle\psi|\rho|\phi\rangle$ , with statistical error decreasing as  $N^{-1/2}$ . Notice that the measurability of the density operator matrix element depends only on the boundedness of the matrix element of the Kernel function, and that no adjustable parameters are needed in the procedure, which thus is unbiased.

The general procedure for noise deconvolution will be presented in Sect. 2.4.9. However, we give here the main result for the density matrix reconstruction. As shown in Sect. 2.3.2, the effect of the efficiency in homodyne detectors is a Gaussian convolution of the ideal probability  $p(x, \varphi)$ , as

$$p_\eta(x, \varphi) = \sqrt{\frac{2\eta}{\pi(1-\eta)}} \int_{-\infty}^{+\infty} dx' e^{-\frac{2\eta}{1-\eta}(x-x')^2} p(x', \varphi). \quad (2.94)$$

The tomographic reconstruction procedure still holds upon replacing  $p(x, \varphi)$  with  $p_\eta(x, \varphi)$ , so that

$$\rho = \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} dx \ p_\eta(x, \varphi) \ K_\eta(X_\varphi - x), \quad (2.95)$$

where now the Kernel function is

$$K_\eta(x) = \frac{1}{2} \text{Re} \int_0^{+\infty} k \ dk e^{\frac{1-\eta}{8\eta}k^2 + ikx}. \quad (2.96)$$

In fact, by taking the Fourier transform of both members of (2.94), one can easily check that

$$\begin{aligned}\rho &= \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} dx p_\eta(x, \varphi) K_\eta(X_\varphi - x) \\ &= \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} dx p(x, \varphi) K(X_\varphi - x).\end{aligned}\quad (2.97)$$

Notice that the anti-Gaussian in (2.96) causes a much slower convergence of the Monte Carlo integral (2.95): the statistical fluctuation will increase exponentially for decreasing detector efficiency  $\eta$ . In order to achieve good reconstructions with non-ideal detectors, then one has to collect a larger number of data.

It is clear from (2.93) that the measurability of the density matrix depends on the chosen representation and on the quantum efficiency of the detectors. For example, for the reconstruction of the density matrix in the Fock basis the Kernel functions are given by

$$\begin{aligned}\mathcal{R}_\eta[|n\rangle\langle n+d|](x, \varphi) &= \int_{-\infty}^{+\infty} \frac{dk |k|}{4} e^{\frac{1-\eta}{8\eta} k^2 - ikx} \langle n+d | e^{ikX_\varphi} | n \rangle \quad (2.98) \\ &= e^{id(\varphi + \frac{\pi}{2})} \sqrt{\frac{n!}{(n+d)!}} \int_{-\infty}^{+\infty} dk |k| e^{\frac{1-2\eta}{2\eta} k^2 - i2kx} k^d L_n^d(k^2),\end{aligned}$$

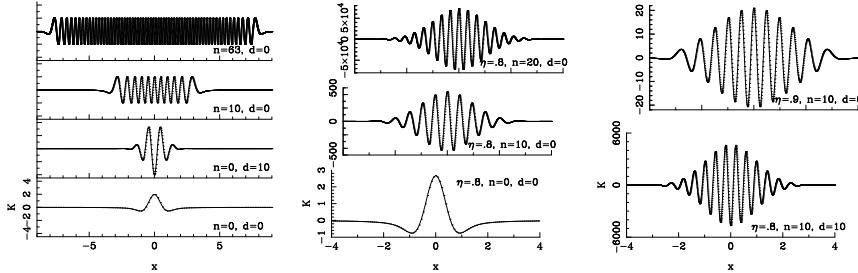
where  $L_n^d(x)$  denotes the generalized Laguerre polynomials. Notice that the Kernel function is bounded only for  $\eta > 1/2$ , while below this threshold the method would give unbounded statistical errors. However, this bound is well below the values that are reasonably achieved in the lab, where actual homodyne detectors have efficiency ranging between 70% and 90% [11, 68]. The kernel functions for the matrix elements  $\mathcal{R}[|n\rangle\langle n+d|](x, \varphi)$  satisfy the following relation [69, 70, 78]

$$\mathcal{R}[|n\rangle\langle n+d|](x, \varphi) = e^{id\varphi} [4x u_n(x) v_{n+d}(x)] \quad (2.99)$$

$$-2\sqrt{n+1} u_{n+1}(x) v_{n+d}(x) - 2\sqrt{n+d+1} u_n(x) v_{n+d+1}(x), \quad (2.100)$$

which can be effectively used to compute the kernel numerically. The functions  $u_j(x)$  and  $v_j(x)$  in Eq.(2.100) are the normalizable and unnormalizable eigenfunctions of the harmonic oscillator with eigenvalue  $j$ , respectively. The noise from quantum efficiency can be unbiased via the inversion of the Bernoulli convolution, which holds for  $\eta > 1/2$  [71]. In Fig. 2.3 the  $x$ -dependent part of the Kernel function is reported for different values of  $n$ ,  $d$  and the quantum efficiency  $\eta$ .

Notice that (2.89) cannot be used for unbounded operators. However, the Kernel functions can be derived also for some unbounded operators, as shown in [16]. In Table 2.1 we report the estimator  $\mathcal{R}_\eta[O](x, \varphi)$  for some operators



**Fig. 2.3.** The  $x$ -dependent part of the Kernel function  $\mathcal{R}[|n\rangle\langle n+d|](x, \varphi)$  for different values of  $n, d$   $\eta < 1$ . Notice the very different ranges.

$O$ . The operator  $\hat{W}_s$  gives the generalized Wigner function  $W_s(\alpha, \alpha^*)$  for ordering parameter  $s$  through the relation in Eq. (2.11). From the expression of  $\mathcal{R}_\eta[\hat{W}_s](x, \varphi)$  it follows that by homodyning with quantum efficiency  $\eta$  one can measure the generalized Wigner function only for  $s < 1 - \eta^{-1}$ : in particular the usual Wigner function for  $s = 0$  cannot be measured for any quantum efficiency.

$O$	$\mathcal{R}_\eta[O](x, \varphi)$
$a^\dagger^n a^m$	$e^{i(m-n)\varphi} \frac{H_{n+m}(\sqrt{2\eta}x)}{\sqrt{(2\eta)^{n+m}} \binom{n+m}{n}}$
$a$	$2e^{i\varphi} x$
$a^2$	$e^{2i\varphi} (4x^2 - 1/\eta)$
$a^\dagger a$	$2x^2 - \frac{1}{2\eta}$
$(a^\dagger a)^2$	$\frac{8}{3}x^4 - \frac{4-2\eta}{\eta}x^2 + \frac{1-\eta}{2\eta^2}$
$\hat{W}_s = \frac{2}{\pi(1-s)} \left( \frac{s+1}{s-1} \right)^{a^\dagger a}$	$\int_0^\infty dt \frac{2e^{-t}}{\pi(1-s) - \frac{1}{\eta}} \cos \left( 2\sqrt{\frac{2t}{(1-s) - \frac{1}{\eta}}} x \right)$
$ n\rangle\langle n+d $	$\mathcal{R}_\eta[ n\rangle\langle n+d ](x, \varphi)$ in Eq. (2.98)

**Table 2.1.** Estimator  $\mathcal{R}_\eta[O](x, \varphi)$  for some operators  $O$  (From Ref. [16]).

#### 2.4.6 Some generalizations

Using condition (2.78) one can see that the Glauber formula can be generalized to

$$O = \int_{\mathbb{C}} \frac{d^2\alpha}{\pi} \text{Tr} [OF_1D(\alpha)F_2] F_2^{-1}D^\dagger(\alpha)F_1^{-1}, \quad (2.101)$$

where  $F_1$  and  $F_2$  are two generic invertible operators. By choosing  $F_1^\dagger = F_2 = S(\zeta)$ , where  $S(\zeta)$  is the squeezing operator

$$S(\zeta) = \exp \left[ \frac{1}{2} \left( \zeta^2 a^{\dagger 2} - \zeta^{*2} a^2 \right) \right], \quad \zeta \in \mathbb{C}, \quad (2.102)$$

we obtain the tomographic resolution

$$\langle O \rangle = \int_0^\pi \frac{d\varphi}{\pi} \int_{-\infty}^{+\infty} dx p_\zeta(x, \varphi) \text{Tr}[OK(X_{\varphi\zeta} - x)], \quad (2.103)$$

in terms of the probability distribution of the generalized squeezed quadrature operators

$$X_{\varphi\zeta} = S^\dagger(\zeta) X_\varphi S(\zeta) = \frac{1}{2} [(\mu e^{i\varphi} + \nu e^{-i\varphi}) a^\dagger + (\mu e^{-i\varphi} + \nu^* e^{i\varphi}) a] \quad (2.104)$$

with  $\mu = \cosh |\zeta|$  and  $\nu = \sinh |\zeta| \exp(2i \arg[\zeta])$ . Such an estimation technique has been investigated in detail in [72].

A different estimation technique can be obtained by choosing in (2.101)  $F_1 = I$ , the identity operator, and  $F_2 = (-)^{a^\dagger a}$ , the parity operator. In this case one gets

$$O = \int_{\mathbb{C}} \frac{d^2\alpha}{\pi} \text{Tr} \left[ OD^\dagger(\alpha) (-)^{a^\dagger a} \right] (-)^{a^\dagger a} D(\alpha). \quad (2.105)$$

Changing variable to  $\alpha = 2\beta$  and using the relation

$$(-)^{a^\dagger a} D(2\beta) = D^\dagger(\beta) (-)^{a^\dagger a} D(\beta) \quad (2.106)$$

it follows

$$\langle O \rangle = 4 \int_{\mathbb{C}} \frac{d^2\beta}{\pi} \text{Tr} \left[ OD^\dagger(\beta) (-)^{a^\dagger a} D(\beta) \right] \text{Tr} \left[ D(\beta) \rho D^\dagger(\beta) (-)^{a^\dagger a} \right] \quad (2.107)$$

Hence, it is possible to estimate  $\langle O \rangle$  by repeated measurement of the parity operator on displaced versions of the state under investigation. An approximated implementation of this technique for a single mode radiation field has been suggested in [73, 74] through the measurement of the photon number probability on states displaced by a beam splitter. A similar scheme has been used for the experimental determination of the motional quantum state of a trapped atom [15]. In comparison with the approximated methods, (2.107) allows to directly obtain the Kernel  $\mathcal{R}[O](\alpha)$  for any operator  $O$  for which the trace exists. For instance, the reconstruction of the density matrix in the Fock representation is obtained by averaging the Kernel

$$\begin{aligned} \mathcal{R}[|n\rangle\langle n+d|](\alpha) &= 4 \langle n+d | D^\dagger(\alpha) (-)^{a^\dagger a} D(\alpha) | n \rangle \\ &= 4 (-)^{n+d} \sqrt{\frac{n!}{(n+d)!}} (2\alpha)^d e^{-2|\alpha|^2} L_n^d(4|\alpha|^2), \end{aligned} \quad (2.108)$$

without the need of artificial cut-off in the Fock space [15].

### 2.4.7 Quantum estimation for spin systems

The so-called spin tomography [20, 30, 66, 67] allows one to reconstruct the quantum state of a spin system from measurements of the spin in different directions, *i.e.* the quorum is the set of operators  $\mathbf{S} \cdot \mathbf{n}$ , where  $\mathbf{S}$  is the spin operator and  $\mathbf{n} \equiv (\cos \varphi \sin \vartheta, \sin \varphi \sin \vartheta, \cos \vartheta)$  is a unit vector. Various different quorums may be constructed by exploiting different sets of directions.

The easiest choice is to consider all possible directions. The procedure to derive the tomographic formulas for this quorum is analogous to the one employed in Sect. 2.4.5 for homodyne tomography. The reconstruction formula for spin tomography for the estimation of an arbitrary operator  $O$  writes

$$\langle O \rangle = \sum_{m=-s}^s \int_{\Omega} \frac{d\mathbf{n}}{4\pi} p(m, \mathbf{n}) \mathcal{R}[O](m, \mathbf{n}), \quad (2.109)$$

where  $p(m, \mathbf{n})$  is the probability of obtaining the eigenvalue  $m$  when measuring the spin along direction  $\mathbf{n}$ ,  $\mathcal{R}[O](m, \mathbf{n})$  is the tomographic Kernel for the operator  $O$ , and  $\Omega$  is the unit sphere. In this case the operators  $C(\lambda)$  of (2.73) are given by the set of projectors over the eigenstates  $|m, \mathbf{n}\rangle$  of the operators  $\mathbf{S} \cdot \mathbf{n}$ . Notice that it is a set of irreducible operators in the system Hilbert space  $\mathcal{H}$ . In order to find the dual basis  $B$ , one must consider the unitary operators obtained by exponentiating the quorum, *i.e.*  $D(\psi, \mathbf{n}) = \exp(i\psi \mathbf{S} \cdot \mathbf{n})$ , which satisfy the bi-orthogonality condition (2.76). In fact,  $D(\psi, \mathbf{n})$  constitutes a unitary irreducible representation of the group  $SU(2)$ , and the bi-orthogonality condition is just the orthogonality relations between the matrix elements of the group representation [75], *i.e.*

$$\int_R dg D_{jr}(g) D_{tk}^\dagger(g) = \frac{V}{d} \delta_{jk} \delta_{tr}, \quad (2.110)$$

where  $D$  is a unitary irreducible representation of dimension  $d$ ,  $dg$  is the group Haar invariant measure, and  $V = \int_R dg$ . For  $SU(2)$ , with the  $2s+1$  dimension unitary irreducible representation  $D(\psi, \mathbf{n})$ , Haar's invariant measure is  $\sin^2 \frac{\psi}{2} \sin \vartheta d\vartheta d\varphi d\psi$ , and  $\frac{V}{d} = \frac{4\pi^2}{2s+1}$ . The bi-orthogonality condition writes

$$\frac{2s+1}{4\pi^2} \int_{\Omega} d\mathbf{n} \int_0^{2\pi} d\psi \sin^2 \frac{\psi}{2} \langle j | e^{i\psi \mathbf{n} \cdot \mathbf{S}} | r \rangle \langle t | e^{-i\psi \mathbf{n} \cdot \mathbf{S}} | k \rangle = \delta_{jk} \delta_{tr}, \quad (2.111)$$

and hence the spin tomography identity is given by

$$O = \frac{2s+1}{4\pi^2} \int_{\Omega} d\mathbf{n} \int_0^{2\pi} d\psi \sin^2 \frac{\psi}{2} \text{Tr} [OD^\dagger(\psi, \mathbf{n})] D(\psi, \mathbf{n}). \quad (2.112)$$

Notice the analogy between (2.112) and Glauber's formula (2.89). In fact, both homodyne and spin tomography can be derived in the domain of Group

Tomography [20], where the underlying group structure is given by the Weyl-Heisenberg group and the SU(2) group, respectively. Formula (2.109) is obtained from (2.112) through the expectation value calculated on the eigenstates of  $\mathbf{S} \cdot \mathbf{n}$ . Thus, the explicit form of the tomographic Kernel is obtained as

$$\mathcal{R}[O](m, \mathbf{n}) = \frac{2s+1}{\pi} \int_0^{2\pi} d\psi \sin^2 \frac{\psi}{2} \text{Tr} [O e^{-i\psi \mathbf{S} \cdot \mathbf{n}}] e^{i\psi m}. \quad (2.113)$$

As already noticed, there are other possible quorums for spin tomography. For example, for spin  $s = \frac{1}{2}$  systems, a self-dual basis for the operator space is given by the identity and the Pauli matrices. In fact, from the property  $\sigma_\alpha \sigma_{\alpha'} = \delta_{\alpha\alpha'} I$  ( $\alpha, \alpha' = x, y, z$ ), both the bi-orthogonality relation (2.76) and the trace condition (2.78) follow. In this case the reconstruction formula writes

$$\langle O \rangle = \frac{1}{2} \text{Tr} [O] + \frac{1}{2} \sum_{\alpha=x,y,z} \sum_{m=\pm\frac{1}{2}} m p(m, \mathbf{n}_\alpha) \text{Tr} [O \sigma_\alpha]. \quad (2.114)$$

In the case of generic  $s$  spin system, Weigert has also shown [66] that by choosing  $(2s+1)^2$  arbitrary directions for  $\mathbf{n}$ , it is possible to obtain (in almost all cases) a quorum of projectors  $|s, \mathbf{n}_j\rangle \langle s, \mathbf{n}_j|$  ( $j = 1, \dots, (2s+1)^2$ ), where  $|s, \mathbf{n}_j\rangle$  is the eigenstate pertaining to the maximum eigenvalue  $s$  of  $\mathbf{S} \cdot \mathbf{n}_j$ .

#### 2.4.8 Quantum estimation for a free particle

The state of a moving packet can be inferred from position measurement at different times [76]. Assuming a particle with unit mass and using normalized unit  $\hbar/2 = 1$ , the free Hamiltonian is given by the square of momentum operator  $H_F = p^2$ . In terms of the eigenvectors  $|x\rangle$  of the position operator and of the selfadjoint operator

$$R(x, \tau) = e^{-ip^2\tau} |x\rangle \langle x| e^{ip^2\tau}, \quad (2.115)$$

the probability density of the position of the free particle at time  $\tau$  is obtained as  $p(x, \tau) = \text{Tr}[\rho R(x, \tau)]$ . The operators  $R(x, \tau)$  provide a self-dual basis, and an arbitrary particle state can be written as

$$\rho = \int_{\mathbb{R}} \int_{\mathbb{R}} dx d\tau p(x, \tau) R(x, \tau). \quad (2.116)$$

#### 2.4.9 Noise deconvolution and adaptive tomography

In the following we will briefly review 1) the noise deconvolution scheme [20, 77], that allows one to eliminate the experimental noise that arises from imperfect detection and lossy devices; 2) the adaptive tomography technique [22] that allows one to tune the unbiased tomographic Kernels to the experimental data one obtains, in order to reduce the statistical noise.

### Noise deconvolution

Essentially it is possible to eliminate detection noise when it is possible to invert the noise map. A noise process is described by a trace preserving completely positive map  $\Gamma$  acting on the Hilbert space of operators. The noise can be deconvolved at the data analysis if

- the inverse of  $\Gamma$  exists, namely  $\Gamma^{-1} : \mathcal{L}(\mathcal{H}) \rightarrow \mathcal{L}(\mathcal{H})$ , with  $\Gamma^{-1}[\Gamma[O]] = O$ , for  $\forall O \in \mathcal{L}(\mathcal{H})$ .
- the estimator  $\mathcal{E}_\lambda[O](Q_\lambda) = \text{Tr}[O B^\dagger(\lambda)] C(\lambda)$  is in the domain of  $\Gamma^{-1}$
- the map  $\Gamma^{-1}[\mathcal{E}_\lambda[O](Q_\lambda)]$  is a function of  $Q_\lambda$ .

If the above conditions are met, we can recover the “ideal” expectation value  $\langle O \rangle$  that we would get without noise. This is achieved by replacing  $\mathcal{E}_\lambda[O](Q_\lambda)$  with  $\Gamma^{-1}[\mathcal{E}_\lambda[O](Q_\lambda)]$ , and evaluating the ensemble average with the state  $\Gamma^\tau(\rho)$ , namely the state affected by the noise ( $\Gamma^\tau$  represents the dual map, that provides the evolution in the Schroedinger picture). Hence, one has

$$\langle O \rangle = \int_A d\lambda \text{Tr}[\Gamma^{-1}[\mathcal{E}_\lambda[O](Q_\lambda)] \Gamma^\tau(\rho)]. \quad (2.117)$$

Consider for example the noise arising from non-unity quantum efficiency  $\eta$  of homodyne detectors. Recall that the ideal probability density is replaced by a Gaussian convolution with rms  $\Delta_\eta^2 = (1 - \eta)/(4\eta)$ . Then, the map  $\Gamma_\eta$  acts on the quorum as follows

$$\begin{aligned} \Gamma_\eta[e^{ikX_\varphi}] &= \int_{-\infty}^{+\infty} dx e^{ikx} \Gamma_\eta[|x\rangle\langle x|] \\ &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dx' e^{ikx} e^{-\frac{(x-x')^2}{2\Delta^2}} [|x'\rangle\langle x'|] = e^{-\frac{1}{2}\Delta^2 k^2} e^{ikX_\varphi}. \end{aligned} \quad (2.118)$$

Of course one has

$$\Gamma_\eta^{-1}[e^{ikX_\varphi}] = e^{\frac{1}{2}\Delta^2 k^2} e^{ikX_\varphi}. \quad (2.119)$$

In terms of the Fourier transform of the Kernel function

$$\tilde{\mathcal{R}}[O](y, \varphi) = \int_{-\infty}^{+\infty} \frac{dx}{2\pi} e^{ixy} \mathcal{R}[O](x, \varphi), \quad (2.120)$$

one has

$$\tilde{\mathcal{R}}_\eta[O](y, \varphi) = e^{\frac{1}{2}\Delta^2 y^2} \tilde{\mathcal{R}}[O](y, \varphi). \quad (2.121)$$

We have implicitly applied the above result in Sect. 2.4.5, where the effect of non-unity quantum efficiency for reconstructing the density matrix elements was discussed. The use of the modified Kernel function in (2.96) and the origin of the bound  $\eta > 1/2$  are indeed due to the invertibility domain of  $\Gamma_\eta$ .

Another simple example of noise deconvolution can be given for a spin 1/2 system. Consider the map that describes the “depolarizing channel”

$$\Gamma_p[O] = (1-p)O + \frac{p}{2} \text{Tr}[O] I, \quad 0 \leq p \leq 1. \quad (2.122)$$

This map can be inverted for  $p \neq 1$  as follows

$$\Gamma_p^{-1}[O] = \frac{1}{1-p} \left( O - \frac{p}{2} \text{Tr}[O] I \right). \quad (2.123)$$

Then (2.114) can be replaced with

$$\langle O \rangle = \frac{1}{2} \text{Tr}[O] + \frac{1}{2(1-p)} \sum_{m=\pm\frac{1}{2}} \sum_{\alpha=x,y,z} m p_p(m, \mathbf{n}_\alpha) \text{Tr}[O \sigma_\alpha], \quad (2.124)$$

where now  $p_p(m, \mathbf{n}_\alpha)$  represents the probability of outcome  $m$  when measuring  $\sigma_\alpha$  on the noisy state  $\Gamma_p[\rho]$ .

### Adaptive tomography

The idea of adaptive tomography is that the tomographic null estimators of (2.85) can be used to reduce the statistical noise arising from the fact that in a real experiment the data sample is always finite. The addition of a null estimator in the ideal case of infinite statistics does not change the measured quantities, since its mean value is zero. In the realistic case of finite statistics, the mean values are statistical variables. As shown in Sect. 2.4.4, by dividing data into statistical blocks and evaluating the tomographic averages, one finds mean values that are Gaussian distributed. Thus, one can look for a procedure to reduce the variance of such distribution. In fact, consider the class of equivalent estimators for  $O$

$$\mathcal{E}'_\lambda[O](Q_\lambda) = \mathcal{E}_\lambda[O](Q_\lambda) + \sum_{i=1}^M \nu_i \mathcal{N}_i(Q_\lambda). \quad (2.125)$$

Each estimator in the class  $\mathcal{E}'$  is identified by the coefficient vector  $\nu$ . The variance of the tomographic averages can be evaluated as

$$\overline{\Delta^2 \mathcal{E}'[O]} = \overline{\Delta^2 \mathcal{E}[O]} + 2 \sum_{i=1}^M \nu_i \overline{\mathcal{N}_i \mathcal{E}[O]} + \sum_{i,j=1}^M \nu_i \nu_j \overline{\mathcal{N}_i \mathcal{N}_j}, \quad (2.126)$$

where  $\overline{F} \equiv \langle \int_A d\lambda F(Q_\lambda) \rangle$ , and

$$\overline{\Delta^2 \mathcal{E}[O]} = \overline{\mathcal{E}^2[O]} - \overline{\mathcal{E}[O]}^2. \quad (2.127)$$

Minimizing  $\overline{\Delta^2 \mathcal{E}'[O]}$  with respect to the coefficients  $\nu_i$ , one obtains the equation

$$\sum_{j=1}^M \nu_j \overline{\mathcal{N}_i \mathcal{N}_j} = -\overline{\mathcal{E}[O] \mathcal{N}_i}, \quad (2.128)$$

which can be solved starting from the estimated mean values, with the vector  $\boldsymbol{\nu}$  as unknown. Notice that the obtained vector  $\boldsymbol{\nu}$  will depend on the experimental data, and has to be calculated with the above procedure for any new set of data.

In summary, the adaptive tomographic algorithm consists in the following steps:

- Find the null estimators  $\mathcal{N}_i(Q_\lambda)$  ( $i = 1, \dots, M$ ) for the quorum which is being used in the experiment.
- Execute the experiment and collect the input data.
- Calculate, using the obtained data, the mean values  $\overline{\mathcal{N}_i \mathcal{N}_j}$  and  $\overline{\mathcal{E}[O] \mathcal{N}_i}$ , and solve the linear system (2.128), to obtain  $\boldsymbol{\nu}$ .
- Use the vector  $\boldsymbol{\nu}$  obtained in the previous step to build the ‘optimized estimator’  $\mathcal{E}'[O](Q_\lambda) = \mathcal{E}[O](Q_\lambda) + \sum_i \nu_i \mathcal{N}_i(Q_\lambda)$ . Using the data collected in the first step, the mean value  $\langle O \rangle$  is now evaluated as

$$\langle O \rangle = \int_A d\lambda \langle \mathcal{E}'_\lambda [O](Q_\lambda) \rangle, \quad (2.129)$$

where the optimized estimator has been used.

- For each new set of data the whole procedure must be repeated, as  $\boldsymbol{\nu}$  is dependent on the data.

Notice that also the mean values are changed in the adaptive tomographic process: null estimators do not change mean values only in the limiting case of infinite statistics. In fact, the mean values are changed in such a way as to reduce the dispersion of the data. Examples of simulations of the adaptive technique that efficiently reduce statistical noise of homodyne tomographic reconstructions can be found in [22]. In homodyne tomography null estimators are obtained as linear combinations of the following functions

$$\mathcal{N}_{k,n}(X_\varphi) = X_\varphi^k e^{\pm i(k+2n)\varphi}, \quad k, n \geq 0. \quad (2.130)$$

One can easily check that such functions have zero average over  $\varphi$ , independently on  $\rho$ . Hence, for every operator  $O$  one actually has an equivalence class of infinitely many unbiased estimators, which differ by a linear combination of functions  $\mathcal{N}_{k,n}(X_\varphi)$ . It is then possible to minimize the rms error in the equivalence class by the least-squares method, obtaining in this way an optimal estimator that is *adapted* to the particular set of experimental data.

## 2.5 Multimode homodyne tomography

The generalization of homodyne tomography from a single-mode to a multimode field is straightforward, the estimator of simple operator tensors

$O = O_1 \otimes O_2 \otimes \dots \otimes O_n$  being just the product of the estimators of each single-mode operator  $O_1, O_2, \dots, O_n$ . By linearity, one then obtains also the estimator for arbitrary multimode operators. Such a simple generalization, however, requires a separate homodyne detector for each mode, which is unfeasible when the modes of the field are not spatio-temporally separated. This is the case, for example of pulsed fields, for which a general multimode tomographic method is especially needed, also due to the problem of mode matching between the local oscillator and the detected fields (determined by their relative spatio-temporal overlap) [83], which produces a dramatic reduction of the overall quantum efficiency.

In this Section we review the general method of [17] for homodyning observables of a multimode electromagnetic field using a *single* local oscillator (LO), providing the rule to evaluate the estimator of an arbitrary multimode operator. The expectation value of the operator can then be obtained by averaging the estimator over the homodyne outcomes that are collected using a single LO whose mode randomly scans all possible linear combinations of incident modes. We will then specifically consider some observables for a two-mode field in a state corresponding to a twin-beam produced by parametric downconversion, and prove the reliability of the method on the basis of computer simulations.

Finally, we report some experimental results [84] obtained in the Prem Kumar's lab at Northwestern University. Such experiment actually represents the first measurement of the joint photon-number probability distribution of the twin-beam state.

### 2.5.1 The general method

The Hilbert-Schmidt operator expansion in (2.89) can be generalized to any number of modes as follows

$$\begin{aligned} O &= \int_{\mathbb{C}} \frac{d^2 z_0}{\pi} \int_{\mathbb{C}} \frac{d^2 z_1}{\pi} \dots \int_{\mathbb{C}} \frac{d^2 z_M}{\pi} \text{Tr} \left\{ O \exp \left[ \sum_{l=0}^M (-z_l a_l^\dagger + z_l^* a_l) \right] \right\} \\ &\quad \times \exp \left[ \sum_{l=0}^M (z_l a_l^\dagger - z_l^* a_l) \right], \end{aligned} \quad (2.131)$$

where  $a_l$  and  $a_l^\dagger$ , with  $l = 0, \dots, M$  and  $[a_l, a_{l'}^\dagger] = \delta_{ll'}$ , are the annihilation and creation operators of  $M + 1$  independent modes, and  $O$  now denotes an operator over all modes. Using the following hyper-spherical parameterization for  $z_l \in \mathbb{C}$

$$\begin{aligned}
z_0 &= \frac{i}{2} k u_0(\boldsymbol{\theta}) e^{i\psi_0} \doteq \frac{i}{2} k e^{i\psi_0} \cos \theta_1 , \\
z_1 &= \frac{i}{2} k u_1(\boldsymbol{\theta}) e^{i\psi_1} \doteq \frac{i}{2} k e^{i\psi_1} \sin \theta_1 \cos \theta_2 , \\
z_2 &= \frac{i}{2} k u_2(\boldsymbol{\theta}) e^{i\psi_2} \doteq \frac{i}{2} k e^{i\psi_2} \sin \theta_1 \sin \theta_2 \cos \theta_3 , \\
&\dots \\
z_{M-1} &= \frac{i}{2} k u_{M-1}(\boldsymbol{\theta}) e^{i\psi_{M-1}} \doteq \frac{i}{2} k e^{i\psi_{M-1}} \sin \theta_1 \sin \theta_2 \dots \sin \theta_{M-1} \cos \theta_M , \\
z_M &= \frac{i}{2} k u_M(\boldsymbol{\theta}) e^{i\psi_M} \doteq \frac{i}{2} k e^{i\psi_M} \sin \theta_1 \sin \theta_2 \dots \sin \theta_{M-1} \sin \theta_M ,
\end{aligned} \tag{2.132}$$

where  $k \in [0, \infty)$ ;  $\psi_l \in [0, 2\pi]$  for  $l = 0, 1, \dots, M$ ; and  $\theta_l \in [0, \pi/2]$  for  $l = 1, 2, \dots, M$ , (2.131) can be rewritten as follows:

$$\begin{aligned}
O &= \int d\mu[\boldsymbol{\psi}] \int d\mu[\boldsymbol{\theta}] \int_0^{+\infty} dk \left(\frac{k}{2}\right)^{2M+1} \frac{1}{M!} \\
&\times \text{Tr}[O e^{-ikX(\boldsymbol{\theta}, \boldsymbol{\psi})}] e^{ikX(\boldsymbol{\theta}, \boldsymbol{\psi})} .
\end{aligned} \tag{2.133}$$

Here we have used the notation

$$\int d\mu[\boldsymbol{\psi}] \doteq \prod_{l=0}^M \int_0^{2\pi} \frac{d\psi_l}{2\pi} , \tag{2.134}$$

$$\int d\mu[\boldsymbol{\theta}] \doteq 2^M M! \prod_{l=1}^M \int_0^{\pi/2} d\theta_l \sin^{2(M-l)+1} \theta_l \cos \theta_l , \tag{2.135}$$

$$X(\boldsymbol{\theta}, \boldsymbol{\psi}) = \frac{1}{2} [A^\dagger(\boldsymbol{\theta}, \boldsymbol{\psi}) + A(\boldsymbol{\theta}, \boldsymbol{\psi})] , \tag{2.136}$$

$$A(\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{l=0}^M e^{-i\psi_l} u_l(\boldsymbol{\theta}) a_l . \tag{2.137}$$

From the parameterization in (2.133), one has  $\sum_{l=0}^M u_l^2(\boldsymbol{\theta}) = 1$ , and hence  $[A(\boldsymbol{\theta}, \boldsymbol{\psi}), A^\dagger(\boldsymbol{\theta}, \boldsymbol{\psi})] = 1$ , namely  $A(\boldsymbol{\theta}, \boldsymbol{\psi})$  and  $A^\dagger(\boldsymbol{\theta}, \boldsymbol{\psi})$  themselves are annihilation and creation operators of a bosonic mode. By scanning all values of  $\theta_l \in [0, \pi/2]$  and  $\psi_l \in [0, 2\pi]$ , all possible linear combinations of modes  $a_l$  are obtained.

For the quadrature operator  $X(\boldsymbol{\theta}, \boldsymbol{\psi})$  in (2.136), one has the following identity for the moments generating function

$$\langle e^{ikX(\boldsymbol{\theta}, \boldsymbol{\psi})} \rangle = \exp \left( \frac{1-\eta}{8\eta} k^2 \right) \int_{-\infty}^{+\infty} dx e^{ikx} p_\eta(x; \boldsymbol{\theta}, \boldsymbol{\psi}) , \tag{2.138}$$

where  $p_\eta(x; \boldsymbol{\theta}, \boldsymbol{\psi})$  denotes the homodyne probability distribution of the quadrature  $X(\boldsymbol{\theta}, \boldsymbol{\psi})$  with quantum efficiency  $\eta$ . Generally,  $\eta$  can depend on the mode itself, i.e., it is a function  $\eta = \eta(\boldsymbol{\theta}, \boldsymbol{\psi})$  of the selected mode. In the

following, for simplicity, we assume  $\eta$  to be mode independent, however. By taking the ensemble average on each side of (2.133) and using (2.138) one has

$$\langle O \rangle = \int d\mu[\psi] \int d\mu[\theta] \int_{-\infty}^{+\infty} dx p_\eta(x; \theta, \psi) \mathcal{R}_\eta[O](x; \theta, \psi), \quad (2.139)$$

where the estimator  $\mathcal{R}_\eta[O](x; \theta, \psi)$  has the following expression

$$\begin{aligned} \mathcal{R}_\eta[O](x; \theta, \psi) &= \frac{\kappa^{M+1}}{M!} \int_0^{+\infty} dt e^{-(1-\frac{\kappa}{2})t+2i\sqrt{\kappa t}x} t^M \\ &\times \text{Tr}[O e^{-2i\sqrt{\kappa t}X(\theta, \psi)}], \end{aligned} \quad (2.140)$$

with  $\kappa = 2\eta/(2\eta - 1)$ . Equations (2.139) and (2.140) allow one to obtain the expectation value  $\langle O \rangle$  for any unknown state of the radiation field by averaging over the homodyne outcomes of the quadrature  $X(\theta, \psi)$  for  $\theta$  and  $\psi$  randomly distributed according to  $d\mu[\psi]$  and  $d\mu[\theta]$ . Such outcomes can be obtained by using a single LO that is prepared in the multimode coherent state  $\otimes_{l=0}^M |\gamma_l\rangle$  with  $\gamma_l = e^{i\psi_l} u_l(\theta) K/2$  and  $K \gg 1$ . In fact, in this case the rescaled zero-frequency photocurrent at the output of a balanced homodyne detector is given by

$$I = \frac{1}{K} \sum_{l=0}^M (\gamma_l^* a_l + \gamma_l a_l^\dagger), \quad (2.141)$$

which corresponds to the operator  $X(\theta, \psi)$ . In the limit of a strong LO ( $K \rightarrow \infty$ ), all moments of the current  $I$  correspond to the moments of  $X(\theta, \psi)$ , and the exact measurement of  $X(\theta, \psi)$  is then realized. Notice that for modes  $a_l$  with different frequencies, in the d.c. photocurrent in (2.141) each LO with amplitude  $\gamma_l$  selects the mode  $a_l$  at the same frequency (and polarization). For less-than-unity quantum efficiency, (2.138) holds.

Equation (2.140) can be applied to some observables of interest. In particular, one can estimate the matrix element  $\langle \{n_l\} | R | \{m_l\} \rangle$  of the multimode density operator  $R$ . This will be obtained by averaging the estimator

$$\begin{aligned} \mathcal{R}_\eta[|\{m_l\}\rangle\langle\{n_l\}|](x; \theta, \psi) &= e^{-i\sum_{l=0}^M (n_l - m_l)\psi_l} \frac{\kappa^{M+1}}{M!} \\ &\times \prod_{l=0}^M \left\{ [-i\sqrt{\kappa} u_l(\theta)]^{\mu_l - \nu_l} \sqrt{\frac{\nu_l!}{\mu_l!}} \right\} \\ &\times \int_0^{+\infty} dt e^{-t+2i\sqrt{\kappa t}x} t^{M+\sum_{l=0}^M (\mu_l - \nu_l)/2} \prod_{l=0}^M L_{\nu_l}^{\mu_l - \nu_l} [\kappa u_l^2(\theta)t], \end{aligned} \quad (2.142)$$

where  $\mu_l = \max(m_l, n_l)$ ,  $\nu_l = \min(m_l, n_l)$ , and  $L_n^\alpha(z)$  denotes the generalized Laguerre polynomial. For diagonal matrix elements, (2.142) simplifies to

$$\begin{aligned} \mathcal{R}_\eta[|\{n_l\}\rangle\langle\{n_l\}|](x; \boldsymbol{\theta}, \boldsymbol{\psi}) = \\ \frac{\kappa^{M+1}}{M!} \int_0^{+\infty} dt e^{-t+2i\sqrt{\kappa t}x} t^M \prod_{l=0}^M L_{n_l}[\kappa u_l^2(\boldsymbol{\theta})t] \end{aligned} \quad (2.143)$$

with  $L_n(z)$  denoting the customary Laguerre polynomial in  $z$ . Using the following identity [79]

$$\begin{aligned} L_n^{\alpha_0+\alpha_1+\dots+\alpha_M+M}(x_0+x_1+\dots+x_M) \\ = \sum_{i_0+i_1+\dots+i_M=n} L_{i_0}^{\alpha_0}(x_0)L_{i_1}^{\alpha_1}(x_1)\dots L_{i_M}^{\alpha_M}(x_M), \end{aligned} \quad (2.144)$$

from (2.143) one can easily derive the estimator of the probability distribution of the total number of photons  $N = \sum_{l=0}^M a_l^\dagger a_l$

$$\mathcal{R}_\eta[|n\rangle\langle n|](x; \boldsymbol{\theta}, \boldsymbol{\psi}) = \frac{\kappa^{M+1}}{M!} \int_0^{+\infty} dt e^{-t+2i\sqrt{\kappa t}x} t^M L_n^M[\kappa t], \quad (2.145)$$

where  $|n\rangle$  denotes the eigenvector of  $N$  with eigenvalue  $n$ . Notice that the estimator in (2.143) does not depend on the phases  $\psi_l$ ; only the knowledge of the angles  $\theta_l$  is needed. For the estimator in (2.145), even the angles  $\theta_l$  can be unknown.

Now we specialize to the case of only two modes  $a$  and  $b$  (i.e.,  $M=1$  and  $\boldsymbol{\theta}$  is a scalar  $\theta$ ). The joint photon-number probability distribution is obtained by averaging

$$\begin{aligned} \mathcal{R}_\eta[|n, m\rangle\langle n, m|](x; \theta, \psi_0, \psi_1) = \\ \kappa^2 \int_0^{+\infty} dt e^{-t+2i\sqrt{\kappa t}x} t L_n(\kappa t \cos^2 \theta) L_m(\kappa t \sin^2 \theta). \end{aligned} \quad (2.146)$$

The estimator (2.145) of the probability distribution of the total number of photons can be written as

$$\mathcal{R}_\eta[|n\rangle\langle n|](x; \theta, \psi_0, \psi_1) = \kappa^2 \int_0^{+\infty} dt e^{-t+2i\sqrt{\kappa t}x} t L_n^1[\kappa t]. \quad (2.147)$$

For the total number of photons one can also derive the estimator of the moment generating function, using the generating function for the Laguerre polynomials [79]. One obtains

$$\mathcal{R}_\eta[z^{a^\dagger a + b^\dagger b}](x; \theta, \psi_0, \psi_1) = \frac{1}{(z + \frac{1-z}{\kappa})^2} \Phi\left(2, \frac{1}{2}; -\frac{1-z}{z + \frac{1-z}{\kappa}} x^2\right). \quad (2.148)$$

For the first two moments one obtains the simple expressions

$$\begin{aligned} \mathcal{R}_\eta[a^\dagger a + b^\dagger b](x; \theta, \psi_0, \psi_1) &= 4x^2 + \frac{2}{\kappa} - 2, \\ \mathcal{R}_\eta[(a^\dagger a + b^\dagger b)^2](x; \theta, \psi_0, \psi_1) &= 8x^4 + \left(\frac{24}{\gamma} - 20\right)x^2 + \frac{6}{\gamma^2} - \frac{10}{\gamma} + 4. \end{aligned} \quad (2.149)$$

It is worth noting that analogous estimators of the photon-number difference between the two modes are singular and one needs a cutoff procedure, similar to the one used in [85] for recovering the correlation between the modes by means of the customary two-mode tomography. In fact, in order to extract information pertaining to a single mode only one needs a delta-function at  $\theta = 0$  for mode  $a$ , or  $\theta = \pi/2$  for mode  $b$ , and, in this case, one could better use the standard one-mode tomography by setting the LO to the proper mode of interest.

Finally, we note that for two-mode tomography the estimators can be averaged by the integral

$$\begin{aligned} \langle O \rangle &= \int_0^{2\pi} \frac{d\psi_0}{2\pi} \int_0^{2\pi} \frac{d\psi_1}{2\pi} \int_{-1}^1 \frac{d(\cos 2\theta)}{2} \int_{-\infty}^{+\infty} dx p_\eta(x; \theta, \psi_0, \psi_1) \\ &\times \mathcal{R}_\eta[O](x; \theta, \psi_0, \psi_1) \end{aligned} \quad (2.150)$$

over the random parameters  $\cos(2\theta)$ ,  $\psi_0$ , and  $\psi_1$ . For example, in the case of two radiation modes having the same frequency but orthogonal polarizations,  $\theta$  represents a random rotation of the polarizations, whereas  $\psi_0$  and  $\psi_1$  denote the relative phases between the LO and the two modes, respectively.

### 2.5.2 Numerical results for two-mode fields

In this section we report some Monte-Carlo simulations from [17] to judge the experimental working conditions for performing the single-LO tomography on two-mode fields. We focus our attention on the twin-beam state, usually generated by spontaneous parametric downconversion, namely

$$|\Psi\rangle = S(\chi)|0\rangle_a|0\rangle_b = \sqrt{1 - |\xi|^2} \sum_{n=0}^{\infty} \xi^n |n\rangle_a |n\rangle_b, \quad (2.151)$$

where  $S(\chi) = \exp(\chi a^\dagger b^\dagger - \chi^* ab)$  and  $\xi = e^{i \arg \chi} \tanh |\chi|$ . The parameter  $\xi$  is related to the average number of photons per beam  $\bar{n} = |\xi|^2 / (1 - |\xi|^2)$ . For the simulations we need to derive the homodyne probability distribution  $p(x; \theta, \psi_0, \psi_1)$  which is given by

$$\begin{aligned} p(x; \theta, \psi_0, \psi_1) &= \text{Tr}[U^\dagger |x\rangle_{aa}\langle x| \otimes 1_b U |\Psi\rangle\langle\Psi|] \\ &= {}_a\langle 0|_b\langle 0| S^\dagger(\chi) U^\dagger [|x\rangle_{aa}\langle x| \otimes 1_b] U S(\chi) |0\rangle_a |0\rangle_b, \end{aligned} \quad (2.152)$$

where  $|x\rangle_a$  is the eigenvector of the quadrature  $x = \frac{1}{2}(a^\dagger + a)$  with eigenvalue  $x$  and  $U$  is the unitary operator achieving the mode transformation

$$U^\dagger \begin{pmatrix} a \\ b \end{pmatrix} U = \begin{pmatrix} e^{-i\psi_0} \cos \theta & e^{-i\psi_1} \sin \theta \\ -e^{i\psi_1} \sin \theta & e^{i\psi_0} \cos \theta \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}. \quad (2.153)$$

In the case of two radiation modes having the same frequency but orthogonal polarizations—the case of Type II phase-matched parametric amplifier—

(2.152) gives the theoretical probability of outcome  $x$  for the homodyne measurement at a polarization angle  $\theta$  with respect to the polarization of the  $a$  mode, and with  $\psi_0$  and  $\psi_1$  denoting the relative phases between the LO and the two modes, respectively. The probability in (2.152) is given by [17]

$$p(x; \theta, \psi_0, \psi_1) \frac{1}{\sqrt{2\pi\Delta^2(\theta, \psi_0, \psi_1)}} \exp\left(-\frac{x^2}{2\Delta^2(\theta, \psi_0, \psi_1)}\right), \quad (2.154)$$

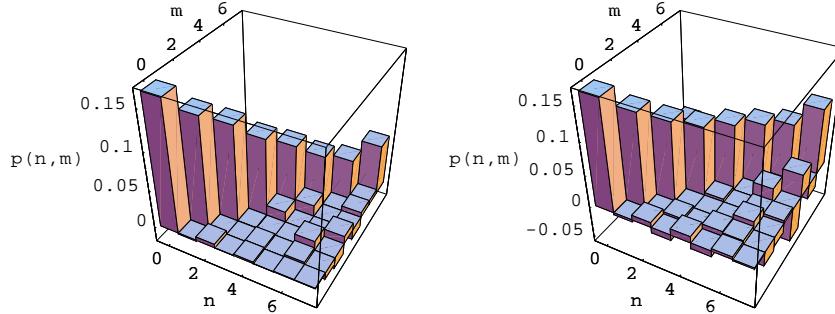
where the variance  $\Delta^2(\theta, \psi_0, \psi_1)$  reads

$$\Delta^2(\theta, \psi_0, \psi_1) = \frac{1 + |\xi|^2 + 2|\xi| \sin 2\theta \cos(\psi_0 + \psi_1 - \arg \xi)}{4(1 - |\xi|^2)}. \quad (2.155)$$

Taking into account the Gaussian convolution that results from less-than-unity quantum efficiency, the variance just increases as

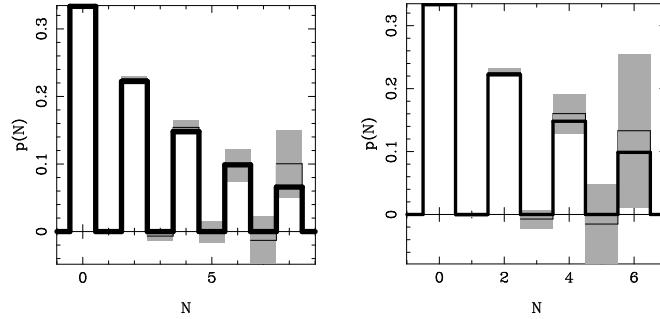
$$\Delta^2(\theta, \psi_0, \psi_1) \rightarrow \Delta_\eta^2(\theta, \psi_0, \psi_1) = \Delta^2(\theta, \psi_0, \psi_1) + \frac{1 - \eta}{4\eta}. \quad (2.156)$$

Notice that the probability distribution in (2.154) corresponds to a squeezed vacuum for  $\theta = \frac{\pi}{4}$  and  $\psi_0 + \psi_1 - \arg \xi = 0$  or  $\pi$ .



**Fig. 2.4.** Two-mode photon-number probability  $p(n, m)$  of the twin-beam state in (2.151) for average number of photons per beam  $\bar{n} = 5$  obtained by a Monte-Carlo simulation with the estimator in (2.146) and random parameters  $\cos 2\theta$ ,  $\psi_0$ , and  $\psi_1$ . On the left: quantum efficiency  $\eta = 1$  and  $10^6$  data samples were used in the reconstruction. On the right:  $\eta = 0.9$ , and  $5 \times 10^6$  data samples (From [17]).

We study the tomographic measurement of the joint photon-number probability distribution and the probability distribution for the total number of photons with use of the estimators in (2.146) and (2.147), respectively. The estimators have been numerically evaluated by applying the Gauss method for calculating the integral in (2.142), which results in a fast and sufficiently precise algorithm with use of just 150 evaluation points.

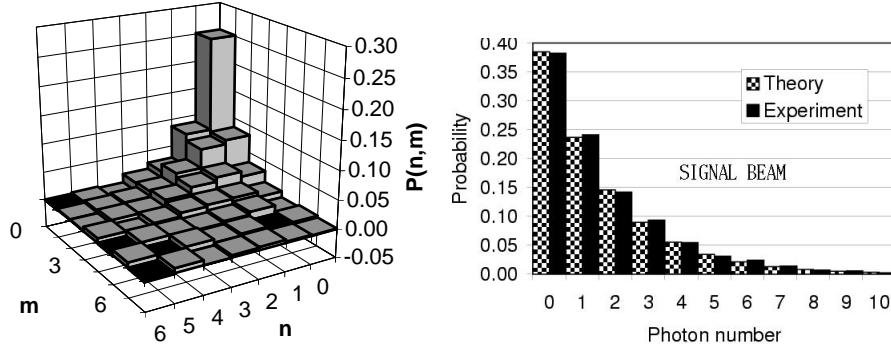


**Fig. 2.5.** Probability distribution for the total number of photons of the twin beams in (2.151) for average number of photons per beam  $\bar{n} = 2$  obtained using the estimator in (2.147). Samples of  $10^7$  data with quantum efficiency  $\eta = 0.9$  (on the left), and of  $2 \times 10^7$  data with  $\eta = 0.8$  (on the right). The theoretical probability (thick solid line) is superimposed onto the result of the Monte-Carlo experiment; the latter is shown by the thin solid line. Notice the dramatic increase of errors (in gray shade) versus  $N$  and for smaller  $\eta$  (From [17]).

In Fig. 2.4 a Monte-Carlo simulation of the joint photon-number probability distribution is reported. The simulated values compare very well with the theoretical ones. In [17] a careful analysis of the statistical errors has been done for various twin-beam states by constructing histograms of deviations of the results from different simulated experiments from the theoretical ones. In comparison to the customary two-LO tomography of [85], where for  $\eta = 1$  the statistical errors saturate for increasingly large  $n$  and  $m$ , here we have statistical errors that are slowly increasing versus  $n$  and  $m$ . This is due to the fact that the range of the estimators in (2.146) increases versus  $n$  and  $m$ . Overall we find that for any given quantum efficiency the statistical errors are generally slightly larger than those obtained with the two-LO method. The convenience of using a single LO then comes with its own price tag.

By using the estimator in (2.147) the probability distribution for the total number of photons  $N$  of the twin beams has been also constructed (Fig. 2.5). Notice the dramatic increase of error bars versus  $N$  and for smaller  $\eta$ .

The first experimental results of a measurement of the joint photon-number probability distribution for a two-mode quantum state created by a nondegenerate optical parametric amplifier has been presented in [84]. In this experiment, however, the twin beams are detected separately by two balanced-homodyne detectors. Some experimental results are reported in Fig. 2.6. As expected for parametric fluorescence, the experiment has shown a measured joint photon-number probability distribution that exhibited up to 1.9 dB of quantum correlation between the two modes, with thermal marginal distributions.



**Fig. 2.6.** Left: Measured joint photon-number probability distribution for the twin-beam state with average number of photons per beam  $\bar{n} = 1.5$  and 400000 samples. Right: marginal distribution for the signal beam for the same data. The theoretical distribution is also shown. Very similar results are obtained for the idler beam (From [84]).

## 2.6 Maximum-likelihood method in quantum estimation

Quantum estimation of states, observables and parameters is, from very basic principles, matter of statistical inference from a population sampling, and the most comprehensive quantum estimation procedure is quantum tomography. As we have shown in Sect. 3, the expectation value of an operator is obtained by averaging an estimator over the experimental data of a “quorum” of observables. The method is very general and efficient, however, in the averaging procedure, we have fluctuations which result in relatively large statistical errors.

Another relevant strategy, the maximum-likelihood (ML) method, can be used in quantum estimation. The ML strategy [86, 87] is an entirely different approach to quantum state measurement compared to the standard quantum-tomographic techniques. The ML procedure consists in finding the quantum state, or the value of the parameters, that are most likely to generate the observed data. This idea can be quantified and implemented using the concept of the likelihood functional.

A detailed account of ML methods in quantum estimation will be given in the Chapter by Z. Hradil *et al* in this book. Here we will review the applications of ML proposed in [23] and [31] to the quantum state reconstruction, with examples for both radiation and spin systems, and, finally, considering the ML estimation for the relevant class of Gaussian states in quantum optics.

As regards state estimation, the ML method estimates the quantum state as a whole. Such a procedure incorporates *a priori* knowledge about relations between elements of the density matrix. This guarantees positivity and normalization of matrix, with the result of a substantial reduction of statistical errors. Regarding the estimation of specific parameters, we notice that

in many cases the resulting estimators are efficient, unbiased and consistent, thus providing a statistically reliable determination.

As we will show, by using the ML method only small samples of data are required for a precise determination. However, ML may not always be the optimal solution of the tomographic problem. Besides being biased due to the Hilbert space truncation—even though the bias can be very small if, from other methods, we know where to truncate—it cannot be generalized to the estimation of any ensemble average, but just of a set of parameters from which the density matrix depends. In addition, for increasing number of parameters the method has exponential complexity.

### 2.6.1 Maximum likelihood principle

Here we briefly review the theory of the maximum-likelihood (ML) estimation of a single parameter. The generalization to several parameters, as for example the elements of the density matrix, is straightforward. The only point that should be carefully analyzed is the parameterization of the multidimensional quantity to be estimated. In the next section the specific case of the density matrix will be discussed.

Let  $p(x|\lambda)$  the probability density of a random variable  $x$ , conditioned to the value of the parameter  $\lambda$ . The form of  $p$  is known, but the true value of  $\lambda$  is unknown, and will be estimated from the result of a measurement of  $x$ . Let  $x_1, x_2, \dots, x_N$  be a random sample of size  $N$ . The joint probability density of the independent random variable  $x_1, x_2, \dots, x_N$  (the global probability of the sample) is given by

$$\mathcal{L}(x_1, x_2, \dots, x_N | \lambda) = \prod_{k=1}^N p(x_k | \lambda), \quad (2.157)$$

and is called the likelihood function of the given data sample (hereafter we will suppress the dependence of  $\mathcal{L}$  on the data). The maximum-likelihood estimator (MLE) of the parameter  $\lambda$  is defined as the quantity  $\lambda_{ml} \equiv \lambda_{ml}(\{x_k\})$  that maximizes  $\mathcal{L}(\lambda)$  for variations of  $\lambda$ , namely  $\lambda_{ml}$  is given by the solution of the equations

$$\frac{\partial \mathcal{L}(\lambda)}{\partial \lambda} = 0; \quad \frac{\partial^2 \mathcal{L}(\lambda)}{\partial \lambda^2} < 0. \quad (2.158)$$

The first equation is equivalent to  $\partial L / \partial \lambda = 0$  where

$$L(\lambda) = \log \mathcal{L}(\lambda) = \sum_{k=1}^N \log p(x_k | \lambda) \quad (2.159)$$

is the so-called log-likelihood function.

In order to obtain a measure for the confidence interval in the determination of  $\lambda_{ml}$  we consider the variance

$$\sigma_\lambda^2 = \int \left[ \prod_k dx_k p(x_k|\lambda) \right] [\lambda_{ml}(\{x_k\}) - \lambda]^2 . \quad (2.160)$$

In terms of the Fisher information

$$F = \int dx \left[ \frac{\partial p(x|\lambda)}{\partial \lambda} \right]^2 \frac{1}{p(x|\lambda)} , \quad (2.161)$$

it is easy to prove that

$$\sigma_\lambda^2 \geq \frac{1}{NF} , \quad (2.162)$$

where  $N$  is the number of measurements. The inequality in (2.162) is known as the Cramér-Rao bound [88] on the precision of the ML estimation. Notice that this bound holds for any functional form of the probability distribution  $p(x|\lambda)$ , provided that the Fisher information exists  $\forall \lambda$  and  $\partial_\lambda p(x|\lambda)$  exists  $\forall x$ . When an experiment has "good statistics" (i.e. for a large enough data sample) the Cramér-Rao bound is saturated.

### 2.6.2 ML quantum state estimation

In this section we review the method of the maximum likelihood estimation of the quantum state of [23], focusing attention to the cases of homodyne and spin tomography.

We consider an experiment consisting of  $N$  measurements performed on identically prepared copies of a given quantum system. Each measurement is described by a positive operator-valued measure (POVM). The outcome of the  $i$ th measurement corresponds to the realization of a specific element of the POVM used in the corresponding run, and we denote this element by  $\Pi_i$ . The likelihood is here a functional of the density matrix  $\mathcal{L}(\rho)$  and is given by the product

$$\mathcal{L}(\rho) = \prod_{i=1}^N \text{Tr}(\rho \Pi_i) , \quad (2.163)$$

which represents the probability of the observed data. The unknown element of the above expression, which we want to infer from data, is the density matrix describing the measured ensemble. The estimation strategy of the ML technique is to maximize the likelihood functional over the set of the density matrices. Several properties of the likelihood functional are easily found, if we restrict ourselves to finite dimensional Hilbert spaces. In this case, it can be easily proved that  $\mathcal{L}(\rho)$  is a concave function defined on a convex and closed set of density matrices. Therefore, its maximum is achieved either on a single isolated point, or on a convex subset of density matrices. In the latter case, the experimental data are insufficient to provide a unique estimate for the density matrix using the ML strategy. On the other hand, the existence of a

single maximum allows us to assign unambiguously the ML estimate for the density matrix.

The ML estimation of the quantum state, despite its elegant general formulation, results in a highly nontrivial constrained optimization problem, even if we resort to purely numerical means. The main difficulty lies in the appropriate parameterization of the set of all density matrices. The parameter space should be of the minimum dimension in order to preserve the maximum of the likelihood function as a single isolated point. Additionally, the expression of quantum expectation values in terms of this parameterization should enable fast evaluation of the likelihood function, as this step is performed many times in the course of numerical maximization.

For such purpose one introduces [23] a parameterization of the set of density matrices which provides an efficient algorithm for maximization of the likelihood function. We represent the density matrix in the form

$$\rho = T^\dagger T , \quad (2.164)$$

which automatically guarantees that  $\rho$  is positive and Hermitian. The remaining condition of unit trace  $\text{Tr}\rho = 1$  will be taken into account using the method of Lagrange multipliers. In order to achieve the minimal parameterization, we assume that  $T$  is a complex lower triangular matrix, with real elements on the diagonal. This form of  $T$  is motivated by the Cholesky decomposition known in numerical analysis [89] for arbitrary non negative Hermitian matrix. For an  $M$ -dimensional Hilbert space, the number of real parameters in the matrix  $T$  is  $M + 2M(M - 1)/2 = M^2$ , which equals the number of independent real parameters for a Hermitian matrix. This confirms that such parameterization is minimal, up to the unit trace condition.

In numerical calculations, it is convenient to replace the likelihood functional by its natural logarithm, which of course does not change the location of the maximum. Thus the log-likelihood function subjected to numerical maximization is given by

$$L(T) = \sum_{i=1}^N \ln \text{Tr}(T^\dagger T \Pi_i) - \lambda \text{Tr}(T^\dagger T) , \quad (2.165)$$

where  $\lambda$  is a Lagrange multiplier accounting for normalization of  $\rho$ . Writing  $\rho$  in terms of its eigenvectors  $|\psi_\mu\rangle$  as  $\rho = \sum_\mu y_\mu^2 |\psi_\mu\rangle\langle\psi_\mu|$ , with real  $y_\mu$ , the maximum likelihood condition  $\partial L/\partial y_\nu = 0$  reads

$$\lambda y_\nu = \sum_{i=1}^N [y_\nu \langle\psi_\nu|\Pi_i|\psi_\nu\rangle / \text{Tr}(\rho\Pi_i)] , \quad (2.166)$$

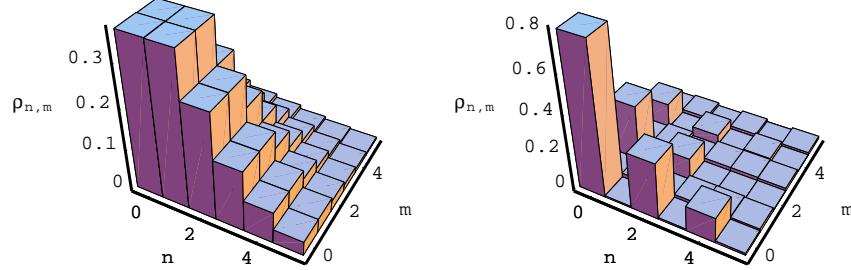
which, after multiplication by  $y_\nu$  and summation over  $\nu$ , yields  $\lambda = N$ . The Lagrange multiplier then equals the total number of measurements  $N$ .

This formulation of the maximization problem allows one to apply standard numerical procedures for searching the maximum over the  $M^2$  real parameters of the matrix  $T$ . The examples presented below use the downhill simplex method [90].

The first example is the ML estimation of a single-mode radiation field. The experimental apparatus used in this technique is the homodyne detector. According to Sect. 2.3.3 the homodyne measurement is described by the positive operator-valued measure

$$\mathcal{H}(x; \varphi) = \sqrt{\frac{2\eta}{\pi(1-\eta)}} \exp\left[-\frac{2\eta}{1-\eta}(X_\varphi - x)^2\right], \quad (2.167)$$

where  $\eta$  is the detector efficiency, and  $X_\varphi = (a^\dagger e^{i\varphi} + a e^{-i\varphi})/2$  is the quadrature operator at phase  $\varphi$ .



**Fig. 2.7.** Reconstruction of the density matrix of a single-mode radiation field by the ML method. The plot shows the matrix elements of a coherent state (left) with  $\langle a^\dagger a \rangle = 1$  photon, and for a squeezed vacuum (right) with  $\langle a^\dagger a \rangle = 0.5$  photon. A sample of 50000 simulated homodyne data for quantum efficiency  $\eta = 80\%$  has been used (From [23]).

After  $N$  measurements, we obtain a set of pairs  $(x_i; \varphi_i)$ , where  $i = 1, \dots, N$ . The log-likelihood functional is given by (2.165) with  $\Pi_i \equiv \mathcal{H}(x_i; \varphi_i)$ . Of course, for a light mode it is necessary to truncate the Hilbert space to a finite dimensional basis. We shall assume that the highest Fock state has  $M - 1$  photons, i.e. that the dimension of the truncated Hilbert space is  $M$ . For the expectation  $\text{Tr}[T^\dagger T \mathcal{H}(x; \varphi)]$  it is necessary to use an expression which is explicitly positive, in order to protect the algorithm against occurrence of small negative numerical arguments of the logarithm function. A simple derivation yields

$$\text{Tr}[T^\dagger T \mathcal{H}(x; \varphi)] = \sqrt{\eta} \sum_{k=0}^{M-1} \sum_{j=0}^k \left| \sum_{n=0}^{k-j} \langle k | T | n+j \rangle B_{n+j,n} \langle n | \sqrt{\eta} x \rangle e^{in\varphi} \right|^2 \quad (2.168)$$

where

$$B_{n+j,n} = \left[ \binom{n+j}{n} \eta^n (1-\eta)^j \right]^{1/2}, \quad (2.169)$$

and

$$\langle n | x \rangle = \left( \frac{2}{\pi} \right)^{1/4} \frac{1}{\sqrt{2^n n!}} H_n(\sqrt{2} x) \exp(-x^2) \quad (2.170)$$

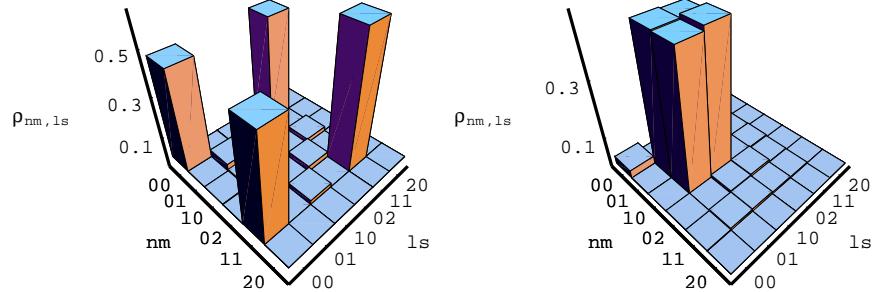
are the eigenstates of the harmonic oscillator in the position representation— $H_n(x)$  being the  $n$ th Hermite polynomial.

The ML technique can be applied to reconstruct the density matrix in the Fock basis from Monte Carlo simulated homodyne statistics. Fig. 2.7 depicts the matrix elements of the density operator as obtained for a coherent state and a squeezed vacuum, respectively. Remarkably, only 50000 homodyne data have been used for quantum efficiency  $\eta = 80\%$ . We recall that in quantum homodyne tomography the statistical errors are known to grow rapidly with decreasing efficiency  $\eta$  of the detector [78, 91]. In contrast, the elements of the density matrix reconstructed using the ML approach remain bounded, as the whole matrix must satisfy positivity and normalization constraints. This results in much smaller statistical errors. As a comparison one could see that the same precision of the reconstructions in Fig. 2.7 could be achieved using  $10^7$ – $10^8$  data samples with conventional quantum tomography. On the other hand, in order to find numerically the ML estimate we need to set *a priori* the cut-off parameter for the photon number, and its value is limited by increasing computation time.

Another relevant example is the reconstruction of the quantum state of two-mode field using single-LO homodyning. Here, the full joint density matrix can be measured by scanning the quadratures of all possible linear combinations of modes. For two modes the measured quadrature operator is given by

$$X(\theta, \psi_0, \psi_1) = \frac{1}{2} (ae^{-i\psi_0} \cos \theta + be^{-i\psi_1} \sin \theta + \text{h.c.}), \quad (2.171)$$

where  $(\theta, \psi_0, \psi_1) \in S^2 \times [0, 2\pi]$ ,  $S^2$  being the Poincaré sphere and one phase ranging between 0 and  $2\pi$ . In each run these parameters are chosen randomly. The POVM describing the measurement is given by the right-hand side of (2.167), with  $X_\varphi$  replaced by  $X(\theta, \psi_0, \psi_1)$ . An experiment for the two orthogonal states  $|\Psi_1\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$  and  $|\Psi_2\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$  has been simulated, in order to reconstruct the density matrix in the two-mode Fock basis using the ML technique. The results are reported in Fig. 2.8.



**Fig. 2.8.** ML reconstruction of the density matrix of a two-mode radiation field. On the left the matrix elements obtained for the state  $|\Psi_1\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ ; on the right for  $|\Psi_2\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ . For  $|\Psi_1\rangle$  we used 100000 simulated homodyne data and  $\eta = 80\%$ ; for  $|\Psi_2\rangle$  we used 20000 data and  $\eta = 90\%$  (From [23]).

The ML procedure can also be applied for reconstructing the density matrix of spin systems. For example, let us consider  $N$  repeated preparations of a pair of spin-1/2 particles. The particles are shared by two parties. In each run, the parties select randomly and independently from each other a direction along which they perform a spin measurement. The obtained result is described by the joint projection operator (spin coherent states [92])  $\mathcal{F}_i = |\Omega_i^A, \Omega_i^B\rangle \langle \Omega_i^A, \Omega_i^B|$ , where  $\Omega_i^A$  and  $\Omega_i^B$  are the vectors on the Bloch sphere corresponding to the outcomes of the  $i$ th run, and the indices  $A$  and  $B$  refer to the two particles. As in the previous examples, it is convenient to use an expression for the quantum expectation value  $\text{Tr}(T^\dagger T \mathcal{F}_i)$  which is explicitly positive. The suitable form is

$$\text{Tr}(T^\dagger T \mathcal{F}_i) = \sum_{\mu} |\langle \mu | T | \Omega_i^A, \Omega_i^B \rangle|^2 , \quad (2.172)$$

where  $|\mu\rangle$  is an orthonormal basis in the Hilbert space of the two particles.

Summarizing, the ML technique can be used to estimate the density matrix of a quantum system. With respect to conventional quantum tomography this method has the great advantage of needing much smaller experimental samples, making experiments with low data rates feasible, however with a truncation of the Hilbert space dimension. We have shown that the method is general and the algorithm has solid methodological background, its reliability being confirmed in a number of Monte Carlo simulations. However, for increasing dimension of Hilbert spaces the method has exponential complexity.

### 2.6.3 Gaussian-state estimation

In this Section we review the ML determination method of [31] for the parameters 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. Apart from an irrelevant phase, we consider Wigner functions of the form

$$W(x, y) = \frac{2\Delta^2}{\pi} \exp \left\{ -2\Delta^2 [e^{-2r}(x - \text{Re}\mu)^2 + e^{2r}(y - \text{Im}\mu)^2] \right\} \quad (2.173)$$

and the ML technique with homodyne detection is applied to estimate the four real parameters  $\Delta, r, \text{Re}\mu$  and  $\text{Im}\mu$ . The four parameters provide the number of thermal, squeezing and coherent-signal photons in the quantum state as follows

$$\begin{aligned} n_{th} &= \frac{1}{2} \left( \frac{1}{\Delta^2} - 1 \right), \\ n_{sq} &= \sinh^2 r, \\ n_{coh} &= |\mu|^2. \end{aligned} \quad (2.174)$$

The density matrix  $\rho$  corresponding to the Wigner function in (2.173) writes

$$\rho = D(\mu) S(r) \frac{1}{n_{th} + 1} \left( \frac{n_{th}}{n_{th} + 1} \right)^{a^\dagger a} S^\dagger(r) D^\dagger(\mu), \quad (2.175)$$

where  $S(r) = \exp[r(a^2 - a^{\dagger 2})/2]$  and  $D(\mu) = \exp(\mu a^\dagger - \mu^* a)$  denote the squeezing and displacement operators, respectively.

The theoretical homodyne probability distribution at phase  $\varphi$  with respect to the local oscillator can be evaluated using (2.7), and is given by the Gaussian

$$\begin{aligned} p(x, \varphi) &= \sqrt{\frac{2\Delta^2}{\pi(e^{2r} \cos^2 \varphi + e^{-2r} \sin^2 \varphi)}} \\ &\times \exp \left\{ -\frac{2\Delta^2}{e^{2r} \cos^2 \varphi + e^{-2r} \sin^2 \varphi} [x - \text{Re}(\mu e^{-i\varphi})]^2 \right\}. \end{aligned} \quad (2.176)$$

The log-likelihood function (2.159) for a set of  $N$  homodyne outcomes  $x_i$  at random phase  $\varphi_i$  then writes as follows

$$\begin{aligned} L &= \sum_{i=1}^N \frac{1}{2} \log \frac{2\Delta^2}{\pi(e^{2r} \cos^2 \varphi_i + e^{-2r} \sin^2 \varphi_i)} \\ &- \frac{2\Delta^2}{e^{2r} \cos^2 \varphi_i + e^{-2r} \sin^2 \varphi_i} [x_i - \text{Re}(\mu e^{-i\varphi_i})]^2. \end{aligned} \quad (2.177)$$

The ML estimators  $\Delta_{ml}, r_{ml}, \text{Re}\mu_{ml}$  and  $\text{Im}\mu_{ml}$  are found upon maximizing (2.177) versus  $\Delta, r, \text{Re}\mu$  and  $\text{Im}\mu$ .

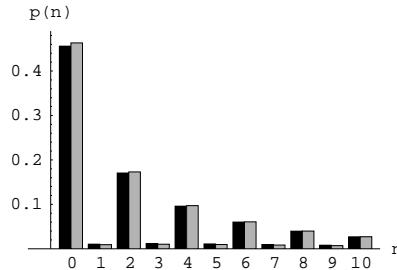
In order to evaluate globally the state reconstruction, one considers the normalized overlap  $\mathcal{O}$  between the theoretical and the estimated state

$$\mathcal{O} = \frac{\text{Tr}[\rho \rho_{ml}]}{\sqrt{\text{Tr}[\rho^2] \text{Tr}[\rho_{ml}^2]}}. \quad (2.178)$$

Notice that  $\mathcal{O} = 1$  iff  $\rho = \rho_{ml}$ . Through Monte-Carlo simulations, one always finds a value around unity, typically with statistical fluctuations over the third digit, for number of data samples  $N = 50000$ , quantum efficiency at homodyne detectors  $\eta = 80\%$ , and state parameters with the following ranges:  $n_{th} < 3$ ,  $n_{coh} < 5$ , and  $n_{sq} < 3$ . Also with such a small number of data samples, the quality of the state reconstruction is so good that other physical quantities that are theoretically evaluated from the experimental values of  $\Delta_{ml}, r_{ml}, \text{Re}\mu_{ml}$  and  $\text{Im}\mu_{ml}$  are inferred very precisely. For example, in [31] the photon number probability of a squeezed thermal state has been evaluated, which is given by the integral

$$\langle n|\rho|n\rangle = \int_0^{2\pi} \frac{d\phi}{2\pi} \frac{[C(\phi, n_{th}, r) - 1]^n}{C(\phi, n_{th}, r)^{n+1}}, \quad (2.179)$$

with  $C(\phi, n_{th}, r) = (n_{th} + \frac{1}{2})(e^{-2r} \sin^2 \phi + e^{2r} \cos^2 \phi) + \frac{1}{2}$ . The comparison of the theoretical and the experimental results for a state with  $n_{th} = 0.1$  and  $n_{sq} = 3$  is reported in Fig. 2.9. The statistical error of the reconstructed number probability affects the third decimal digit, and is not visible on the scale of the plot.



**Fig. 2.9.** Photon-number probability of a squeezed-thermal state (thermal photons  $n_{th} = 0.1$ , squeezing photons  $n_{sq} = 3$ ). Compare the reconstructed probabilities by means of the maximum likelihood method and homodyne detection (gray histogram) with the theoretical values (black histogram). Number of data samples  $N = 50000$ , quantum efficiency  $\eta = 80\%$ . The statistical error affects the third decimal digit, and it is not visible on the scale of the plot (From [31]).

The estimation of parameters of Gaussian Wigner functions through the ML method allows one to estimate the parameters in quadratic Hamiltonians of the generic form

$$H = \alpha a + \alpha^* a^\dagger + \varphi a^\dagger a + \frac{1}{2} \xi a^2 + \frac{1}{2} \xi^* a^{\dagger 2}. \quad (2.180)$$

In fact, the unitary evolution operator  $U = e^{-iHt}$  preserves the Gaussian form of an input state with Gaussian Wigner function. In other words, one can use a known Gaussian state to probe and characterize an optical device described by a Hamiltonian as in (2.180). Assuming  $t = 1$  without loss of generality, the Heisenberg evolution of the radiation mode  $a$  is given by

$$U^\dagger a U = \gamma a + \delta a^\dagger + \mu, \quad (2.181)$$

with

$$\begin{aligned} \gamma &= \cos(\sqrt{\varphi^2 - |\xi|^2}) - i \frac{\varphi}{\sqrt{\varphi^2 - |\xi|^2}} \sin(\sqrt{\varphi^2 - |\xi|^2}), \\ \delta &= -i \frac{\xi^*}{\sqrt{\varphi^2 - |\xi|^2}} \sin(\sqrt{\varphi^2 - |\xi|^2}), \\ \mu &= \frac{\varphi \alpha^* - \xi^* \alpha}{\varphi^2 - |\xi|^2} (\cos(\sqrt{\varphi^2 - |\xi|^2}) - 1) - i \frac{\alpha^*}{\sqrt{\varphi^2 - |\xi|^2}} \sin(\sqrt{\varphi^2 - |\xi|^2}). \end{aligned} \quad (2.182)$$

For an input state  $\rho$  with known Wigner function  $W_\rho(\beta, \beta^*)$ , the corresponding output Wigner function writes

$$\begin{aligned} W_{U\rho U^\dagger}(\beta, \beta^*) &= \\ W_\rho[(\beta - \mu)\gamma^* - (\beta^* - \mu^*)\delta, (\beta^* - \mu^*)\gamma - (\beta - \mu)\delta^*]. \end{aligned} \quad (2.183)$$

Hence, by estimating the parameters  $\gamma, \delta, \mu$  and inverting (2.182), one obtains the ML values for  $\alpha, \varphi$ , and  $\xi$  of the Hamiltonian in (2.180). The present example can be used in practical applications for the estimation of the gain of a phase-sensitive amplifier or equivalently to estimate a squeezing parameter.

## Acknowledgments

This work has been co-sponsored by: the Italian *Ministero dell'Istruzione, dell'Università e della Ricerca* (MIUR) under the Cofinanziamento 2003 *New generation of quantum measuring devices for photon quantum information technology*, the *Istituto Nazionale di Fisica della Materia* under the project PRA-2002-CLON, and by the European Community program ATE-SIT (Contract No. IST-2000-29681). M. G. A. P. is research fellow at *Collegio Alessandro Volta*.

## References

1. W. Heisenberg: Zeit. für Physik **43**, 172 (1927); W. Heisenberg: *The physical principle of Quantum Theory* (Univ. Chicago Press, Dover, NY, 1930).
2. J. von Neumann: *Mathematical Foundations of Quantum Mechanics* (Princeton Univ. Press, NJ, 1955).
3. W. K. Wootters and W. H. Zurek: Nature **299**, 802 (1982); H. P. Yuen: Phys. Lett. A **113**, 405 (1986).
4. G. M. D'Ariano and H. P. Yuen: Phys. Rev. Lett. **76**, 2832 (1996).
5. U. Fano: Rev. Mod. Phys. **29**, 74 (1957), Sec. 6.
6. D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani: Phys. Rev. Lett. **70**, 1244 (1993); M. G. Raymer, M. Beck, and D. F. McAlister: Phys. Rev. Lett. **72**, 1137 (1994); D. T. Smithey, M. Beck, J. Cooper, and M. G. Raymer: Phys. Rev. A **48**, 3159 (1993).
7. K. Vogel and H. Risken: Phys. Rev. A **40**, 2847 (1989).
8. G. M. D'Ariano, C. Macchiavello and M. G. A. Paris: Phys. Rev. A **50**, 4298 (1994).
9. G. M. D'Ariano, U. Leonhardt and H. Paul: Phys. Rev. A **52**, R1801 (1995).
10. M. Munroe, D. Boggavarapu, M. E. Anderson, M. G. Raymer: Phys. Rev. A **52**, R924 (1995).
11. S. Schiller, G. Breitenbach, S. F. Pereira, T. MHuller, J. Mlynek: Phys. Rev. Lett. **77**, 2933 (1996); G. Breitenbach, S. Schiller, and J. Mlynek: Nature **387**, 471 (1997).
12. U. Janicke and M. Wilkens: J. Mod. Opt. **42**, 2183 (1995); S. Wallentowitz and W. Vogel: Phys. Rev. Lett. **75**, 2932 (1995); S. H. Kienle, M. Freiberger, W. P. Schleich, and M. G. Raymer: In *Experimental Metaphysics: Quantum Mechanical Studies for Abner Shimony*, ed by Cohen S. et al. (Kluwer, Lancaster, 1997) p. 121.
13. T. J. Dunn, I. A. Walmsley, and S. Mukamel: Phys. Rev. Lett. **74**, 884 (1995).
14. C. Kurtsiefer, T. Pfau, and J. Mlynek: Nature **386**, 150 (1997).
15. D. Leibfried, D. M. Meekhof, B. E. King, C. Monroe, W. M. Itano, and D. J. Wineland: Phys. Rev. Lett. **77**, 4281 (1996).
16. G. M. D'Ariano: In *Quantum Communication, Computing, and Measurement*, ed by O. Hirota, A. S. Holevo and C. M. Caves (Plenum Publishing, New York and London, 1997) p. 253.
17. G. M. D'Ariano, P. Kumar, and M. F. Sacchi: Phys. Rev. A **61**, 013806 (2000).
18. G. M. D'Ariano: In *Quantum Communication, Computing, and Measurement*, ed by Kumar P., D'Ariano G. M. Hirota O. (Kluwer Academic/Plenum Publishers, New York and London, 2000) p. 137.
19. M. Paini: preprint quant-ph/0002078.
20. G. M. D'Ariano: Phys. Lett. A **268**, 151 (2000).
21. G. Cassinelli, G. M. D'Ariano, E. De Vito, and A. Leviero: J. Math. Phys. **41**, 7940 (2000).
22. G. M. D'Ariano and M. G. A. Paris: Acta Phys. Slov. **48**, 191 (1998); Phys. Rev. A **60**, 518 (1999).
23. K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi: Phys. Rev. A **61**, R010304 (2000).
24. G. M. D'Ariano, P. Kumar, and M. F. Sacchi: Phys. Rev. A **59**, 826 (1999).

25. G. M. D'Ariano, P. Kumar, C. Macchiavello, L. Maccone, and N. Sterpi: Phys. Rev. Lett. **83**, 2490 (1999).
26. G. M. D'Ariano, L. Maccone and M. G. A. Paris: Phys. Lett. A **276**, 25 (2000); J. Phys. A **34**, 93 (2001).
27. G. M. D'Ariano and P. Lo Presti: Phys. Rev. Lett. **86**, 4195 (2001).
28. E. P. Wigner: Phys. Rev. **40**, 749 (1932).
29. K. E. Cahill and R. J. Glauber: Phys. Rev. **177**, 1857 (1969); Phys. Rev. **177**, 1882 (1969).
30. G. M. D'Ariano, L. Maccone, and M. Paini: J. Opt. B **5**, 77 (2003).
31. G. M. D'Ariano, M. G. A. Paris and M. F. Sacchi: Phys. Rev. A **62**, 023815 (2000); Phys. Rev. A **64**, 019903(E) (2001).
32. G. M. D'Ariano and M. F. Sacchi: Nuovo Cimento **112B**, 881 (1997).
33. *The Physics of Phase Space*, ed by Y. S. Kim and W. W. Zachary (Springer, Berlin, 1986).
34. C. W. Gardiner: *Quantum noise* (Springer-Verlag, Berlin, 1991).
35. H. Weyl: *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).
36. K. E. Cahill: Phys. Rev. **138**, B1566 (1965).
37. G. M. D'Ariano, M. Fortunato and P. Tombesi: Nuovo Cimento **110B**, 1127 (1995).
38. C. T. Lee: Phys. Rev. A **44**, R2775 (1991); Phys. Rev. A **52**, 3374 (1995).
39. P. L. Kelley and W. H. Kleiner: Phys. Rev. **136**, 316 (1964).
40. H. P. Yuen and V. W. S. Chan: Opt. Lett. **8**, 177 (1983).
41. G. L. Abbas, V. W. S. Chan, S. T. Yee: Opt. Lett. **8**, 419 (1983); IEEE J. Light. Tech. **LT-3**, 1110 (1985).
42. G. M. D'Ariano: Nuovo Cimento **107B**, 643 (1992).
43. G. M. D'Ariano: Quantum estimation theory and optical detection. In: *Quantum Optics and the Spectroscopy of Solids*, ed by T. Hakioğlu and A. S. Shumovsky (Kluwer, Dordrecht, 1997) pp. 139-174.
44. K. Wodkiewicz and J. H. Eberly: JOSA B **2**, 458 (1985).
45. G. M. D'Ariano: Int. J. Mod. Phys. B **6**, 1291 (1992).
46. H. P. Yuen and J. H. Shapiro: IEEE Trans. Inf. Theory **24**, 657 (1978); **25**, 179 (1979).
47. H. P. Yuen and J. H. Shapiro: IEEE Trans. Inf. Theory **26**, 78 (1980).
48. G. M. D'Ariano, P. Lo Presti and M. F. Sacchi: Phys. Lett. A **272**, 32 (2000).
49. G. M. D'Ariano and M. F. Sacchi: Phys. Rev. A **52**, R4309 (1995).
50. C. W. Helstrom: *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
51. E. Arthurs and J. L. Kelly, Jr.: Bell System Tech. J. **44**, 725 (1965).
52. H. P. Yuen: Phys. Lett. **91A**, 101 (1982).
53. E. Arthurs and M. S. Goodman: Phys. Rev. Lett. **60**, 2447 (1988).
54. G. M. D'Ariano, C. Macchiavello, and M. G. A. Paris: Phys. Lett. A **198**, 286 (1995).
55. M. G. A. Paris: Phys. Rev. A **53**, 2658 (1996).
56. R. Baltin: J. Phys. A **16**, 2721 (1983); Phys. Lett. A **102**, 332 (1984).
57. M. G. A. Paris: Opt. Comm. **124**, 277 (1996).
58. J. H. Shapiro and S. S. Wagner: IEEE J. Quant. Electron. **QE-20**, 803 (1984).
59. J. H. Shapiro: IEEE J. Quant. Electron. **QE-21**, 237 (1985).
60. N. G. Walker: J. Mod. Opt. **34**, 15 (1987).

61. Y. Lai and H. A. Haus: *Quantum Opt.* **1**, 99 (1989).
62. M. G. A. Paris, A. Chizhov, and O. Steuernagel: *Opt. Comm.* **134**, 117 (1986).
63. A. Zucchetti, W. Vogel, and D. G. Welsch: *Phys. Rev. A* **54**, 856 (1996).
64. G. M. D'Ariano: *Phys. Lett. A* **300**, 1 (2002).
65. T. Opatrny and D. -G. Welsch: In *Prog. Opt.*, vol. XXXIX, ed by E. Wolf (Elsevier, Amsterdam, 1999) p. 63.
66. S. Weigert: *Phys. Rev. Lett.* **84**, 802 (2000).
67. H. F. Hofmann, S. Takeuchi: preprint quant-ph/0310003.
68. C. Kim and P. Kumar: *Phys. Rev. Lett.* **73**, 1605 (1994).
69. Th. Richter: *Phys. Lett. A* **211**, 327 (1996).
70. U. Leonhardt, M. Munroe, T. Kiss, M. G. Raymer, and Th. Richter: *Opt. Comm.* **127**, 144 (1996).
71. T. Kiss, U. Herzog, and U. Leonhardt: *Phys. Rev. A* **52**, 2433 (1995); G. M. D'Ariano and C. Macchiavello: *Phys. Rev. A* **57**, 3131 (1998); T. Kiss, U. Herzog, and U. Leonhardt: *Phys. Rev. A* **57**, 3131 (1998).
72. G. M. D'Ariano, S. Mancini, V. I. Manko, and P. Tombesi: *Q. Opt.* **8**, 1017 (1996).
73. K. Banaszek and K. Wodkiewicz: *Phys. Rev. Lett.* **76**, 4344 (1996).
74. T. Opatrny and D. -G. Welsch: *Phys. Rev. A* **55**, 1462 (1997).
75. F. D. Murnaghan: *The theory of group representation* (Johns Hopkins Press, 1938), p. 216.
76. U. Leonhardt and M. G. Raymer: *Phys. Rev. Lett.* **76**, 1985 (1996).
77. G. M. D'Ariano: *Acta Phys. Slov.* **49**, 513 (1999).
78. G. M. D'Ariano, C. Macchiavello, and N. Sterpi: *Quantum Semiclass Opt.* **9**, 929 (1997).
79. I. S. Gradshteyn and I. M. Ryzhik: *Table of integrals, series, and products* (Academic Press, New York, 1980).
80. Th. Richter: *Phys. Rev. A* **53**, 1197 (1996).
81. A. Orlowsky, A. Wünsche: *Phys. Rev. A* **48**, 4617 (1993).
82. A. S. Holevo: *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland Publishing, Amsterdam, 1982).
83. J. H. Shapiro and A. Shakeel: *JOSA B* **14**, 232 (1997).
84. M. Vasilyev, S.-K. Choi, P. Kumar, and G. M. D'Ariano: *Phys. Rev. Lett.* **84**, 2354 (2000).
85. G. M. D'Ariano, M. Vasilyev, and P. Kumar: *Phys. Rev. A* **58**, 636 (1998).
86. Z. Hradil: *Phys. Rev. A* **55**, R1561 (1997); in the context of phase measurement, see S. L. Braunstein, A. S. Lane, and C. M. Caves: *Phys. Rev. Lett.* **69**, 2153 (1992).
87. K. Banaszek: *Phys. Rev. A* **57**, 5013 (1998).
88. H. Cramer: *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946).
89. A. S. Householder: *The Theory of Matrices in Numerical Analysis* (Blaisdell, New York, 1964) Sec. 5.2.
90. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery: *Numerical Recipes in Fortran* (Cambridge Univ. Press, Cambridge, 1992) Sec. 10.4.
91. G. M. D'Ariano and M. G. A. Paris: *Phys. Lett. A* **233**, 49 (1997).
92. F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas: *Phys. Rev. A* **6**, 2211 (1972).
93. F. Natterer: *The mathematics of computerized tomography* (Wiley, New York, 1986).

94. P. Mansfield and P. G. Morris: *NMR Imaging in Biomedicine* (Academic Press, New York, 1982).



## 3 Maximum-Likelihood Methods in Quantum Mechanics

Zdeněk Hradil<sup>1</sup>, Jaroslav Řeháček<sup>1</sup>, Jaromír Fiurášek<sup>1,2</sup>, and Miroslav Ježek<sup>1</sup>

<sup>1</sup> Department of Optics, Palacky University, Olomouc, Czech Republic  
[hradil@aix.upol.cz](mailto:hradil@aix.upol.cz)

<sup>2</sup> Ecole Polytechnique, CP 165, Université Libre de Bruxelles, 1050 Bruxelles, Belgium

*Yet your mistrust cannot make me a traitor:  
Tell me whereon the likelihood depends.*

W. Shakespeare: As you like it

### 3.1 Introduction

Our world is controlled by chance and everybody must cope with it. Events may but need not happen. Success and surviving depends on the ability of any living creature to set the risk associated with a particular alternative and to chose the optimal one. Chance has probably played a very important role in the formation of life on the Earth as Darwin noticed centuries ago. But chance manifests itself even in less sophisticated aspects of everyday life. Insurance companies, share markets, lotteries, bookmakers, and betting agencies count on chance. Chance has been commercialized and converted to money.

It is intriguing to note that the chance in classical physics is caused by the enormous complexity of the world, due to which deterministic solutions are intractable in reality. In this sense, the origin of chance may be explained and quantified. The notion of chance acquires dramatically different meaning in the quantum domain, since quantum theory is intrinsically a probabilistic theory. Chance expressed by probabilities is here an indispensable ingredient of any effect and deserves special attention. All information about the quantum world comes through the probabilities. According to the Copenhagen interpretation of quantum theory, probabilities are coded in a quantum state describing the available knowledge about the system. Quantum theory handles observable events on the most fundamental level currently available, predicting the statistics of quantum phenomena.

The statistical nature of the quantum world is revealed in the experimental reality. When performing repeatedly an experiment on the ensemble of identically prepared systems, the experimenter cannot deterministically control their result due to the unavoidable fluctuations. But what can be

observed is the statistics of certain events, expressed either in the form of probabilities or in the form of average values and their moments. Quantum theory answers the question: What statistics can be expected provided that the quantum state is known? This demonstrates the central role played by the quantum state. Its knowledge makes it possible to predict statistical results of any measurement performed on the same system. The determination of the quantum state itself represents an inverse problem, and as it is well known, inverse problems are more involved than direct ones. The inverse problem of determining the quantum state is called quantum state reconstruction or quantum tomography.

Though the history of this problem can be traced back to the early days of quantum mechanics, namely to the Pauli problem [1], experimental utilization had to wait until quantum optics has opened a new era. The theoretical predictions of Vogel and Risken [2] were closely followed by the experimental realization of the suggested algorithm by Smits et al. [3]. Since that time many improvements and new techniques have been proposed, an overview can be found in Ref. [4]. Reconstruction of quantum states is now considered a standard technique used in various branches of contemporary physics [5–7]. Methods of this kind treat the problem of quantum state reconstruction as a special case of Radon transformation used in medicine X-ray tomography. Chapter by Raymer and Beck of this volume is devoted to various aspects of this wide spread treatment, which will be called *standard* in the following.

The purpose of this chapter is to contribute to the mosaic of quantum state reconstruction techniques presented in the book, and to present an alternative to standard schemes. The idea behind this is surprisingly simple. In accordance with the elucidation of the role of chance in our world, it could be paraphrased, loosely speaking, as the following strategy: Always bet on the most likely interpretation. This method, is known in mathematical statistics as the maximum likelihood (ML) estimation. It was proposed in the 1920s by R. Fisher.

The principle of maximum likelihood is not a rule that requires justification – it does not need to be proved. Nowadays it is widely used in many applications in optics, noise analysis, communications, geophysics, nuclear physics etc. Among the most curious and spectacular applications that we can find, is the discovery of tunnels in the demilitarized zone between North and South Korea [8]. What makes this technique so attractive and powerful is its efficiency. It plays the prominent role among the other estimation techniques since it may reach the ultimate limit of accuracy predicted by Cramer-Rao bound. In this chapter we will stress out another remarkable feature, namely that ML estimation is predetermined for solving inverse quantum problems. It naturally reproduces the structure of quantum theory such as the closure or uncertainty relations. This fundamental reason justifies its usage, though it leads to nonlinear algorithms and its numerical implementation is demanding on computing resources.

In Section 3.2 various reconstruction techniques will be compared. Particularly, the essence of the standard quantum state reconstruction will be briefly reviewed pointing out weak points of this technique. This will serve as motivation for the ML estimation, which will be further advocated as a versatile and natural tool for solving quantum problems. For the sake of completeness, ML quantum-state tomography will be briefly compared with alternative statistical approaches to this problem, such as Bayesian and Maximum Entropy (MaxEnt) estimations.

In Section 3.3 basics of ML estimation of the quantum state will be presented using the Jensen inequality. Since this will serve as a paradigm for further applications, an elementary exposition on the introductory level will be given. As a result, the extremal equation will be derived and its interpretation will be given and compared to standard reconstructions.

In Section 3.3.4 the ML technique will be treated as a problem of a statistical distance. The extremal equation will be re-derived using variational approach, which is more suitable for further generalizations. The numerical aspects of the extremal solution will be detailed on the example of a reconstruction of two entangled qubits.

Section 3.4 will link the problem of the state reconstruction with the operational information. As will be shown, the information content acquired in the course of measurement may be quantified by means of Fisher information matrix. Loosely speaking, it grasps the width of the likelihood functional.

In Section 3.5 we will apply the technique of ML estimation to the reconstruction of quantum processes. Section 3.6 will address the estimation of quantum measurements, and finally Section 3.7 will consider the problem of quantum discrimination motivated by the ML estimation technique.

## 3.2 Overview of quantum-state reconstruction techniques

To address the problem of quantum state reconstruction, let us consider a generic quantum measurement. The formulation will be developed for the case of finite dimensional quantum systems. The reader can think of a spin 1/2 system for the sake of simplicity. Applications to infinite dimensional Hilbert spaces will be mentioned separately.

Assume that we are given a finite number  $N$  of identical copies of the system, each in the same but unknown quantum state described by the density operator  $\rho$ . Given those systems our task is to identify the unknown *true* state  $\rho$  as accurately as possible from the results of measurements performed on them. For simplicity we will assume all measurements in the sense of von Neumann. The generalization to a positive valued operator measure (POVM) is straightforward.

Let us consider, for concreteness, that  $N$  particles prepared in the same state have been observed in  $M$  different output channels of the apparatus.

For spin one-half particles those channels could be for instance the  $M = 6$  output channels of the Stern-Gerlach apparatus oriented along  $x$ ,  $y$ , and  $z$  directions, respectively.

Provided that each particular output

$$|y_j\rangle\langle y_j|, \quad j = 1, \dots, M \quad (3.1)$$

has been registered  $n_j$  times,  $\sum_j n_j = N$ , the relative frequencies are given as  $f_j = n_j/N$ . Using these ingredients, the true state  $\rho$  is to be inferred. For the sake of simplicity, the measurement performed will be assumed as complete, i.e.

$$H \equiv \sum_j |y_j\rangle\langle y_j| = 1. \quad (3.2)$$

This condition is satisfied for Stern-Gerlach measurements mentioned above. Later we will release this condition and consider incomplete measurements too. The probabilities of occurrences of various outcomes are generated by the true quantum state  $\rho$  according to the well-known quantum rule,

$$p_j = \langle y_j | \rho | y_j \rangle. \quad (3.3)$$

Let us now briefly discuss various approaches to the quantum state reconstruction.

### 3.2.1 Standard reconstruction

If the probabilities  $p_j$  of getting a sufficient number of different outcomes  $|y_j\rangle$  were known, it would be possible to determine the true state  $\rho$  directly by inverting the linear relation (3.3). This is the philosophy behind the “standard” quantum tomographic techniques. For example, in the rather trivial case of a spin one half particle, the probabilities of getting three linearly independent projectors determine the unknown state uniquely. Here, however, a serious problem arises. Since only a finite number of systems can be investigated, there is no way to find out those probabilities. The only data one has at his or her disposal are the relative frequencies  $f_j$ , which sample the *principally* unknowable probabilities  $p_j$ . It is obvious that for a small number of runs, the true probabilities  $p_j$  and the corresponding detected frequencies  $f_j$  may differ substantially. As a result of this, the modified realistic problem,

$$f_j = \langle y_j | \rho | y_j \rangle, \quad (3.4)$$

has generally no solution on the space of semi-positive definite hermitian operators describing physical states. This linear equation for the unknown density matrix may be solved for example by means of pattern functions, see e.g. [4, 9], what could be considered as a typical example of the standard approach suffering from the above mentioned drawbacks. In the case

of homodyne detection the discrete index  $j$  should be replaced by the pair of continuous variables  $x$  and  $\phi$  denoting the detected value and phase of a quadrature operator. The quantum state of a single-mode quantum field represented by the Wigner function  $W(\alpha)$  can then be formally written as the following integral transformation

$$W(\alpha) = \int d\phi dx K(x, \phi) f(x, \phi), \quad (3.5)$$

where  $K(x, \phi)$  denotes a kernel of the integral transformation. Notice, that such a standard solution is linear with respect to data. No matter how ingeniously the inversions in the standard reconstruction schemes are done, there always remains some problems caused by the application of this treatment in quantum theory. All measurements in practice are inevitably limited as far as the amount and accuracy of data is concerned. The continuous parameter can be scanned only in a limited number of positions, and such a measurement done with a finite ensemble will always be affected by fluctuations. Moreover, the kernels involved in the standard reconstruction of the type (3.5) are often singular. When the standard approach is applied to real data, serious problems with positivity of reconstructed density matrix may appear. Though these simple techniques give us a rough picture of the unknown state, they are not able to provide the full quantum description.

### 3.2.2 ML estimation

Having measurements done and their results registered, the experimenter's knowledge about the measured system is increased. Since quantum theory is probabilistic, it makes little sense to pose the question: "What quantum state is determined by that measurement?" A more appropriate question is: "What quantum states seem to be most likely for that measurement?"

Quantum theory predicts the probabilities of individual detections, see Eq. (3.3). From them one can construct the total joint probability of registering data  $\{n_j\}$ . We assume that the input system (particle) is always detected in one of  $M$  output channels, and this is repeated  $N$  times. Subsequently, the overall statistics of the experiment is multinomial,

$$\mathcal{L}(\rho) = \frac{N!}{\prod_i n_i!} \prod_j \langle y_j | \rho | y_j \rangle^{n_j}, \quad (3.6)$$

where  $n_j = Nf_j$  denotes the rate of registering a particular outcome  $j$ . In the following we will omit the multinomial factor from expression (3.6) as it has no influence on the results. Physically, the quantum state reconstruction corresponds to a synthesis of various measurements done under different experimental conditions, performed on the ensemble of identically prepared systems. For example, the measurement might be a subsequent recording of

an unknown spin of the neutron (polarization of the photon) using different settings of the Stern Gerlach apparatus, or the recording of the quadrature operator of light in rotated frames in quantum homodyne tomography. The likelihood functional  $\mathcal{L}(\rho)$  quantifies the degree of belief in the hypothesis that for a particular data set  $\{n_j\}$  the system was prepared in the quantum state  $\rho$ . The ML estimation simply selects the state for which the likelihood attains its maximum value on the manifold of density matrices.

The mathematical formulation of ML estimation in quantum theory will be worked out below showing that it provides a feasible statistical inversion. This is guaranteed by the convexity of both the set of density matrices and likelihood functional. As will be demonstrated, provided that the standard method yields a physically sound (semipositive) solution, the same density matrix is also a ML solution. On the contrary to standard approaches, ML principle is a highly nonlinear fitting procedure. Let us clarify some physical reasons for why it is so.

For simplicity, let us consider the textbook example of a spin 1/2 system and simple Stern-Gerlach (SG) detections. Let us assume that several projections have been registered. How to use all of them without discarding any information? Some observations may appear incompatible with each other due to the fluctuations and noises involved. That is why the overdetermined system of equations (3.4) is inapplicable in the presence of noise. Moreover, different SG projections cannot be treated on equal footing. They are observing various “faces” of the spin system. Such measurements, even when done with equal numbers of particles, possess different errors. It is not difficult to see why. The measured frequencies fluctuate around the true probabilities according to the binomial statistics, and significantly, their root-mean square errors depend on the overlap between the projections and the true (but unknown!) spin state. Various SG measurements are also incompatible in the sense of quantum theory, because projections to different directions do not commute. Such data cannot be obtained in the same measurement, but may be collected subsequently in the course of repeated detections, where, in general, different observations will be affected by different errors. But these errors in turn depend on the unknown state. The estimation procedure must therefore predict the unknown state and consider the data fluctuations simultaneously. This indicates the nonlinearity of the optimal estimation algorithm. As will be demonstrated later, ML estimation does this job.

### 3.2.3 Other statistical approaches

In this short subsection let us mention two other statistically motivated approaches to the quantum state reconstruction problem. The heart of the Bayesian statistical inference is Bayes’s rule [10]

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}, \quad (3.7)$$

where  $p(D|H)$  is the probability of data  $D$  given hypothesis  $H$ ,  $p(H)$  is the prior probability that hypothesis  $H$  is true,  $P(H|D)$  is the posterior probability that hypothesis  $H$  is true given data  $D$ , and the normalization  $p(D) = \sum_H p(D|H)p(H)$  is the marginal probability of data  $D$ . Bayes's rule shows how the acquired data  $D$  updates our knowledge,  $p(H) \rightarrow p(H|D)$ . This formulation is classical in the sense that it involves classical probabilities. There were attempts to extend it to quantum domain [11]. On the conceptual level this can be done easily associating hypothesis  $H$  with a quantum state,  $H \rightarrow \rho$ , and data  $D$  with a particular output of the measurement. Invoking the linearity of the quantum theory, the quantum state associated with the posterior distribution  $p(\rho)$  on the space of density matrices reads

$$\rho = \int d\rho p(\rho)\rho. \quad (3.8)$$

The linkage of the Bayesian reasoning to ML seems to be clear – it represents just a rough approximation replacing a probability distribution by a single peak localized at the most likely state. This is true, but the problem of quantum Bayesian inference is more complicated than it looks. Namely, it is not clear what should be chosen as the proper measure for the integration over the manifold of mixed states. Beside this, when resorting to Bayesian estimation the inferred state is not going to be the maximally likely one in general. This is a bit counterintuitive, since why it should be better to throw away the chance and diminish the probability of success resorting to a less probable interpretation?

Another statistically motivated approach is the Maximum entropy inference devised by Jaynes [12] and since then applied to many physical problems [13]. Its formulation hinges on the entropy as a measure of information. Loosely speaking, the MaxEnt principle looks for maximally unbiased solutions fulfilling the given set of constraints. In tomographic applications, the constraints depend on the registered data. In this sense the MaxEnt reconstruction provides the most pessimistic guess compatible with the observations made. Its mathematical formulation is conceptually simple. Entropy of a quantum state is defined as follows,

$$S = -k\text{Tr}[\rho \log \rho], \quad (3.9)$$

$k$  being an arbitrary constant (Boltzmann constant  $k_B$  in statistical physics, here we will set it to  $k = 1$ ). As constraints we usually fix the expectation values of some set of observables or projectors,

$$a_i = \text{Tr}[\rho A_i], \quad (3.10)$$

$a_i$  are real numbers inferred from the measurement, and  $A_i$  are the corresponding observables (q-numbers). The normalization,  $\text{Tr}\rho = 1$ , can be taken as an additional constraint with  $a_0 = 1$  and  $A_0 = \hat{1}$ . MaxEnt solution can always be written in the form,

$$\rho = \exp\left[\sum_i \lambda_i A_i\right], \quad (3.11)$$

where the Lagrange multipliers  $\lambda_i$  can be determined from the constraints (3.10) by solving a set of coupled nonlinear equations. This can be difficult, particularly when the operators  $A_i$  do not commute.

To compare the MaxEnt and ML estimations let us have a closer look on the constraints corresponding to the sampled probabilities ( $A_i = |i\rangle\langle i|$ ). It is not difficult to see that such constraints are equivalent to problem (3.4) of the standard reconstruction scheme. Consider the case of a  $p$ -dimensional Hilbert space, where the unknown quantum state is uniquely specified by  $p^2 - 1$  parameters. This is also the maximum number of constraints which may be satisfied simultaneously. But even in the case of under-determined problems, some constraints may still appear as inconsistent due to the fluctuations present in the observed data. This is a serious obstacle for the straightforward application of the MaxEnt principle to the sampled probabilities. To get a physically correct reconstruction, the constraints have to be released, see Bužek in this volume.

### 3.3 ML quantum-state estimation

#### 3.3.1 Extremal equation

The mathematical formulation of ML estimation will be developed here. The likelihood functional corresponding to detected data (3.6) should be maximized on the manifold of density matrices. This will guarantee getting a correct and physically sound interpretation of obtained results. No additional assumptions are needed. Here we will formulate the ML extremal equation for a generic case of projective measurements  $|y_i\rangle\langle y_i|$ ; generalization to measurements described by elements of a probability valued operator measure is straightforward as will be shown later.

ML estimation consist in maximizing the likelihood functional,

$$\mathcal{L}(\rho) = \prod_i \langle y_i | \rho | y_i \rangle^{n_i}. \quad (3.12)$$

Here, and in the following, we omit the unimportant multinomial multiplicative factor. In order to maximize Eq. (3.12), the Jensen inequality between the geometric and arithmetic averages will be adopted,

$$\prod_i \left[ \frac{x_i}{a_i} \right]^{f_i} \leq \sum_i f_i \frac{x_i}{a_i}, \quad (3.13)$$

where  $x_i \geq 0$  and  $a_i > 0$  are auxiliary positive nonzero numbers and  $f_i = n_i/N$ . The equality sign is reached if and only if all the numbers  $x_i/a_i$  are

equal. We will use boldface to denote vectors, i.e.  $\mathbf{a}$  and  $\mathbf{x}$ . Setting  $x_i = \langle y_i | \rho | y_i \rangle$ , the dimension of these vectors is given by the number of independent projections, parameters  $a_i$  being a subject of further considerations. This inequality may be easily adopted for the maximization of likelihood since,

$$(\mathcal{L}(\rho))^{1/N} = \prod_i \left( \langle y_i | \rho | y_i \rangle \right)^{f_i} \leq \prod_j a_j^{f_j} \text{Tr}\{\rho R(\mathbf{y}, \mathbf{a})\} \quad (3.14)$$

$$\leq \lambda(\mathbf{y}, \mathbf{a}) \prod_i a_i^{f_i}. \quad (3.15)$$

Operator  $R$  is a positive semi-definite operator defined by its expansion into the measured, generally non-orthogonal projectors as:

$$R(\mathbf{y}, \mathbf{a}) = \sum_i \frac{f_i}{a_i} |y_i\rangle\langle y_i|. \quad (3.16)$$

Relation (3.14) follows simply from the definition of the likelihood (3.12) and from the Jensen inequality (3.13) for any allowed vector  $\mathbf{a}$ . The last inequality (3.15) represents the maximization of the previous expression over all density matrices. Parameter  $\lambda(\mathbf{y}, \mathbf{a})$  representing the largest eigenvalue of the operator  $R$  defines the upper bound of the likelihood. Let us clarify under what conditions the equality sign will be achieved in the chain of inequalities (3.14) and (3.15). Fixing the parameters  $a_i$ , the most general extremal state  $\rho_e$  should have its support in the subspace corresponding to the maximal eigenvalue  $\lambda$ ,

$$R(\mathbf{y}, \mathbf{a})\rho_e = \lambda(\mathbf{y}, \mathbf{a})\rho_e. \quad (3.17)$$

This guarantees the equality sign in the inequality (3.15). Setting further,  $a_i = \langle y_i | \rho_e | y_i \rangle$ , the equality sign in (3.14) will be achieved too. Put together, the extremal equation for the ML density matrix reads,

$$R\rho = \rho, \quad (3.18)$$

where the operator

$$R = \sum_i \frac{f_i}{\langle y_i | \rho | y_i \rangle} |y_i\rangle\langle y_i| \quad (3.19)$$

is state dependent, Lagrange multiplier  $\lambda = 1$  by normalization, and the subscript “e” was omitted for brevity. Notice also, that there are many equivalent forms of this extremal equation. Particularly, the extremal state is seen to commute with the operator  $R$ ,  $[R, \rho] = 0$ , since  $\rho$  is hermitian. An alternative form of extremal equation reads,

$$\sum_i \frac{f_i}{\langle y_i | \rho | y_i \rangle} \Pi_i = 1_\rho, \quad (3.20)$$

where  $1_\rho$  denotes the identity operator defined on the support of the (unknown) extremal density matrix, and  $\Pi_i = |y_i\rangle\langle y_i|$  are POVM elements corresponding to measurement.

### 3.3.2 Example of simple measurements

Let us show some special solutions linking the ML estimation to standard reconstruction schemes. The latter, based on linear inversion, is obtained as the solution of equations  $f_i = \langle y_i | \rho | y_i \rangle$ . If a semi-positive definite solution exists, such a density matrix will also maximize the ML functional. This follows from the Gibbs inequality, [14]

$$\sum_i f_i \ln \langle y_i | \rho | y_i \rangle \leq \sum_i f_i \ln f_i. \quad (3.21)$$

Consider the textbook example [15] of a measurement represented by a single nondegenerate hermitian operator  $A$ ,  $A|a\rangle = a|a\rangle$ . Its discrete orthonormal eigenvectors,  $\langle a|a' \rangle = \delta_{aa'}$  provide the closure relation  $\sum_a |a\rangle\langle a| = \hat{1}$ . The probability of measuring a particular value of  $A$  on a system that has been prepared in a quantum state  $\rho$  is  $p_a = \langle a | \rho | a \rangle$ . When this measurement is repeated on  $N$  identical copies of the system, each outcome  $a$  occurring  $n_a$  times, the relative frequencies  $f_a = \frac{n_a}{N}$  will sample the true probabilities  $p_a$  reproducing them only in the limit  $N \rightarrow \infty$ . Experimenter's knowledge can then be expressed in the form of a diagonal density matrix

$$\rho_{est} = \sum_a f_a |a\rangle\langle a|. \quad (3.22)$$

This state is apparently semi-positive definite, and as a consequence of inequality (3.21) it also maximizes the likelihood functional. In formula (3.22) we have just stated the result of the experiment  $\{n_a\}$  in one special way; in fact, no tomography is involved there. Similar knowledge is obtained by compatible observations, i.e. by measurements described by commuting operators with a common diagonalizing basis  $\{|a\rangle\}$ .

In order to find also non-diagonal elements of the unknown density matrix, some other measurements non-compatible with  $A$  should be adopted. This can be done by projecting the unknown state onto a set of non-commuting eigenvectors of operators  $A_j$ ,  $j = 1, 2, \dots$ . Such a measurement is going to give us more information than just the diagonal elements of the density matrix in some a priori *given* basis. Maximum-likelihood synthesis and interpretation of such general observations will be given in the following section.

### 3.3.3 Interpretation of the ML solution

An intriguing property of ML extremal equation (3.20) is that it resembles the closure relation  $\sum_j |y_j\rangle\langle y_j| = \hat{1}$ . Indeed, in terms of re-normalized projectors  $\Pi'_i = (f_i/p_i)\Pi_i$ , the extremal equation reads

$$\sum_i \Pi'_i = 1_\rho. \quad (3.23)$$

Moreover, the expectation values of such re-normalized projectors taken with the extremal state reproduce identically the detected frequencies

$$\text{Tr}(\rho \Pi_i') \equiv f_i, \quad \forall i. \quad (3.24)$$

Maximum likelihood estimation thus can be interpreted as a renormalization of observations done, which apart from being able to fit the registered data exactly provides a resolution of unity and thus represents a genuine quantum measurement.

This can now be contrasted to the standard formulation of the estimation problem which is based on the solution of a linear set of equations,

$$\text{Tr}(\rho \Pi_i) = f_i, \quad \forall i, \quad (3.25)$$

while the closure relation,

$$\sum_i \Pi_i \equiv 1, \quad (3.26)$$

is fulfilled as identity. Notice that roles of the closure relation and probability rule in ML approach are *reversed* with respect to that in the standard formulation.

Let us stress once more that the standard reconstruction has often no solution on the set of density matrices, whereas the ML formulation has always a solution. If the standard reconstruction yields a density matrix, then this state is also maximum-likely quantum state, but not vice versa.

Another interesting feature of the ML reconstruction technique is that the method itself, through the closure relation (3.23), defines “the field of view” of the tomographic scheme as the subspace, where the reconstruction is done. No such interpretation of the standard solution is possible. We consider this mutual relationship between the standard and ML reconstructions as a strong argument in favor of the latter method.

In view of the above analysis one may wonder why standard methods are so popular. The answer is that in many applications, the ensembles of measured systems are so huge, that fluctuations, which scale as  $1/N$ , can entirely be neglected. Standard methods are then preferred due to their simplicity and much lower computing costs.

As will be shown later in Sec. 3.4, the ignoring of the positivity constraint always lowers the accuracy of the estimation procedure. However, this undesirable effect may be suppressed by enlarging the ensemble of probed particles, at least in applications where the positivity is not crucial. That is why the X-ray tomography in medicine is such a wonderful tool, though the standard back projection based on the deterministic inverse Radon transformation is used on fluctuating data. However, in applications with much smaller count rates, such as in neutron radiography discussed by Badurek *et al.* in this volume, the ill posed nature of inverse problems fully reveals itself by the appearance of severe artifacts. The same may and usually does happen in quantum tomography.

### 3.3.4 Maximum likelihood as a statistical distance

Probabilistic interpretation of quantum theory suggests that a statistical treatment of the observed data [17–19] is more natural and appropriate than the deterministic one. Maximum likelihood estimation [17], which is the main subject of this chapter looks for the most likely quantum state. The likelihood Eq. (3.12) quantifies our degree of belief in the given hypothesis  $\rho$ . However, one can also think of the likelihood as a sort of statistical distance  $D[\mathbf{f}, \mathbf{p}(\rho)]$  between the theory  $p_j$  and data  $f_j$ . Via its minimization we find the density matrix  $\rho_e$  that generates through Eq. (3.3) probabilities  $p_j$  lying as “close” to the observed frequencies  $f_j$  as possible. This statistical distance is known in the theory of statistics as the relative entropy or Kullback-Leibler divergence [21]:

$$D[\mathbf{f}, \mathbf{p}] = - \sum_j f_j \ln p_j. \quad (3.27)$$

Indeed, adopting metric (3.27) is equivalent to finding the maximum of the likelihood functional (3.12).

At first sight it might seem that there is no reason to *prefer* one particular metric to another one – different metrics leading to different results. This ambiguity can be resolved by considering the formal description of the reconstruction process [20]. If the whole measurement and subsequent reconstruction is looked at as a single generalized measurement, see Sec. 3.3.3, then the *relation* between the actually performed measurement and resulting probability operator measure becomes particularly simple and easy to interpret if Kulback-Leibler distance (3.27) is adopted.

ML methods are well-known in the field of inverse problems and they have found many applications in reconstructions and estimations so far [22–25]. Unfortunately, except in most simple cases, the maximization of the likelihood functional is a challenging problem on its own.

In the following we will show two different but related ways to maximize the likelihood (3.12). The first one is based on splitting the complex problem of estimating the density operator into two simpler tasks: finding the optimal eigenvalues and eigenvectors. The second approach relies on a direct application of the calculus of variations [17, 26].

Both these routes will provide alternative derivations of the extremal equation (3.18) of Sec. 3.3.1, and give us some hints how to solve it iteratively.

### 3.3.5 Maximization of the likelihood

In the classical signal processing an important role is played by linear and positive (LinPos) problems [27, 28]. Since these are closely related to the problem of quantum state reconstruction it is worthwhile to recall how the positive and linear problems can be dealt with using the ML approach.

Let us consider that the probabilities  $p_j$  of getting outcomes  $y_j$  are given by the following linear and positive relation

$$p_j = \sum_i r_i h_{ij}, \quad \mathbf{p}, \mathbf{r}, \mathbf{h} > \mathbf{0}. \quad (3.28)$$

Here  $\mathbf{r}$  is the vector describing the “state” of the system. For example, the reconstruction of a one-dimensional object from the noiseless detection of its blurred image could be accomplished by inverting the relation (3.28), where  $\mathbf{r}$  and  $\mathbf{p}$  would be the normalized intensities of the object and image, and  $\mathbf{h}$  would describe the blurring mechanism. Again here the presence of noise ( $f_j \neq p_j$ ) tends to spoil the positivity of the reconstructed intensity  $\mathbf{r}$ .

The solution to LinPos problems in the sense of ML can be found using the expectation-maximization (EM) algorithm [27, 28],

$$r_i^{(n)} = r_i^{(n-1)} \sum_j \frac{h_{ij} f_j}{p_j(\mathbf{r}^{(n-1)})}, \quad (3.29)$$

which if initialized with a positive vector  $\mathbf{r}^0$  ( $r_i^0 > 0 \forall i$ ) is guaranteed to converge to the global minimum of the Kullback-Leibler divergence  $D[\mathbf{f}, \mathbf{p}]$  for any input data  $\mathbf{f}$ .

This algorithm is convenient from the point of view of the numerical analysis. It is certainly much more convenient than the direct multidimensional maximization of the corresponding ML functional  $\ln \mathcal{L} = \sum_j f_j \ln p_j$  [29]. This brings us back to the problem of quantum state reconstruction. It would be nice to have a similar iterative algorithm for dealing with the problem (3.4), or equivalently for maximizing the ML functional (3.12). On the one hand it is clear that the problem of quantum state reconstruction is not a LinPos problem, since the quantum rule (3.3) cannot be rewritten to the form of Eq. (3.28) with a known positive kernel  $\mathbf{h}$ . As a consequence of this, the EM algorithm cannot be straightforwardly applied here. On the other hand the reconstruction of the elements of the density matrix becomes a LinPos problem if the eigenbasis diagonalizing the density matrix is known. In this case the unknown density matrix can be parametrized as follows

$$\rho = \sum_k r_k |\phi_k\rangle\langle\phi_k|, \quad \rho|\phi_k\rangle = r_k |\phi_k\rangle, \quad (3.30)$$

where  $r_i$  are eigenvalues of  $\rho$ , the only parameters which remain to be determined from the performed measurement. Using Eq. (3.30), relation (3.3) can easily be rewritten to the form of LinPos problem Eq. (3.28).

Unlike the standard quantum measurement of a single observable, a series of sequential measurements of many non-commuting operators determines not only the diagonal elements of  $\rho$  but also the diagonalizing basis itself. This hints on splitting the quantum state reconstruction into two subsequent steps: the reconstruction of the eigenvectors of  $\rho$  in a fixed basis, which represents

the classical part of the problem, followed by the “rotation” of the basis  $\{|\phi_i\rangle\}$  in the “right” direction using the unitary transformation

$$|\phi'_k\rangle\langle\phi'_k| = U|\phi_k\rangle\langle\phi_k|U^\dagger. \quad (3.31)$$

Its infinitesimal form reads

$$U \equiv e^{i\epsilon G} \approx 1 + i\epsilon G. \quad (3.32)$$

Here  $G$  is a hermitian generator of the unitary transformation and  $\epsilon$  is a positive real number small enough to make the second equality in Eq. (3.32) approximately satisfied.

Consider now the total change of the log-likelihood caused by the change of diagonal elements of density matrix and rotation of basis. Keeping the normalization condition  $\text{Tr}\rho = 1$ , the first order contribution to the variation reads

$$\delta \ln \mathcal{L} = \sum_k \delta r_k (\langle\phi_k|R|\phi_k\rangle - 1) + i\epsilon \text{Tr}\{G[\rho, R]\}. \quad (3.33)$$

Operator  $R$  was defined earlier in Eq. (3.19). We remind the reader that it is a semi-positive definite operator depending on  $\rho$ .

Inspection of Eq. (3.33) reveals a simple two-step strategy to make the likelihood of the new state  $\rho'$  as high as possible [within the limits of the validity of the linearization (3.32), of course.] In the first step, the first term on the right hand side of Eq. (3.33) is maximized by estimating the eigenvalues of the density matrix keeping its eigenvectors  $|\phi_k\rangle$  constant. The iterative EM algorithm (3.29) described above can straightforwardly be applied to this LinPos problem. As the second step, the likelihood can further be increased by making the second term on the right hand side of Eq. (3.33) positive. This is accomplished by a suitable choice of the generator  $G$  of the unitary transformation (3.32). Recalling the natural norm induced by the scalar product defined on the space of operators,  $(A, B) = \text{Tr}\{A^\dagger B\}$ , the generator  $G$  may be chosen as

$$G = i[\rho, R]. \quad (3.34)$$

This choice guarantees the non-negativity of its contribution to the likelihood and is optimal in the sense of the above introduced scalar product. Notice that this derivation holds only if the second order contribution in  $\epsilon$  to (3.32) and (3.33) is negligible. From this, an upper bound on the value of  $\epsilon$  can be derived. Parameter  $\epsilon$  can then be adaptively changed in each  $U$  step in order to minimize the computing time.

Now we have at our disposal all ingredients comprising the EMU quantum state reconstruction algorithm [30]. Starting from some strictly positive density matrix  $\rho^0$ , this initial guess is improved, first by finding new eigenvalues using the EM iterative algorithm (3.29), and then again by finding new eigenvectors by a suitable unitary ( $U$ ) transformation of the

old ones according to Eqs. (3.31-3.32) and (3.34). Repetition of these two steps, each monotonically increasing the likelihood of the current estimate, resembles climbing a hill. Convexity of the likelihood functional (3.12),  $\mathcal{L}(\alpha\rho_1 + (1 - \alpha)\rho_2) \geq \alpha\mathcal{L}(\rho_1) + (1 - \alpha)\mathcal{L}(\rho_2)$ ,  $\alpha \in \langle 0, 1 \rangle$ , then guarantees that the global maximum is always attained eventually.

The EMU algorithm naturally leads to the previously introduced extremal equation for the density matrix [17, 26]. The stationary point of EMU algorithm is characterized by the vanishing variation of the log likelihood (3.33). Since the variations  $\delta r_k, \epsilon$  are arbitrary parameters, this is equivalent to the Lagrange-Euler equation for density matrix,

$$R\rho_e = \rho_e, \quad (3.35)$$

that was already derived in Sec. 3.3.1 by other means.

For the sake of completeness, let us present one more derivation based on the calculus of variations [26]: For generality let us assume that a generalized measurement described by POVM elements  $\{\Pi_j\}$ ,  $\sum_j \Pi_j = 1$  was performed repeatedly  $N$  times. The theoretical probability of observing outcome  $j$  is  $p_j = \text{Tr}\rho\Pi_j$ . Let us now maximize the logarithm of the likelihood functional on the space of positive operators subject to normalization  $\text{Tr}\rho = 1$ . The latter constraint will be dealt with using an undetermined Lagrange multiplier  $\lambda$  while the positivity is ensured by decomposing  $\rho$  as follows,  $\rho = A^\dagger A$ . A necessary condition for  $\rho$  to be the maximum likely state is now that the functional,

$$F = \sum_j f_j \ln (\text{Tr}\{A^\dagger A\Pi_j\}) - \lambda \text{Tr}\{A^\dagger A\} \quad (3.36)$$

is stationary in the independent variable  $A$ . Varying  $A$  to  $A + \delta A$ , where  $\delta A$  is a small variation, the value of  $F$  will change as follows:

$$\delta F = \sum_j \frac{f_j}{p_j} \text{Tr}\{\Pi_j A^\dagger \delta A\} - \lambda \text{Tr}\{A^\dagger \delta A\}. \quad (3.37)$$

Since  $\rho$  is an extremal point,  $\delta F$  must vanish for all  $\delta A$ , which means that

$$\sum_j \frac{f_j}{p_j} \Pi_j A^\dagger = \lambda A^\dagger. \quad (3.38)$$

On the left hand side we recognize the operator  $R$  introduced before in Eq. (3.19) with projectors  $|y_j\rangle\langle y_j|$  replaced by POVM elements  $\Pi_j$ . Multiplying this equation by  $A$  from the right side we get

$$R\rho = \lambda\rho. \quad (3.39)$$

Lagrange multiplier  $\lambda$  is determined simply by taking trace and using the normalization condition, which yields  $\lambda = 1$ . This completes the alternative derivation of the extremal equation (3.35) for the maximum likelihood quantum-state estimate.

Except in the most simple cases it is impossible to solve Eq. (3.35) by analytical means. This is, of course, a consequence of its strong nonlinearity in  $\rho$ . To find the solution numerically one can use the above described EMU algorithm which combines the classical EM algorithm with a rotation of the eigenbasis of  $\rho$  in each step. Another possibility is to apply iterations directly to the extremal equation (3.35). Since for a general  $\rho$ , the left hand side of Eq. (3.35) is not even a hermitian operator it is first necessary to rewrite Eq. (3.35) to an explicitly positive semidefinite form. Notice that if Eq. (3.35) holds so will its adjoint  $\rho = \rho R$ . Using this in Eq. (3.35) we get

$$R\rho R = \rho, \quad (3.40)$$

where on both sides are now positive semidefinite operators. Starting from some unbiased density matrix such as the maximally mixed state, Eq. (3.40) can be used to find the first approximation, and this procedure can be repeated until the stationary point of transformation (3.40) is attained. Unfortunately, unlike for classical EM algorithm the convergence of this quantum algorithm is not guaranteed in general. However it is not difficult to prove its convergence for sufficiently small steps when the transformation  $R$  is “diluted” as follows  $R \rightarrow (I + \alpha R/2)/(1 + \alpha/2)$ , where  $I$  is the identity operator and  $\alpha$  is a small positive number. Considering now the  $i$ th step of the algorithm (3.40), the current density matrix  $\rho^i$  is transformed as follows:

$$\rho^{i+1} = (1 - \alpha)\rho^i + \alpha \frac{\rho^i R^i + R^i \rho^i}{2}. \quad (3.41)$$

Terms quadratic in  $\alpha$  have been neglected. Denoting

$$\rho' = \frac{\rho^i R^i + R^i \rho^i}{2}, \quad (3.42)$$

it is seen that  $(i + 1)$ th iteration  $\rho^{i+1}$  is a convex combination of  $\rho^i$  and  $\rho'$ . Loosely speaking it is obtained by moving a “distance”  $\alpha$  from the old density matrix along the line connecting it with a “density matrix”  $\rho'$ . Now we will show that  $\mathcal{L}(\rho^{i+1}) \geq \mathcal{L}(\rho^i)$ , which is the same as  $\ln \mathcal{L}(\rho^{i+1}) \geq \ln \mathcal{L}(\rho^i)$ . Evaluating log-likelihood at  $\rho^{i+1}$ , expanding it in  $\alpha$  and neglecting higher order terms we get

$$\ln \mathcal{L}(\rho^{i+1}) = \ln \mathcal{L}(\rho^i) + \alpha [\text{Tr}\{R^i \rho'\} - 1]. \quad (3.43)$$

It remains to show that the last term on the right hand side is positive:

$$\text{Tr}\{R^i \rho'\} = \text{Tr}\{R^i \rho^i R^i\} = \text{Tr}\{R^i \rho^i R^i\} \text{Tr}\{\rho^i\} \geq \text{Tr}^2\{R^i \rho^i\} = 1. \quad (3.44)$$

In the second to last step the Cauchy-Schwarz inequality has been used, and the last step follows from the definition of operator  $R$ . So we proved that a sufficiently small step always increases the likelihood. Since this functional

has no side minima (convexity), the algorithm converges monotonically to the global maximum.

All results discussed so far have to be modified in the case of incomplete detections. Provided that  $H \neq 1$  in Eq. (3.2), the closure relation may be always recovered in the form

$$\sum_j H^{-1/2} |y_j\rangle\langle y_j| H^{-1/2} = 1. \quad (3.45)$$

This corresponds to the renormalization of the true probabilities  $p_j = \langle y_j | \rho | y_j \rangle$  as follows:  $p_j \rightarrow p_j / \sum_i p_i$ . This formulation incorporates the case of incomplete detection. Notice, that the extremal equation (3.35) remains valid, provided we introduce renormalized operators

$$R \rightarrow R' = (H')^{-1/2} R (H')^{-1/2}, \quad \rho_e \rightarrow \rho'_e = (H')^{1/2} \rho_e (H')^{1/2},$$

where we defined  $H' = H / \sum_j p_j$ . All the conclusions derived for complete measurements may be extended to this case of incomplete measurement as well.

This formulation coincides with the ML estimation provided that the experiment is governed by Poissonian statistics. Assume that  $n_i$  samples the mean number of particles  $np_i$ , where  $p_i$  is as before the prediction of quantum theory for a detection in the  $i$ -th channel and  $n$  is an unknown mean number of particles. The relevant part of log-likelihood corresponding to the Poissonian statistics reads  $\ln \mathcal{L} \propto \sum_i n_i \ln(np_i) - n \sum_i p_i$ . The extremal equation for  $n$  can easily be formulated as the condition  $n = \sum_i n_i / \sum_i p_i$ . On inserting this estimate of the unknown mean number of Poissonian particles into the log-likelihood we reproduce the renormalized likelihood function.

### 3.3.6 Maximum likelihood as a statistical distance continued ...

Having shown several ways how to maximize likelihood we can now come back to the justification of the Kullback-Leibler divergence as a plausible statistical distance. Let us assume the existence of a quantum measure  $D(\mathbf{d}, \mathbf{p})$ , which parameterizes the distance between measured data and probabilities predicted by quantum theory. We will search for the state(s) located in the closest neighborhood of the data. Repeating steps leading to Eqs. (3.38) and (3.39) we get a new extremal equation,

$$\sum_j \frac{\partial D}{\partial p_j} \Pi_j \rho = \lambda \rho. \quad (3.46)$$

Of course, different measures  $D$  yield different extremal equations whose solutions are also different in general. Lagrange multiplier  $\lambda$  is determined from the normalization condition  $\text{Tr} \rho = 1$  as before,

$$\lambda = \sum_i \frac{\partial D}{\partial p_j} p_j.$$

As any composed function  $G[D(\mathbf{f}, \mathbf{p})]$  fulfills the same extremal equation (3.46) with the Lagrange multiplier rescaled as  $\lambda \frac{dG}{dD}$ , we will consider, without the loss of generality, the normalization condition  $\lambda = 1$ .

The extremal equation (3.46) has the form of a decomposition of the identity operator on the subspace, where the density matrix is defined by

$$\sum_i \frac{\partial D}{\partial p_j} \Pi_j = 1_\rho. \quad (3.47)$$

This resembles the definition of a POVM characterizing a new generalized measurement [31]. To link the above extremalization with quantum theory, let us postulate the natural condition for the quantum expectation values,

$$\text{Tr}\left(\frac{\partial D}{\partial p_j} \Pi_j \rho\right) = f_j. \quad (3.48)$$

This assumption seems to be reasonable. The synthesis of sequential non-compatible observations may be regarded as a new measurement scheme, namely the measurement of the quantum state. The quantum measure  $D$  then fulfills the differential equation

$$\frac{\partial D}{\partial p_j} p_j = f_j. \quad (3.49)$$

and singles out the solution in the form

$$D(\mathbf{f}, \mathbf{p}) = \sum_j f_j \ln p_j. \quad (3.50)$$

This is nothing else than the log-likelihood or Kullback–Leibler relative information [21]. Formal requirements of quantum theory, namely the interpretation of the extremal equation as a POVM, result in the concept of maximum likelihood in mathematical statistics. In this sense maximum likelihood estimation may be considered as a new quantum measurement.

### 3.3.7 Example: entangled photons

Let us illustrate ML quantum-state estimation on a simple example of a two-photon entangled state generated by the spontaneous downconversion source of White et al. [32]; for more examples see Badurek *et al.* in this volume. In their experiment, White et al. measured the nominal Bell state  $(|HH\rangle + |VV\rangle)/\sqrt{2}$  along sixteen distinct directions:  $\{|y_j\rangle\} = \{|HH\rangle, |HV\rangle, |VH\rangle, |VV\rangle, |HD\rangle, |HL\rangle, |DH\rangle, |RH\rangle, |DD\rangle, |RD\rangle, |RL\rangle, |DR\rangle, |DV\rangle, |RV\rangle, |VD\rangle, |VL\rangle\}$ .  $H$ ,  $V$ ,  $D$ ,  $R$ , and  $L$  being horizontal, vertical, diagonal,

right circular, and left circular polarization, respectively. Counted numbers of coincidences along these directions can be found in [32]. Let us use their experimental data to estimate the *true* state of entangled photons.

Due to various sources of errors, the true state is expected to differ from the nominal state. Notice that the chosen measurements are not complete, that is  $H = \sum_j |y_j\rangle\langle y_j| \neq \hat{1}$ . This has been taken into account, see the discussion after Eq. (3.45).

Starting from the maximally mixed state  $(|HH\rangle\langle HH| + |VV\rangle\langle VV| + |HV\rangle\langle HV| + |VH\rangle\langle VH|)/4$ , new eigenvalues and eigenvectors of density matrix are found using Eqs. (3.29) and (3.31). This has been repeated until a stationary point of the iteration process has been attained. The diagonal representation of the reconstructed density matrix reads

$$\rho_e^{\text{ML}} = 0.962 |\phi_1\rangle\langle\phi_1| + 0.038 |\phi_2\rangle\langle\phi_2|. \quad (3.51)$$

The other two eigenvalues are zero. The eigenvectors  $|\phi_1\rangle$  and  $|\phi_2\rangle$  are given in Tab. 3.1.

**Table 3.1.** Eigenvectors of the reconstructed density matrix.

	$ \phi_1\rangle$	$ \phi_2\rangle$
$ VV\rangle$	$0.696 - 0.027i$	$0.630 + 0.071i$
$ VH\rangle$	$-0.050 - 0.020i$	$-0.284 + 0.174i$
$ HV\rangle$	$-0.040 + 0.015i$	$-0.150 - 0.247i$
$ HH\rangle$	$0.712 - 0.062i$	$-0.634 - 0.035i$

The reconstructed density matrix (3.51) agrees well with the qualitative reasoning given in [32]. Namely, the reconstructed state is almost a pure state – a slightly rotated Bell state. The apparent incompatibility of the nominal state with the registered data was interpreted in [32] as the result of possible slight misalignments of the axes of analysis systems with respect to the axes of the downconversion source. This is, of course, reflected in the reconstructed state (3.51), which quantifies such misalignments and might serve for hunting down the errors, and calibrating the experimental setup. For such purposes the error analysis of the presented reconstruction technique becomes crucial. The results of numerical simulations suggest that the fidelity of the reconstructed state corresponding to tens of thousands of detections (like in [32]) is typically better than 0.99. More detailed discussion of estimation errors will be presented in the next section.

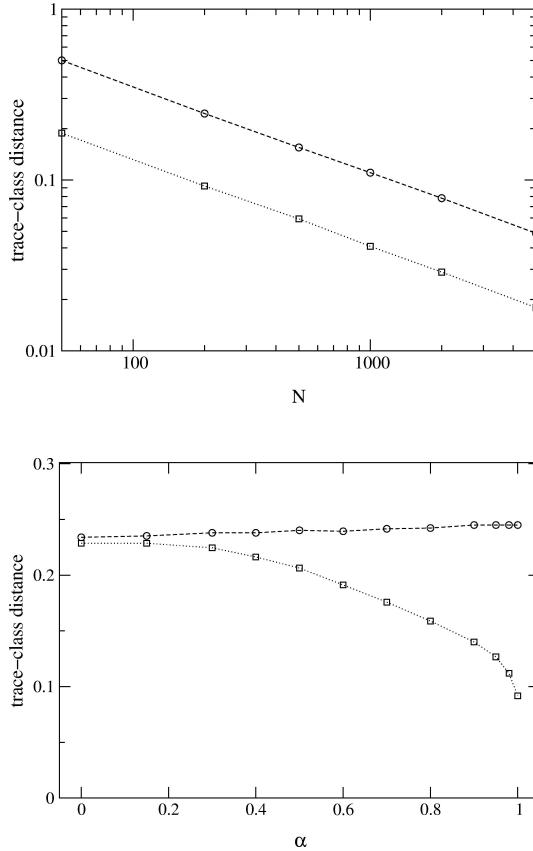
Notice also that the reconstructed density matrix (3.51) is semi-positive definite. This should be contrasted with the result of the standard reconstruction given in [32]. The deterministic inversion of Eq. (3.4) has the following diagonal representation:  $r_1 = 1.022$ ,  $r_2 = 0.068$ ,  $r_3 = -0.065$ ,  $r_4 = -0.024$ . The corresponding eigenvectors need not be specified here. Apparently, the

standard reconstruction is an unphysical, non-positive matrix. It is worth noticing that the negative eigenvalues are comparable in magnitude with the non-diagonal elements of  $\rho_e^{\text{ML}}$  in  $H$ - $V$  basis, see Tab. 3.1. This is a nice example of a situation when standard methods fail even though rather high number of particles (tens of thousands) has been registered. ML reconstruction provides always physically sound results. Moreover, it represents genuine quantum measurement of entangled state.

It is also interesting to see how large an improvement upon the standard direct inversion can be expected if more computationally demanding ML estimation is used. Let us illustrate this again on the simple example of two entangled qubits.

To compare two different estimators some measure of the quality of reconstruction is needed. A natural requirement is that the reconstructed state should come as close to the true one as possible. The word “close” here is however somewhat ambiguous. A number of different measures of distance could be used each leading to slightly different results. Let us here mention Jozsa fidelity (generalized overlap)  $F = \text{Tr}^2\{\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\}$  [33], relative entropy  $E_r = \text{Tr}\{\rho_1 \log \rho_2\}$ , and trace-class distance  $D = \text{Tr}\{|\rho_1 - \rho_2|\}/2$  to name a few. All these measures have good operational meaning. The fidelity is closely related to Wootters statistical distance [34] and also quantifies probability of success of the error-free state discrimination [35, 36].  $E_r$  is linked to the probability of mistaking  $\rho_1$  for  $\rho_2$  in optimum discrimination and large  $N$  limit [37]. The last distance  $D$  quantifies the reliability of the predictions based on the reconstructed state. It measures how much on the average the probabilities of particular outcomes of measurements predicted by  $\rho_1$  differ from the probabilities predicted by  $\rho_2$ .  $D$  has the advantage that unlike the fidelity and relative entropy it defines distance also between two non-positive operators. This is important because the direct inversion often yields unphysical results, as we have seen above, and we want to compare the results as they are without any tampering with them.

The results of two simulations are summarized in Fig. 2. In the upper panel, the errors of the standard and ML estimation are compared for a different number of detected photon pairs (measuring time). One of the four Bell states was chosen as the true state. Notice that the relative ML improvement upon the standard method remains constant over a wide range of intensities and numerically it is close to the Euler number  $e \approx 2.71$ . This is a significant improvement worth of technically more involved ML reconstruction procedure. Lower panel shows the dependence of estimation errors on the purity (or degree of entanglement) of the true state for a fixed intensity of the source of photon pairs. Apparently, ML estimation is significantly better for highly pure states while it is just a shade better for highly mixed states. The improvement increases with moving the true state closer to the boundary of the convex space of physically allowed states. This can be understood as follows: Under the given conditions the standard and ML reconstructions



**Fig. 3.1.** Errors of the ML (squares) and standard (circles) reconstructions in dependence on the (simulated) experimental situation. Upper panel: the source generates Bell states  $\bar{\rho} = (|HH\rangle + |VV\rangle)(\langle HH| + \langle VV|)/2$ ; its intensity  $N$  is the independent variable. Lower panel: the intensity is kept constant at  $N = 200$  photon pairs; the purity and the degree of entanglement are varied,  $\bar{\rho} = (1 + \alpha\sigma_1 \otimes \sigma_1 - \alpha\sigma_2 \otimes \sigma_2 + \alpha\sigma_3 \otimes \sigma_3)/4$ .

would coincide provided the standard reconstruction would fall within the physically allowed region. This is simply because for the 16 measured observables there is a one to one correspondence between the registered data and the (not necessarily positive) operator representing the source. Only when the standard procedure fails to yield a physical state its result differs from the ML estimate. In that case  $\mathcal{L}$  has maximum outside the closed set on which it should be maximized and we thus search for the maximum likely state on the boundary. Obviously the probability that this happens becomes larger for states lying closer to the boundary. Fig. 3.1 demonstrates that the ML reconstruction scheme is particularly efficient in that case.

Here one remark seems to be in order. The chosen system of two qubits is a rather simple one of dimension 4. For more complicated systems the boundary of the set of density matrices becomes a *very* complicated set of states and, loosely speaking, almost any state is located near some border. It is reassuring that ML scheme is capable of handling those cases well.

### 3.4 Estimation errors and Fisher information

In the previous section we have discussed several unique features of ML inference. Among them, the high efficiency of ML estimators is perhaps the most important one. Error of ML estimation is tightly connected to an interesting measure of information named after R. A. Fisher [38]. Let us first illustrate this useful concept on the simple example of the estimation of a single parameter  $\theta$ . Consider a result  $x$  of a measurement occurs with probability  $p(x|\theta)$ . Now suppose that  $y$  has been registered. From this result we are to estimate the true value of parameter  $\theta$ . Any function  $\theta(x)$  of data, called estimator, can be used for this purpose. If this experiment was repeated with the same setting of  $\theta$ , some other value might have been observed which would in general lead to a different estimated value of  $\theta$ . Of course different estimators can yield very different errors when applied to the same measured data. An important result by Cramér and Rao [39, 40] states that no matter in how clever way an estimator is designed, its error is always bounded from below by

$$(\Delta\theta)^2 \geq \frac{1}{F}, \quad (3.52)$$

where  $F$  is the Fisher information defined as

$$F = \left\langle \left( \frac{d}{d\theta} \ln p(x|\theta) \right)^2 \right\rangle_x, \quad (3.53)$$

and the angle brackets denote averaging over data. Inequality (3.52) is called the Cramér-Rao lower bound (CRLB) and its importance stems from the fact that it provides the ultimate resolution of estimation. Another important theorem due to Fisher [38] says that ML estimators attain this bound asymptotically for large amount of registered data (detected particles). So if many observations are collected, ML estimation is the most efficient estimation strategy.

Let us now evaluate the overall performance of ML quantum-state tomography with the help of CRLB. First we will chose a convenient parameterization of the quantum system under observation. Any density matrix can be decomposed in an orthonormal basis  $\{\Gamma_k\}$  of traceless hermitian operators defined on the Hilbert space of the system

$$\rho = 1/p + \sum_{k=1}^{p^2-1} a_k \Gamma_k, \quad (3.54)$$

where  $p$  is the dimension of the Hilbert space. For the system of  $n$  qubits, a natural basis is the one generated by Pauli spin matrices  $\Gamma_k = \sigma_{k_1} \otimes \dots \otimes \sigma_{k_n}/n$ ;  $k = k_1 + 2^1 k_2 + \dots + 2^{n-1} k_n$ ;  $k_1, \dots, k_n = 0, 1, 2, 3$ , though other choices are possible [41]. Generators  $\Gamma_k$  have the property of  $\text{Tr}\{\Gamma_i \Gamma_j\} = \delta_{ij}$ . Real parameters  $a_k$  are now the unknown parameters that are to be estimated from the measurement. As in the case of a single parameter, their errors are bounded by CRLB,

$$\text{Var}(a_k) \geq (\mathbf{F}^{-1})_{kk}, \quad (3.55)$$

where  $\mathbf{F}$  is a multidimensional generalization of the Fisher information,

$$F_{kl} = \left\langle \frac{\partial \log \mathcal{L}}{\partial a_k} \frac{\partial \log \mathcal{L}}{\partial a_l} \right\rangle_{\mathbf{n}}. \quad (3.56)$$

$\mathcal{L}$  is the joint probability of observing data  $\mathbf{n}$ . At the same time this function is the likelihood of the given vector  $\mathbf{a}$  parameterizing the quantum state.

Any measurement can be described by POVM elements  $\{\Pi_j\}$ , where index  $j$  labels independent output channels. Statistics of such a generalized measurement is multinomial, see Eq. (3.6). The relevant part of the total probability now reads,

$$\mathcal{L} = \prod_j (\text{Tr} \rho \Pi_j)^{n_j}. \quad (3.57)$$

Let us evaluate the Fisher information matrix in this most general case. Using (3.54) and (3.57) in definition (3.56), we get

$$F_{kl} = \sum_i \sum_j \frac{\text{Tr}\{\Gamma_k \Pi_i\} \text{Tr}\{\Gamma_l \Pi_j\}}{p_i p_j} \langle n_i n_j \rangle. \quad (3.58)$$

Averaging with respect to multinomial distribution with a total of  $N$  repetitions of the measurement yields  $\langle n_i n_j \rangle = N^2 p_i p_j - N p_i p_j + N p_i \delta_{ij}$ , where  $\delta_{ij}$  is the Kronecker symbol. Only the last term gives a nonzero contribution since  $d(\sum_j p_j)/da_k = 0$  due to the normalization to unity. The Fisher information matrix thus simplifies to

$$F_{kl} = N \sum_j \text{Tr}\{\Gamma_k \Pi_j\} \text{Tr}\{\Gamma_l \Pi_j\} / p_j. \quad (3.59)$$

Unfortunately the variances of  $a_k$  alone do not provide enough information needed for placing “error-bars” on the reconstructed state, nor do they suffice for determining how much the predictions based on it could be in error. This is due to the possible correlations of fluctuations of different  $a_k$  that might arise in the reconstruction procedure. To avoid such correlations let us decompose the state of the system in a new basis  $\Gamma'$  – such that the corresponding Fisher matrix becomes diagonal. This has a clear physical meaning: On registering counts  $\mathbf{n}$ , the probability  $P(\mathbf{n}|\mathbf{a})$  characterizes the likelihood  $\mathcal{L}$  of various states  $\mathbf{a}$ . In the asymptotical limit of a large amount of accumulated data

the likelihood can be approximate by Gaussian distribution. In terms of the Fisher matrix it reads:

$$\log \mathcal{L} = \log P(\mathbf{n}|\mathbf{a}) \approx - \sum_{kl} (a_k - \tilde{a}_k)(a_l - \tilde{a}_l) F_{kl}, \quad (3.60)$$

where  $\tilde{\mathbf{a}}$  specifies the maximum-likelihood solution. Let us define the error volume as the set of density matrices which have likelihoods that do not drop below a certain threshold,  $\mathcal{L} \geq \text{const}$ . According to Eq. (3.60) the error volume is an ellipsoid, whose axes lie in the directions of the diagonalizing basis  $\Gamma'$ , their lengths being given by the eigenvalues of the inverse of Fischer information matrix. Obviously, all the states inside the ellipsoid are still likely solution of our inverse problem. The rotation of the error ellipsoid is described by unitary transformation of the generators,

$$\Gamma'_k = U^\dagger \Gamma_k U = \sum_i u_i^k \Gamma_k. \quad (3.61)$$

Orthogonal  $(p^2 - 1)$  dimensional matrix of coefficients  $u_i^k$  will be denoted by  $\mathbf{U}$  to distinguish it from the operator  $U$ . In the new basis the state of the system reads,

$$\rho - 1/p = \sum_k a'_k \Gamma'_k = \sum_k a'_k U^\dagger \Gamma_k U = \sum_k \sum_i a'_k u_i^k \Gamma_k. \quad (3.62)$$

Similarly, the measurement can be decomposed in the  $\Gamma$  basis,

$$\Pi_j = \sum_k q_k^j \Gamma_k. \quad (3.63)$$

The transformation properties of Fisher information matrix are simple. Using Eq. (3.61) and (3.63) in Eq. 3.59, we find that under the rotation of basis  $F$  transforms as follows:

$$F'_{kl} = \sum_m \sum_n u_m^k \left( \sum_j \frac{q_m^j q_n^j}{p_j} \right) u_n^l. \quad (3.64)$$

The term in parenthesis can be recognized as the old Fisher information matrix, so finally we get

$$F' = \mathbf{U} F \mathbf{U}^T. \quad (3.65)$$

As seen, the transition from the old basis to the new one coinciding with the axes of the error ellipsoid ( $F'$  becomes diagonal) is provided by matrix  $\mathbf{U}$  composed from the eigenvectors of the Fisher matrix. They depend on the measurement that has been done and on the true probabilities which can be (approximatively) calculated from the maximum-likely state. The fluctuations of  $a'_k$ , now being independent, then can be propagated to all the quantities of interest using the standard methods of error analysis.

From this point of view, the synthesis of all the quantum observations is equivalent to the registration of the orthogonal observables  $\Gamma'$  defining those axes. In this new representation the Fisher matrix attains the diagonal form, which means that the estimates of the transformed quantum-state “coordinates”  $\tilde{a}'_k = \text{Tr}\{\bar{\rho}\Gamma'\}$  fluctuate independently. This hints at the possibility to form a single number quantifying the performance of the reconstruction scheme as a whole by adding together those independent errors:

$$\epsilon = \sum_k \langle (\Delta a'_k)^2 \rangle = \sum_k \langle (\Delta a_k)^2 \rangle \geq \text{Tr}\{F^{-1}\}. \quad (3.66)$$

Notice that this number does not depend on the chosen operator basis. Error  $\epsilon$  has an interesting geometrical interpretation: Hilbert-Schmidt distance

$$d(\rho, \bar{\rho}) = \text{Tr}\{(\rho - \bar{\rho})^2\} \quad (3.67)$$

is a natural metric that can be defined on the space of hermitian operators. Let us evaluate the Hilbert-Schmidt distance between the true state  $\bar{\rho}$  and its estimate  $\rho$ . Decomposing both states as  $\bar{\rho} = \sum_k \bar{a}_k \Gamma_k$ , and  $\rho = \sum_k a_k \Gamma_k$ , we get

$$d(\rho, \bar{\rho}) = \sum_k (a_k - \bar{a}_k)^2. \quad (3.68)$$

The mean distance (error) is then given by averaging  $d$  over many repetitions of the estimation procedure, each yielding slightly different estimates of the true state,

$$\langle d \rangle = \sum_k \langle (\Delta a_k)^2 \rangle. \quad (3.69)$$

But according to Eq. (3.66) this quantity is bounded from below by the trace of the inverse Fisher matrix,

$$\langle d \rangle \geq \text{Tr}\{F^{-1}\}. \quad (3.70)$$

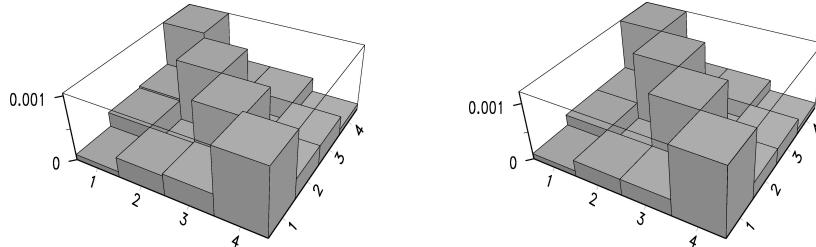
Thus the mean error of the optimal estimation from the measurement of the chosen set of observables is given by the trace of the inverse of the Fisher matrix. In particular,  $\text{Tr}\{F^{-1}\}$  quantifies the performance of ML estimation schemes per particle in the limit of a large amount of collected data  $N \gg 1$ . Its reciprocal then quantifies information about the unknown quantum state acquired in the measurement.

Also notice that since each new observed quantum system contributes equally to the Fisher information, see Eq. (3.59), it is possible to define total information about the quantum state gained per particle as follows:

$$I = \frac{1}{N \text{Tr}\{F^{-1}\}} \quad (3.71)$$

This quantity measures the quality of a given reconstruction scheme. One can use it for comparing performances of different sets of tomographic measurements or for investigating their invariance properties.

As an example, let us compare errors calculated from the Fisher information with actual errors of the quantum estimation of a state of two entangled qubits. The chosen set of 16 projections is the same as that discussed in the previous section; for details see Ref. [32]. The results for a slightly mixed entangled true state and observation of 2000 pairs of photons are shown in Fig. 3.2. Right panel shows variances of the estimated elements of the density matrix in  $|HH\rangle$ ,  $|HV\rangle$ ,  $|VH\rangle$  and  $|VV\rangle$  basis predicted by CRLB. Left panel shows actual errors that were obtained by averaging over many repeated estimations. Notice that the reconstructed values of the diagonal elements are



**Fig. 3.2.** Fig. 1. Squared errors (variances) of the elements of the reconstructed density matrix. Left panel: simulation; right panel: CRLB.  $\bar{\rho} = (1 + 0.8\sigma_1 \otimes \sigma_1 - 0.8\sigma_2 \otimes \sigma_2 + 0.8\sigma_3 \otimes \sigma_3)/4$ ;  $N = 2000$  pairs. We note that the diagonal is situated horizontally.

much more reliable than the rest. This is perhaps caused by the choice of the set of measured observables since the four basis vectors can be found among them. It is also clear that Fisher information in this case gives reliable error estimates. In a completely analogous fashion one could use Fisher information to derive error estimates for more complicated inverse problems such as the estimation of quantum operations or measurement that will be described in the following.

### 3.5 Estimation of quantum processes

The quantum-state tomography discussed in the preceding sections provides a precise quantitative method of characterization of sources of quantum states. However, the preparation of quantum states is often only the initial part of experiments. Typically, one would like to manipulate the quantum states by applying some unitary transformations, or transmit the quantum states via some channel to a distant location. The holy grail of the quantum information science is to construct a quantum computer which would provide an ultimate control over the evolution of quantum states. In practice, errors will unavoidably occur, stemming from decoherence, losses, etc. Thus, the quantum state  $\rho$  will typically become entangled with the environment

and the evolution of  $\rho$  cannot be described as a unitary transformation of  $\rho$ . Therefore, a more general theory of transformations of quantum states must be invoked.

A very general framework is provided by the formalism of quantum operations (also called quantum processes or quantum channels) [42]. Assuming that the environment interacting with  $\rho$  is not initially entangled or correlated with the device that prepares the state  $\rho$ , the most general transformation of a quantum state  $\rho$  allowed by the laws of the quantum mechanics is the so-called linear completely positive (CP) map  $\mathcal{E}$ . The condition that the environment and the state preparation device are not correlated is satisfied in the vast majority of the experiments and the formalism of quantum operations is thus applicable in practically all cases.

Given current interest in the quantum-information processing it is thus of paramount importance to develop tools for characterization of quantum processes. The concept of quantum process tomography has been introduced independently by Nielsen and Chuang [43] and by Poyatos *et al.* [44]. The reconstruction of quantum process has many practical applications, ranging from the probing of quantum communication channels to the evaluation of the performance of quantum gates and eventually debugging of quantum computers [43–50]. Several experimental demonstrations of the quantum process tomography in NMR [51, 52] and quantum optics systems [53–55] have been reported recently.

Most proposed quantum-process reconstruction techniques are based on direct linear inversion of experimental data. These methods are conceptually simple but may yield unphysical results. On the other hand, all necessary properties of the deterministic quantum transformations, namely the complete positivity and trace preservation can be naturally incorporated within the maximum-likelihood approach as the appropriate constraints [56, 57]. In what follows we shall show that the ML estimation of quantum process can be formally formulated in a very similar way as the ML estimation of quantum states.

The discussion is greatly facilitated by the Jamiolkowski isomorphism [58] that associates a positive-semidefinite operator  $E$  with a CP map  $\mathcal{E}$ . This formalism is briefly reviewed below. The main formal difference between estimation of processes and states lies in a higher number of constraints involved in the former case. The trace-preservation condition gives rise to altogether  $d^2$  real constraints, where  $d$  is the dimension of the Hilbert space  $\mathcal{H}$  of the input states. Rigorous ML estimation should take all these constraints into account properly. We shall demonstrate how this can be accomplished in practice and how the ML estimate of  $E$  can be efficiently numerically computed from the acquired experimental data. We will also touch upon several topics of current interest, such as the probing of quantum processes by entangled states and the reconstruction of trace-decreasing CP maps, which describe probabilistic (conditional) quantum operations.

### 3.5.1 Jamiolkowski isomorphism

The deterministic quantum operation  $\mathcal{E}$  must satisfy the following conditions (see [42] for an excellent detailed discussion):

- (i) *Positivity.* If  $\rho_{\text{in}} \geq 0$  then  $\rho_{\text{out}} \equiv \mathcal{E}(\rho_{\text{in}}) \geq 0$ , the map transforms density operators onto density operators.
- (ii) *Complete positivity.* The positivity is not a sufficient condition for  $\mathcal{E}$  to be a physical map. The reason is that the input state  $\rho_{\text{in}}$  may be a part of a maximally entangled state  $|\psi\rangle_{AB}$ . Let  $\mathcal{H}$  and  $\mathcal{K}$  denote the Hilbert spaces of input and output states, respectively, and denote  $d = \dim \mathcal{H}$ . The map  $\mathcal{E}$  must preserve positive semidefiniteness of  $\rho_{AB}$  when applied to one part of  $\rho_{AB}$ . We say that  $\mathcal{E}$  is completely positive iff the extended map  $\mathcal{M} = \mathcal{I}_{h,A} \otimes \mathcal{E}_B$  is positive for all  $h$ , where  $h$  is a dimension of the auxiliary Hilbert space  $\mathcal{H}'_A$ ,  $\mathcal{I}$  denotes an identity operation and  $\mathcal{M}$  acts on operators on the Hilbert space  $\mathcal{H}'_A \otimes \mathcal{H}_B$ . In fact, it suffices to take  $h = d$ .
- (iii) *Trace preservation.* The deterministic maps must preserve the total probability, hence  $\text{Tr}(\rho_{\text{out}}) = \text{Tr}(\rho_{\text{in}})$  must hold for all  $\rho_{\text{in}}$ .

There are several ways how the CP map can be described mathematically. Physically, every linear trace-preserving CP map can be realized as a unitary operation  $U_E$  on an extended Hilbert space  $\mathcal{H} \otimes \mathcal{A}$  of the system and ancilla, where the ancilla is initially prepared in a blank pure state  $|0\rangle_A$ ,

$$\rho_{\text{out}} = \text{Tr}_A[U_E \rho_{\text{in}} \otimes |0\rangle_A \langle 0| U_E^\dagger]. \quad (3.72)$$

There is a big freedom in the choice of  $U_E$  so this description of  $\mathcal{E}$  is not very suitable for our purposes.

Another widely used formalism is the Kraus decomposition which states that every CP map can be expressed in terms of a set of operators  $A_j$  as follows,

$$\rho_{\text{out}} = \sum_j A_j \rho_{\text{in}} A_j^\dagger.$$

The trace preservation condition amounts to the constraint  $\sum_j A_j^\dagger A_j = I$ . The Kraus decomposition is particularly useful when studying quantum noise and quantum error correction. However, the decomposition is not unique and infinitely many sets of operators  $\{A_j\}$  can describe the same map  $\mathcal{E}$ .

The mathematical representation of CP maps that we shall employ relies on the isomorphism between linear CP maps  $\mathcal{E}$  from operators on the Hilbert space  $\mathcal{H}$  to operators on the Hilbert space  $\mathcal{K}$  and positive semidefinite operators  $E$  on the Hilbert space  $\mathcal{H} \otimes \mathcal{K}$ . This representation is sometimes referred to as Jamiolkowski isomorphism. Its main advantage is that it eliminates all the free parameters and provides a compact description of the CP map  $\mathcal{E}$ .

Let us investigate what happens if the map  $\mathcal{E}$  is applied to one part of the maximally entangled state on  $\mathcal{H} \otimes \mathcal{H}$ . Let us define

$$|\psi\rangle_{AB} = \sum_{j=1}^d |j\rangle_A |j\rangle_B \quad (3.73)$$

and consider the positive semidefinite operator  $E$  on  $\mathcal{H} \otimes \mathcal{K}$ ,

$$E = \mathcal{I}_A \otimes \mathcal{E}_B(|\psi\rangle_{AB}\langle\psi|). \quad (3.74)$$

By definition,  $E$  is positive semidefinite,  $E \geq 0$ . One immediately finds that

$$E = \sum_{j,k} |j\rangle\langle k| \otimes \mathcal{E}(|j\rangle\langle k|). \quad (3.75)$$

It is an easy exercise, which we leave for the reader, to verify that the input-output transformation can be expressed in terms of the operator  $E$  as

$$\rho_{\text{out}} = \text{Tr}_{\mathcal{H}}[E \rho_{\text{in}}^T \otimes I], \quad (3.76)$$

where  $T$  denotes transposition in the basis  $|j\rangle$  and  $\text{Tr}_{\mathcal{H}}$  denotes the partial trace over the input Hilbert space. The deterministic quantum transformations preserve the trace of the transformed operators,  $\text{Tr}_{\mathcal{K}}[\rho_{\text{out}}] = \text{Tr}_{\mathcal{H}}[\rho_{\text{in}}]$ . Since this must hold for any  $\rho_{\text{in}}$  the operator  $E$  must satisfy the condition

$$\text{Tr}_{\mathcal{K}}[E] = I_{\mathcal{H}}, \quad (3.77)$$

where  $I_{\mathcal{H}}$  is an identity operator on space  $\mathcal{H}$ . The condition (3.77) effectively represents  $(\dim \mathcal{H})^2$  real constraints as advertised earlier.

### 3.5.2 Reconstruction of trace-preserving CP map

Having established the mathematical formalism, we can now proceed to the reconstruction of the CP map. The experimental setup that we have in mind is as follows. Some sources prepare various input states  $\rho_m$  that are used for the determination of the quantum process. Here we assume that we have full knowledge of  $\rho_m$ . Later on, we will remove this assumption and will analyze a more complex scenario when one simultaneously reconstructs the states and the process. The states  $\rho_m$  are sent through the quantum channel  $\mathcal{E}$ . Measurements described by POVMs  $\Pi_{ml}$  are carried out on each corresponding output state  $\mathcal{E}(\rho_m)$ . Let  $f_{ml}$  denote the relative frequency of detection of the POVM element  $\Pi_{ml}$ . These frequencies approximate the theoretical detection probabilities

$$p_{ml} = \text{Tr}[\mathcal{E}(\rho_m)\Pi_{ml}] = \text{Tr}[E \rho_m^T \otimes \Pi_{ml}], \quad (3.78)$$

where we used Eq. (3.76). The quantum process  $E$  should be reconstructed from the knowledge of the probe states  $\rho_m$  and the measured frequencies  $f_{ml}$ .

With the help of the Jamiołkowski isomorphism we may formulate the exact maximum-likelihood principle for estimated CP map  $E$  in a very simple and transparent form. The estimated operator  $E$  should maximize the constrained log-likelihood functional [56, 57]

$$\mathcal{L}_c[f_{ml}, p_{ml}(E)] = \sum_{m,l} f_{ml} \ln p_{ml} - \text{Tr}[\Lambda E], \quad (3.79)$$

where  $\Lambda = \lambda \otimes I_{\mathcal{K}}$  and  $\lambda$  is the (hermitian) matrix of Lagrange multipliers that account for the trace-preservation condition (3.77). The extremal equations for  $E$  can be obtained by varying functional (3.79) with respect to  $E$ , similarly to the case of quantum state estimation. This leads to the extremal equation

$$E = \Lambda^{-1} K E, \quad (3.80)$$

where the operator  $K$  reads

$$K = \sum_{m,l} \frac{f_{ml}}{p_{ml}} \rho_m^T \otimes \Pi_{ml}. \quad (3.81)$$

Further we have from Eq. (3.80) and Hermicity that  $E = E K \Lambda^{-1}$ . When we insert this expression in the right-hand side of Eq. (3.80), we finally arrive at symmetrical expression suitable for iterations,

$$E = \Lambda^{-1} K E K \Lambda^{-1}. \quad (3.82)$$

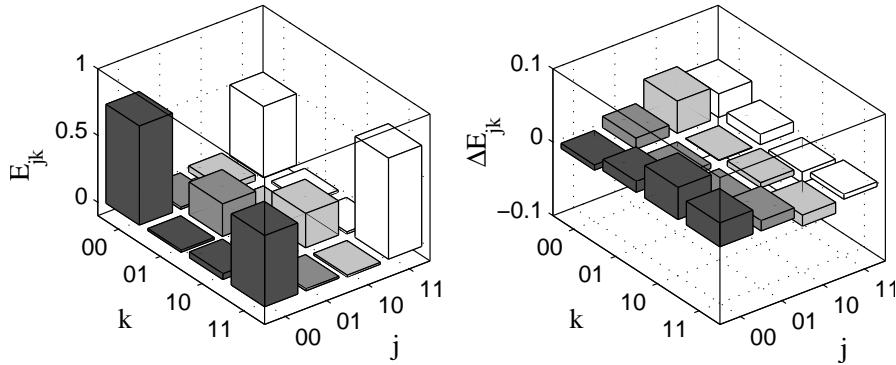
The Lagrange multiplier  $\lambda$  must be determined from the constraint (3.77). On tracing Eq. (3.82) over space  $\mathcal{K}$  we obtain quadratic equation for  $\lambda$  which may be solved as

$$\lambda = (\text{Tr}_{\mathcal{K}}[KEK])^{1/2}, \quad (3.83)$$

where we take the positive root,  $\lambda > 0$ . The system of coupled Eqs. (3.82) and (3.83) may be conveniently solved numerically by means of repeated iterations, starting from some unbiased CP map, for example  $E^{(0)} = I_{\mathcal{H} \otimes \mathcal{K}} / (\dim \mathcal{K})$ . It is important to note that Eq. (3.82) preserves the positive semidefiniteness of  $E$  and also the constraint  $\text{Tr}_{\mathcal{K}}[E] = I_{\mathcal{H}}$  is satisfied at each iteration step.

The feasibility of this reconstruction technique has been confirmed by extensive numerical simulations for various single- and two-qubit CP maps  $\mathcal{E}$ , various sets of probe density matrices  $\rho_m$  and POVMs  $\Pi_{ml}$ . As an illustrative example, we describe the reconstruction of a single-qubit trace preserving CP map. The operator  $E$  is characterized by 16 real parameters. The trace-preservation condition yields four constraints, which leaves 12 independent parameters. The maximum-likelihood estimation then amounts to finding the maximum of the log-likelihood functional in a 12-dimensional space, with highly non-trivial boundary defined by  $E \geq 0$ .

In the numerical simulations, we have considered six different probe states — the eigenstates of three Pauli matrices  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_z$ .  $3N$  copies of each input state are used. On each corresponding output state, a spin projection along axes  $x$ ,  $y$  and  $z$  is measured  $N$  times. The detected frequencies were generated by means of Monte Carlo simulations. The maximum likelihood



**Fig. 3.3.** Reconstruction of a depolarizing channel (3.84) with  $\eta = 0.5$ .  $3N = 300$  copies of each of six input probe states have been used – see text for more details. The left panel shows the matrix elements of the reconstructed operator  $E$  and the right panel displays the difference between reconstructed and exact operators  $\Delta E = E_{\text{est}} - E_{\text{true}}$ . Only real parts of complex elements are shown.

estimate of the true process has been obtained by iteratively solving the nonlinear extremal equations (3.82) and (3.83).

As the first example, consider the partially depolarizing channel (DC)

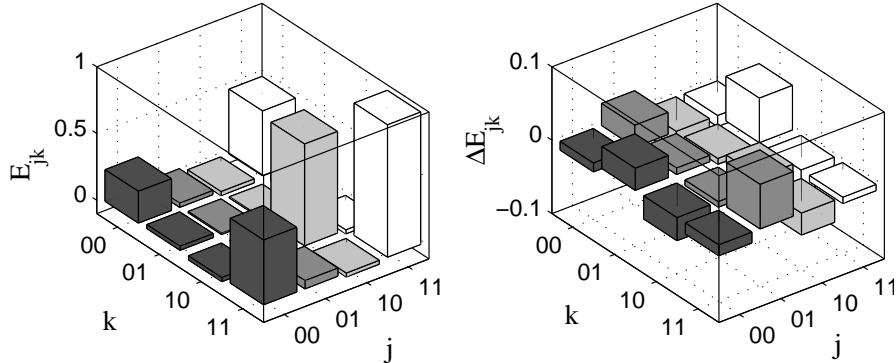
$$E_{DC} = \eta \mathcal{I} + (1 - \eta) \mathcal{O}, \quad (3.84)$$

where  $\mathcal{I}$  is the identity operation while  $\mathcal{O}$  denotes the totally depolarizing channel that maps all density matrices onto maximally mixed state,  $\mathcal{O}(\rho) = I/2$ . The results of the simulated quantum process tomography are shown in Fig. 3.3. For observations on  $N = 100$  copies of each probe state, the reconstruction works very well and the estimated process is typically very close to the true one, as confirmed by the small values of the estimation errors displayed in Fig. 3.3.

Our second example is the amplitude damping channel (ADC) that describes a decay process. This channel naturally arises when the qubit is represented by two levels of an atom, where  $|1\rangle$  is the ground state while  $|0\rangle$  is the excited state. Even in vacuum, the atom may emit a photon and decay to the ground state due to the coupling of the atom to the vacuum fluctuations of the electromagnetic field. If the spontaneous emission occurs with probability  $\eta^2$ , then in the basis  $|00\rangle, |10\rangle, |01\rangle, |11\rangle$  the ADC is described by the operator

$$E_{ADC} = \begin{pmatrix} \eta^2 & 0 & 0 & \eta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 - \eta^2 & 0 \\ \eta & 0 & 0 & 1 \end{pmatrix}. \quad (3.85)$$

An example of the simulated tomography of the amplitude damping channel is given in Fig. 3.4. Similarly as in the case of the depolarizing channel, the



**Fig. 3.4.** Reconstruction of an amplitude damping channel (3.85) with  $\eta = 0.5$ . Similarly as in Fig. 3.3, only real parts of complex elements are shown.

maximum-likelihood estimate is in good agreement with the true process. An important point that should be emphasized is that in both examples, the estimated process is really a trace preserving completely positive map and all the constraints imposed by quantum mechanics are satisfied. In particular,  $E \geq 0$  and  $\text{Tr}_{\mathcal{K}}[E] = I_{\mathcal{H}}$  holds. As discussed in previous sections, the maximum-likelihood estimation is asymptotically (for large number of probes  $N$ ) the optimal estimation method because it saturates the Cramér-Rao bound. All these facts illustrate the advantages of this approach in comparison to simpler data processing techniques.

### 3.5.3 Entangled probes

The above discussed reconstruction of a single-qubit CP map bears striking resemblance to the estimation of the entangled-two qubit state which was discussed in Section 3.3. This similarity is direct consequence of the Jamiołkowski isomorphism. Following the definition (3.74), the operator  $E$  representing the CP map can be in fact prepared physically in the laboratory if we first prepare a maximally entangled state on the Hilbert space  $\mathcal{H} \otimes \mathcal{H}$  and then apply the CP map to one part of this entangled state. In this way, the quantum-process tomography can be transformed to the quantum-state tomography. More generally, this suggests that it may be useful to employ entangled quantum states as probes of an unknown quantum process [48, 49].

Let  $\rho_{m,AB}$  denote the entangled state on the Hilbert space  $\mathcal{H}_A \otimes \mathcal{H}_B$  that serves as a probe of the CP map  $\mathcal{E}$  that is applied to the subsystem  $A$ . A joint generalized measurement described by the POVMs  $\Pi_{ml}$  if performed on the output Hilbert space  $\mathcal{K} \otimes \mathcal{H}_B$ . The log-likelihood functional still has the form (3.79), only the formula for the probability  $p_{ml}$  changes to

$$p_{ml} = \text{Tr}_{\mathcal{H}_A \mathcal{H}_B \mathcal{K}}[(E \otimes I_{\mathcal{H}_B})(\rho_{m,AB}^{T_A} \otimes I_{\mathcal{K}})(I_{\mathcal{H}_A} \otimes \Pi_{ml})], \quad (3.86)$$

where  $T_A$  stands for the partial transposition in the subsystem  $A$ . Consequently, the operator  $K$  appearing in the extremal Eqs. (3.82) and (3.83) must be calculated as follows,

$$K = \sum_{l,m} \frac{f_{ml}}{p_{ml}} \text{Tr}_{\mathcal{H}_B} [(\rho_{m,AB}^{T_A} \otimes I_{\mathcal{K}})(I_{\mathcal{H}_A} \otimes \Pi_{ml})]. \quad (3.87)$$

Apart from these modifications of  $p_{ml}$  and  $K$  one can proceed as before and solve the Eqs. (3.82) and (3.83) by means of repeated iterations.

### 3.5.4 Probabilistic operations

The transformations of the quantum states may be probabilistic. Such a conditional map  $\mathcal{E}$  succeeds with probability

$$q_m \equiv \text{Tr}[\mathcal{E}(\rho_m)] = \text{Tr}[E \rho_m^T \otimes I] \quad (3.88)$$

and fails with probability  $1 - q_m$ . Since  $q_m \leq 1$ , the inequality  $\text{Tr}_{\mathcal{K}}[E] \leq I_{\mathcal{H}}$  must hold. Note that the re-normalized output state  $\rho_{m,\text{out}} = \mathcal{E}(\rho_m)/q_m$  is a nonlinear function of  $\rho_m$ . Probabilistic transformations arise in many areas of quantum information processing and quantum state manipulation. We can mention the conditional generation of quantum states [59–61], entanglement distillation protocols [62, 63], and the probabilistic scheme for quantum computing with linear optics and single photons proposed by Knill, Laflamme and Milburn [64]. Three alternative ways for manipulating CP decreasing maps will be presented.

The tomography of probabilistic CP maps can be in fact re-formulated as a tomography of a trace preserving CP map provided that we know in each case whether the probabilistic map succeeded or failed. The observation of a failure is a valid measurement outcome which can be associated with a POVM element  $\Pi_{\emptyset}$ . Without loss of generality, we may assume that the total output Hilbert space  $\mathcal{K}_{\text{tot}}$  is a direct sum of the Hilbert space  $\mathcal{K}$  and a one-dimensional space  $\mathcal{K}_{\text{fail}}$  spanned by  $|\emptyset\rangle$ ,  $\mathcal{K}_{\text{tot}} = \mathcal{K} \oplus \mathcal{K}_{\text{fail}}$ . Thus whenever the operation fails, we assume that the output state is  $|\emptyset\rangle$ . The state  $|\emptyset\rangle$  serves as a sink for all unsuccessful trials and  $\Pi_{\emptyset} = |\emptyset\rangle\langle\emptyset|$ . The POVM  $\{\Pi_{ml}\}_{l=1}^N$  that describes measurement of the output states in the Hilbert space  $\mathcal{K}$  is completed by adding the element  $\Pi_{\emptyset}$  so that the new POVM satisfies the closure relation on  $\mathcal{K}_{\text{tot}}$ .

Instead of the trace-decreasing map  $\mathcal{E}$  we shall reconstruct the extended trace-preserving map  $\tilde{\mathcal{E}}$  that maps operators on  $\mathcal{H}$  onto operators on  $\mathcal{K}_{\text{tot}}$ . Let  $f_{m\emptyset}$  denote the number of observed failures of the application of the map  $\mathcal{E}$  to the probe state  $\rho_m$ . The constrained log-likelihood functional can be obtained as a simple extension of (3.79),

$$\mathcal{L}_c[\tilde{E}] = \sum_{m,l} f_{ml} \ln p_{ml} + \sum_m f_{m\emptyset} \ln p_{m\emptyset} - \text{Tr}[A \tilde{E}], \quad (3.89)$$

where  $p_{ml} = \text{Tr}[\tilde{E}\rho_m^T \otimes \Pi_{ml}]$  and  $p_{m\emptyset} = \text{Tr}[\tilde{E}\rho_m^T \otimes \Pi_\emptyset]$ . The trace-preserving CP map  $\tilde{\mathcal{E}}$  can be reconstructed with the use of the iterative algorithm described in the Section 3.5.2. From the estimated operator  $\tilde{E}$  we can extract the sought after operator  $E$  by projecting  $\tilde{E}$  onto the subspace  $\mathcal{H} \otimes \mathcal{K}$ .

In many experiments, however, we cannot determine how often the conditional map succeeded. This problem typically arises in the experiments with photon pairs generated by means of spontaneous parametric downconversion. This process is random and we do not know whether the photon pair was generated until we detect it. In certain experimental setups, this prevents us from measuring the number of unsuccessful events  $f_{m\emptyset}$ . Without knowing  $f_{m\emptyset}$ , we cannot use the trick with extension to the trace-preserving map and we must try to estimate directly the trace-decreasing map  $\mathcal{E}$ .

Consequently, the constraint (3.77) should be replaced with  $\text{Tr}_{\mathcal{K}} E \leq I_{\mathcal{H}}$ . If we do not know the frequency of failures,  $E$  can be determined only up to an overall normalization prefactor and the constraint  $\text{Tr}_{\mathcal{K}} E \leq I_{\mathcal{H}}$  is thus irrelevant. In the log-likelihood (3.79) we must replace  $p_{ml}$  with the re-normalized probabilities  $p_{ml}/q_m$ . We thus have to maximize

$$\mathcal{L} = \sum_{l,m} f_{ml} (\ln p_{ml} - \ln q_m), \quad (3.90)$$

under the constraint  $E \geq 0$ .

The extremal equation for  $E$  can be derived by introducing the decomposition  $E = A^\dagger A$  and varying (3.90) with respect to  $A$ . We obtain  $RE = SE$ , where

$$S = \sum_{m,l} \frac{f_{ml}}{q_m} \rho_m^T \otimes I. \quad (3.91)$$

This operator replaces the Lagrange multiplier  $A$ . Since  $S$  is invertible, we can express  $E = S^{-1}RE$  and symmetrize the extremal equation which yields

$$E = a^{-1} S^{-1} R E R S^{-1}, \quad (3.92)$$

where  $a = \text{Tr}[S^{-1}R E R S^{-1}]$  is a normalization factor chosen such that  $\text{Tr}[E] = 1$ .

Alternatively, the ML estimate of  $E$  can be determined by the analogue of the EMU algorithm. This approach can be applied here because we are not bound by the constraints (3.77) and we can thus consider independently the variations of the eigenstates  $r_j^2$  and eigenvalues  $|e_j\rangle$  of  $E = \sum_j r_j^2 |e_j\rangle\langle e_j|$ . Assume small variations  $r_j \rightarrow r_j + \delta r_j$  and  $|e_j\rangle \rightarrow \exp(i\epsilon H)|e_j\rangle$ , where  $H$  is a Hermitean operator. The corresponding variation of  $\mathcal{L}$  is given by

$$\delta\mathcal{L} = 2 \sum_j r_j \delta r_j \sum_{l,m} \langle e_j | X_{ml} | e_j \rangle + i\epsilon \text{Tr} \left( H \sum_{m,l} [E, X_{ml}] \right) \quad (3.93)$$

where

$$X_{ml} = \frac{f_{ml}}{p_{ml}} \rho_m^T \otimes \Pi_{ml} - \frac{f_{ml}}{q_m} \rho_m^T \otimes I. \quad (3.94)$$

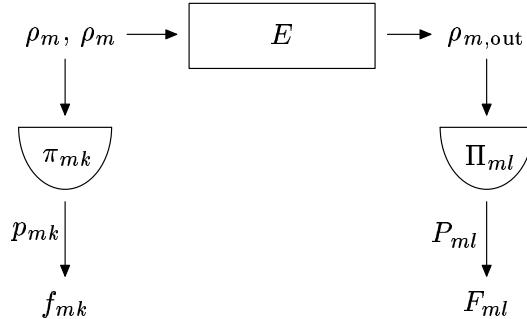
From Eq. (3.93) we can deduce  $\delta r_j$  and  $H$  that will increase the log-likelihood,

$$r_j \rightarrow r_j + \eta r_j \sum_{m,l} \langle e_j | X_{ml} | e_j \rangle, \quad |e_j\rangle \rightarrow \exp(-\epsilon \sum_{m,l} [E, X_{ml}]) |e_j\rangle. \quad (3.95)$$

At each iteration step,  $\eta$  and  $\epsilon$  can be further optimized in order to achieve maximum increase of  $\mathcal{L}$ . Moreover, after each iteration, we can re-normalize  $E \rightarrow E/\text{Tr}[E]$ . Note, that this does not change the value of  $\mathcal{L}$ .

### 3.5.5 Unknown probes

Up to now quantum states and processes have been treated independently. Widely accepted strategy of how to approach a complex problem is to specify some partial subproblems, address them separately and merge the solutions. This technique usually gives a good answer in the technical sense. Though this is possible even in quantum theory, there are no fundamental reasons for such a factorization. To consider the full problem without splitting it into isolated subproblems is technically more advanced but could be advantageous. This strategy can be demonstrated on the synthesis of the problems treated separately in the previous paragraphs. Let us assume the estimation of the generic quantum process (CP map)  $E$  with the help of a set of probe states  $\rho_m$ , identity of which is also unknown [57]. What is only known to the experimenters are the output of certain measurements performed on the ensemble of probe states and on the ensemble of transformed probe states. All the physically relevant results can be derived exclusively from the acquired data, where input states and their transformation are inseparably involved.



**Fig. 3.5.** Scheme of setup for the generalized measurement of quantum process using unknown quantum states as probes.

In accordance with the theory presented above let us consider the set of probe states  $\rho_m$  on the space  $\mathcal{H}$ . By means of unknown quantum process  $E$

these states are transformed onto output states  $\rho_{m,\text{out}}$  in the space  $\mathcal{K}$ . The observation must be more complex now involving the detection on the ensemble of both the input and the output states. For this purpose the corresponding POVM elements will be denoted by  $\pi_{mk}$  and  $\Pi_{ml}$ . The diagram involving detected signals and measurements is shown in Fig. 3.5. Let  $f_{mk}$  denote the relative frequency of detection of the POVM element  $\pi_{mk}$  in the input space  $\mathcal{H}$  and  $F_{ml}$  denote the relative frequency of detection of the POVM element  $\Pi_{ml}$  in the output space  $\mathcal{K}$ . The frequencies  $f_{mk}$ ,  $\sum_k f_{mk} = 1$ , and  $F_{ml}$ ,  $\sum_l F_{ml} = 1$ , approximate the true probabilities  $p_{mk}$  and  $P_{ml}$  of individual outcomes, respectively,

$$p_{mk} = \text{Tr}_{\mathcal{H}}[\rho_m \pi_{ml}], \quad P_{ml} = \text{Tr}_{\mathcal{K}}[\rho_{m,\text{out}} \Pi_{ml}] = \text{Tr}[E(\rho_m^T \otimes \Pi_{ml})], \quad (3.96)$$

where the relation (3.76) was used. The estimated process  $E$  and probe states  $\rho_m$  should maximize the constrained log-likelihood functional

$$\mathcal{L} = \sum_{m,k} f_{mk} \ln p_{mk} + \sum_{m,l} F_{ml} \ln P_{ml} - \sum_m \mu_m \text{Tr}[\rho_m] - \text{Tr}[\Lambda E]. \quad (3.97)$$

The additivity of log likelihood reflects the independence of observations performed on the input and output states. The Lagrange multipliers  $\mu_m$  and  $\Lambda = \lambda \otimes I_{\mathcal{K}}$  fix necessary constraints—the trace normalization of the states,  $\text{Tr}[\rho_m] = 1$ , and the trace-preserving property (3.77) of the process  $E$ .

Maximization of this functional again leads to coupled extremal equations and iterative algorithm that preserves all necessary properties of estimated quantum states  $\rho_m$  and quantum process  $E$ . It can also be shown that combined ML strategies yield results superior to other methods [57].

### 3.6 Estimation of quantum measurements

Let us imagine that we possess an apparatus that performs some measurement on a certain quantum mechanical system. We are not sure which measurement the device carries out and we would like to calibrate it. This problem is in a sense complementary to the reconstruction procedures for quantum states discussed in the previous sections. Here, the role of the states and the measuring apparatus will be inverted and we shall use known states  $\rho_m$  to probe the measuring device.

#### 3.6.1 Calibration of the measuring apparatus

Suppose that the apparatus can respond with  $L$  different measurement outcomes. As is well known from the theory of quantum measurement [31, 65], such a device is completely POVM whose  $L$  elements  $\Pi_l$ ,  $l = 1, \dots, L$  govern the measurement statistics. Let us recall that the probability  $p_{ml}$  of the

measurement readout  $\Pi_l$  when measuring the quantum state with density matrix  $\rho_m$  can be expressed as

$$p_{ml} = \text{Tr}[\Pi_l \rho_m]. \quad (3.98)$$

The POVM elements are positive semi-definite hermitian operators,  $\Pi_l \geq 0$ , which decompose the identity operator,

$$\sum_{l=1}^L \Pi_l = I. \quad (3.99)$$

This ensures the probability normalization  $\sum_{l=1}^L p_{ml} = 1$ .

The general strategy to determine the POVM consists of performing a set of measurements on various known quantum states and then inferring  $\Pi_l$  from the collected experimental data [66]. The POVM can be easily estimated by direct linear inversion of Eq. (3.98). Let  $f_{ml}$  denote the total number of detections of  $\Pi_l$  for the measurements performed on the quantum state  $\rho_m$ . Assuming that the theoretical detection probability  $p_{ml}$  given by Eq. (3.98) can be replaced with the corresponding relative frequency, we may write

$$\text{Tr}[\Pi_l \rho_m] \equiv \sum_{i,j=1}^d \Pi_{l,ij} \rho_{m,ji} = \frac{f_{ml}}{\sum_{l'=1}^L f_{ml'}}, \quad (3.100)$$

where  $d$  is the dimension of the Hilbert space on which the operators  $\Pi_l$  act. Formula (3.100) establishes a system of linear equations for the matrix elements of the operators  $\Pi_l$ . If sufficient amount of data is available then Eq. (3.100) can be inverted (e.g. by the least squares method [67, 68]) and we can determine  $\Pi_l$ . This approach is a direct analogue of linear reconstruction algorithms devised for quantum-state and quantum-process reconstructions. The linear inversion is simple and straightforward, but it does not guarantee the positivity of the reconstructed POVM elements. Consequently, the linear estimation may lead to unphysical POVM, predicting negative probabilities  $p_{ml}$  for certain input quantum states. To avoid such problems, one should resort to a more sophisticated reconstruction strategy. In what follows we show how to calibrate the measuring apparatus with the use of the maximum-likelihood (ML) estimation [69]. The reconstruction of the POVM is thus another example of the remarkable utility and versatility of the estimation methods based on the maximum-likelihood principle.

### 3.6.2 Maximum-likelihood estimation of the POVM

The estimated operators  $\Pi_l$  should maximize the log-likelihood functional [69]

$$\mathcal{L}[\{\Pi_l\}] = \sum_{l=1}^L \sum_{m=1}^M f_{ml} \ln p_{ml}, \quad (3.101)$$

where  $M$  is the number of different quantum states  $\rho_m$  used for the reconstruction and note that  $p_{ml}$  depends on  $\Pi_l$  through Eq. (3.98).

We proceed as before and derive the extremal equations for the most likely POVM. The constraint (3.99) has to be incorporated by introducing a hermitian operator  $\lambda$  whose matrix elements are the Lagrange multipliers. We thus have to find the maximum of the constrained log-likelihood functional

$$\mathcal{L}'[\{\Pi_l\}] = \mathcal{L}[\{\Pi_l\}] - \sum_{l=1}^L \text{Tr}[\lambda \Pi_l]. \quad (3.102)$$

As usual, the extremal equations for  $\Pi_l$  can be derived by introducing the decomposition  $\Pi_l = A_l^\dagger A_l$ , varying  $\mathcal{L}'$  with respect to  $A_l$ , and setting the variations equal to zero. After some algebraic manipulations, one obtains

$$\Pi_l = \lambda^{-1} R_l \Pi_l, \quad (3.103)$$

where

$$R_l = \sum_{m=1}^M \frac{f_{ml}}{p_{ml}} \rho_m.$$

The extremal equation can be symmetrized by substituting  $\Pi_l = \Pi_l R_l \lambda^{-1}$  into the right-hand side of (3.103), and we have

$$\Pi_l = \lambda^{-1} R_l \Pi_l R_l \lambda^{-1}. \quad (3.104)$$

The Lagrange multiplier  $\lambda$  must be calculated self-consistently from the constraint (3.99). This yields

$$\lambda = \left( \sum_{l=1}^L R_l \Pi_l R_l \right)^{1/2}, \quad (3.105)$$

where the positive branch of the square root is taken. The extremal Eqs. (3.104) and (3.105) can be conveniently solved by means of repeated iterations. We emphasize that the conditions  $\Pi_l \geq 0$  and  $\sum_l \Pi_l = I$  are exactly fulfilled at each iteration step  $\Pi_l \rightarrow \lambda^{-1} R_l \Pi_l R_l \lambda^{-1}$ .

If there exists a POVM whose elements  $\Pi_l^e$  exactly solve linear Eqs. (3.100) then the ML estimate coincides with  $\Pi_l^e$ . In this case it holds for all  $l, m$  that

$$p_{ml} = \frac{f_{ml}}{\sum_{l'=1}^L f_{ml'}}. \quad (3.106)$$

On inserting this expression into Eq. (3.103), we find after some algebra that the set of  $L$  Eqs. (3.103) reduces to the formula for the operator of Lagrange multipliers

$$\lambda = \sum_{m=1}^M \sum_{l=1}^L f_{ml} \rho_m. \quad (3.107)$$

The principal advantage of the ML estimation lies in its ability to handle correctly any experimental data and to provide reliable estimates in cases when linear algorithms fail. As noted before, the linear inversions may provide unphysical estimates, namely operators  $\Pi_l$ , which are not positive definite. It should be noted that such a failure of linear inversion is rather typical and can occur with high probability. This is most apparent in the case of von Neumann measurement, when the operators  $\Pi_l$  are rank-one projectors and  $d - 1$  eigenvalues of each  $\Pi_l$  are equal to zero. For sufficiently large number of measured data, the linear estimate of a matrix element of  $\Pi_l$  is random variable with Gaussian distribution centered at the true value. In the basis where the projector  $\Pi_l$  is diagonal, its  $d - 1$  diagonal elements fluctuate around zero. It follows that in most cases at least one diagonal element is negative and the linear inversion yields non-positive  $\{\Pi_l\}$ , which cannot describe any measuring device.

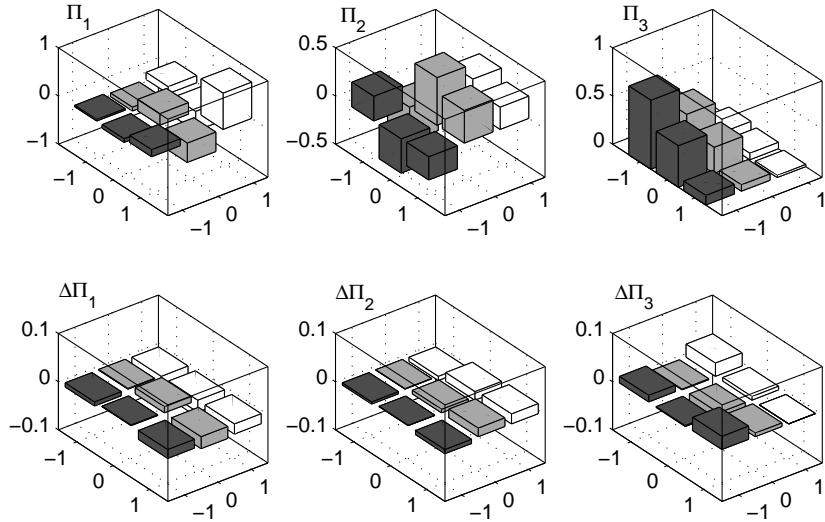
These problems of linear algorithms stem from the difference between recorded relative frequencies and theoretical probabilities, which are assumed to be equal in Eq. (3.100). The frequencies  $f_{ml}$  are fluctuating quantities with multinomial distribution characterized by probabilities  $p_{ml}$ . In the experiment we can, in principle, detect any  $f_{ml}$ . However, some sets of relative frequencies do not coincide with any theoretical probabilities (3.98) calculated for given quantum states  $\rho_m$  used for the calibration. In other words, sometimes there does not exist a POVM that would yield probabilities  $p_{ml}$  equal to detected relative frequencies and a direct linear inversion of Eq. (3.100) may then provide an unphysical result. The observation of several different quantum states by a single measuring apparatus is equivalent with the measurement of several noncommuting observables on many copies of a given quantum state. Thus the ML estimation of the quantum measurement can be interpreted as a synthesis of information from mutually incompatible observations [20, 26].

The determination of the quantum measurement can simplify considerably if we have some reliable *a-priori* information about the apparatus. In particular, the structure of the POVM may be fixed by the superselection rules. As an example let us consider a class of optical detectors that are sensitive only to the number of photons in a single mode of an electromagnetic field. The elements of the POVM describing a phase-insensitive detector are all diagonal in the Fock basis,

$$\Pi_l = \sum_n r_{ln} |n\rangle\langle n| \quad (3.108)$$

and the ML estimation reduces to the determination of the eigenvalues  $r_{ln} \geq 0$ , which is an instance of the LinPos problem. The extremal Eq. (3.103) simplifies to

$$r_{ln} = \frac{r_{ln}}{\lambda_n} \sum_{m=1}^M \frac{f_{ml}}{p_{ml}} \rho_{m,nn}, \quad \lambda_n = \sum_{m=1}^M \sum_{l=1}^L \frac{f_{ml}}{p_{ml}} \rho_{m,nn} r_{ln}. \quad (3.109)$$



**Fig. 3.6.** Example of maximum-likelihood reconstruction of a POVM that characterizes a Stern Gerlach apparatus that measures spin along the  $(1, 0, 1)/\sqrt{2}$  axis. Nine different pure states have been used for the reconstruction, and 100 copies of each state have been measured. The upper panels show the reconstructed POVM elements while the lower panels display the estimation errors  $\Delta\Pi_j$ . Only real parts of complex matrix elements are displayed.

Here  $p_{ml} = \sum_n \rho_{m,nn} r_{ln}$ . We have thus recovered a generalized version of the expectation-maximization algorithm discussed in Sec. 3.3.

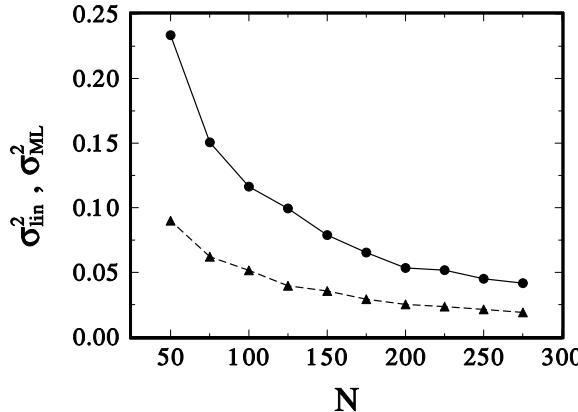
### 3.6.3 Stern-Gerlach apparatus

Let us illustrate the reconstruction of the POVM by means of numerical simulations for Stern-Gerlach apparatus measuring a spin-1 particle. We compare the linear inversion and ML estimation and demonstrate that the ML algorithm outperforms the linear estimation.

Let  $S_x$ ,  $S_y$ , and  $S_z$  denote the operators of spin projections onto axis  $x$ ,  $y$ , and  $z$ , respectively. We choose the three eigenstates of  $S_z$  as the basis states,  $S_z|s_z\rangle = s_z|s_z\rangle$ ,  $s_z = -1, 0, 1$ . In our numerical simulations, nine different pure quantum states are used for the calibration: three eigenstates of  $S_z$  and six superposition states

$$\frac{1}{\sqrt{2}}(|j_z\rangle + |k_z\rangle), \quad \frac{1}{\sqrt{2}}(|j_z\rangle + i|k_z\rangle), \quad (3.110)$$

where  $j_z, k_z = -1, 0, 1$  and  $j_z < k_z$ . The measurement on each state is performed  $N$  times. We consider a Stern-Gerlach apparatus that measures



**Fig. 3.7.** Variances of linear ( $\circ$ ) and ML ( $\triangle$ ) estimates versus the number of measurements  $N$ . The figure shows results for Stern-Gerlach apparatus measuring spin along the axis  $\mathbf{n} = (1, 1, 1)/\sqrt{3}$ .

the projection of the spin component along direction  $\mathbf{n}$ . This von Neumann measurement has three outcomes and the POVM elements  $\Pi_l$  are projectors

$$\Pi_j = |j_{\mathbf{n}}\rangle\langle j_{\mathbf{n}}|, \quad j_{\mathbf{n}} = -1, 0, 1, \quad (3.111)$$

where  $S_{\mathbf{n}}|j_{\mathbf{n}}\rangle = j_{\mathbf{n}}|j_{\mathbf{n}}\rangle$  and  $S_{\mathbf{n}} = n_x S_x + n_y S_y + n_z S_z$ .

We have performed Monte Carlo simulations of the measurements and we have subsequently reconstructed the POVM from the simulated experimental data. The ML estimates  $\Pi_{l,ML}$  were obtained by an iterative solution of the extremal equations (3.103) and (3.105). A typical example of the ML reconstruction of the POVM is given in Fig. 3.6 for  $\mathbf{n} = (1, 0, 1)/\sqrt{2}$ . The linear estimates  $\Pi_{l,lin}$  were found by solving the system of Eqs. (3.100) by means of least squares inversion. In order to compare these two procedures, we define the variances of the estimates as

$$\sigma_{ML}^2 = \left\langle \sum_l \text{Tr} [\Delta \Pi_{l,ML}^2] \right\rangle_{\text{ens}}, \quad \sigma_{lin}^2 = \left\langle \sum_l \text{Tr} [\Delta \Pi_{l,lin}^2] \right\rangle_{\text{ens}}, \quad (3.112)$$

where  $\Delta \Pi_{l,ML} = \Pi_{l,ML} - \Pi_l$ ,  $\Delta \Pi_{l,lin} = \Pi_{l,lin} - \Pi_l$ , and  $\langle \cdot \rangle_{\text{ens}}$  denotes averaging over the ensemble of all possible experimental outcomes.

We have repeated the reconstruction of the POVM for 100 different simulated experimental data and the ensemble averages yielded  $\sigma_{ML}^2$  and  $\sigma_{lin}^2$ . The variances were determined for 10 different  $N$  and the results are shown in Fig. 3.7. We can see that the ML estimates exhibit significantly lower fluctuations than the linear ones. The ML estimation procedure guarantees that the reconstructed POVM elements are positive semidefinite operators. This restriction to physically allowed  $\Pi_l$  significantly improves the reconstruction accuracy. This is a considerable practical advantage of the ML estimation compared to linear inversions.

### 3.7 Discrimination between quantum states

Discrimination between a set of known quantum states  $\rho_j$ ,  $j = 1 \dots N$  can be understood as a limiting case of quantum-state estimation with “multiple-delta-peaked” prior information. The problem becomes more interesting if we are allowed to choose the measurement at our will. The goal is then to optimize the discriminating apparatus with respect to some figure of merit.

It is no wonder that methods and tools similar to those developed in previous sections for the maximization of the likelihood can also be applied to the quantum state discrimination problem as well. This similarity will be pursued in this section.

Like density matrices in quantum estimation, the generalized measurements we optimize here are subject to positivity constraints that make quantum discrimination a highly involved nonlinear problem.

Theoretical and experimental aspects of quantum discrimination are described in detail by Bergou *et al.* and Chefles in this volume. The reader is advised to consult those chapters for basic facts and definitions.

#### 3.7.1 Minimum-error discrimination

The goal of minimum-error discrimination is to maximize the probability,

$$P_S = \sum_{j=1}^N p_j \text{Tr}[\rho_j \Pi_j], \quad (3.113)$$

of guessing the right state out of a set of  $N$  states, where the  $N$ -component POVM  $\{\Pi_j\}$  describes one of Bob’s general guessing strategies, and  $p_j$  is the prior probability of  $\rho_j$  being sent. Let us note in passing, that there are no fundamental reasons for taking error probability as the cost of decisions. Depending on applications, other measures of merit may turn out to be more appropriate, see e.g. [70].

In compact form the problem reads:

$$\begin{aligned} & \text{maximize } P_S \text{ subject to constraints} \\ & \Pi_j \geq 0, \quad j = 1, \dots, N, \\ & \sum_j \Pi_j = 1. \end{aligned} \quad (3.114)$$

Unfortunately, attacking this problem by analytical means has a chance to succeed only in the simplest cases ( $N = 2$ ) [31], or in cases with symmetric or linearly independent states [77–80]. In most situations one must resort to numerical methods. An iterative algorithm for finding the maximum of the average success rate over projection valued POVM was derived in [81]. Later, an iterative algorithm solving the problem (3.114) in its full generality was devised in [82].

We are going to seek the global maximum of the success functional  $P_S$  subject to the constraints given in Eq. (3.114). To take care of the first constraint we will decompose the POVM elements as follows  $\Pi_j = A_j^\dagger A_j$ ,  $j = 1, \dots, N$ . The other constraint (completeness) can be incorporated into our model using the method of uncertain Lagrange multipliers. Putting all these things together, the functional to be maximized becomes

$$\mathcal{L} = \sum_j p_j \text{Tr}\{\rho_j A_j^\dagger A_j\} - \text{Tr}\{\lambda \sum_j A_j^\dagger A_j\}, \quad (3.115)$$

where  $\lambda$  is a hermitian Lagrange operator. This expression is now varied with respect to  $N$  independent variables  $A_j$  to yield a necessary condition for the extremal point in the form of a set of  $N$  extremal equations for the unknown POVM elements,

$$p_j \rho_j \Pi_j = \lambda \Pi_j, \quad j = 1, \dots, N, \quad (3.116)$$

originally derived by Holevo in [83]. For our purposes it is advantageous to bring these equations to an explicitly positive semidefinite form,

$$\Pi_j = p_j^2 \lambda^{-1} \rho_j \Pi_j \rho_j \lambda^{-1}, \quad j = 1, \dots, N. \quad (3.117)$$

Lagrange operator  $\lambda$  is obtained by summing Eq. (3.117) over  $j$ ,

$$\lambda = \left( \sum_j p_j^2 \rho_j \Pi_j \rho_j \right)^{1/2}. \quad (3.118)$$

The iterative algorithm comprised of the  $N + 1$  equations (3.117) and (3.118) provides an elegant and efficient way to optimize Bob's discriminating measurement.

One usually starts from some “unbiased” trial POVM  $\{\Pi_j^0\}$ . After plugging it in Eq. (3.118) the first guess of the Lagrange operator  $\lambda$  is obtained. This operator is, in turn, used in Eq. (3.117) to get the first correction to the initial-guess strategy  $\{\Pi_j^0\}$ . The procedure gets repeated, until, eventually, a stationary point is attained. Both the positivity and completeness of the initial POVM are preserved in the course of iterating [82]. It is interesting to notice that when the initial POVM is chosen to be the maximally ignorant one,  $\Pi_j^0 = 1/N$ ,  $j = 1, \dots, N$ , the first correction is quite similar to the “pretty good” measurement [84]. For equally probable states,  $p_j = 1/N, \forall j$ , they are equivalent. Interestingly enough, the pretty good measurement is known to be optimal in certain cases [80].

Since equations (3.117) and (3.118) represent only a necessary condition for the extreme, one should always check the optimality of the stationary point by verifying the following set of conditions [31, 83]:

$$\lambda - p_j \rho_j \geq 0, \quad j = 1, \dots, N. \quad (3.119)$$

It is worth mentioning that this condition can also be derived from the theory of the semidefinite programming (SDP) [85, 86]. SDP tools also provide an alternative means of solving the problem (3.114) numerically [87]. To see the link between the quantum discrimination and SDP theory let us remind that the dual problem of SDP is defined as follows:

$$\begin{aligned} \text{maximize : } & -\text{Tr}F_0Z, \text{ subject to} \\ & Z \geq 0, \\ & \text{Tr}F_iZ = c_i, \quad i = 1, \dots, n, \end{aligned} \tag{3.120}$$

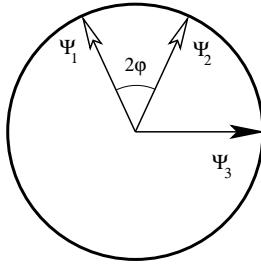
where data are  $n + 1$  hermitian matrices  $F_i$  and a complex vector  $c \in \mathbb{C}^n$ , and  $Z$  is a hermitian variable. Our problem (3.114) reduces to this dual SDP problem upon the following substitutions:

$$\begin{aligned} F_0 &= -\bigoplus_{j=1}^n p_j \rho_j, \quad Z = \bigoplus_{j=1}^n \Pi_j, \\ F_i &= \bigoplus_{j=1}^n \Gamma_i, \quad c_i = \text{Tr}\Gamma_i, \quad i = 1, \dots, h^2. \end{aligned} \tag{3.121}$$

Here operators  $\{\Gamma_i, i = 1, \dots, h^2\}$  comprise an orthonormal operator basis in the  $h^2$ -dimensional space of hermitian operators acting on the Hilbert space of our problem:  $\text{Tr}\Gamma_j\Gamma_k = \delta_{jk}, j, k = 1, \dots, h^2$ . For simplicity, let us take  $\Gamma_1$  proportional to the unity operator, then all  $c_i$  apart from  $c_1$  vanish. In SDP the necessary condition (3.116) for the maximum of the functional (3.115) is called the complementary slackness condition. When inequalities in Eq. (3.119) hold it can be shown to be also sufficient.

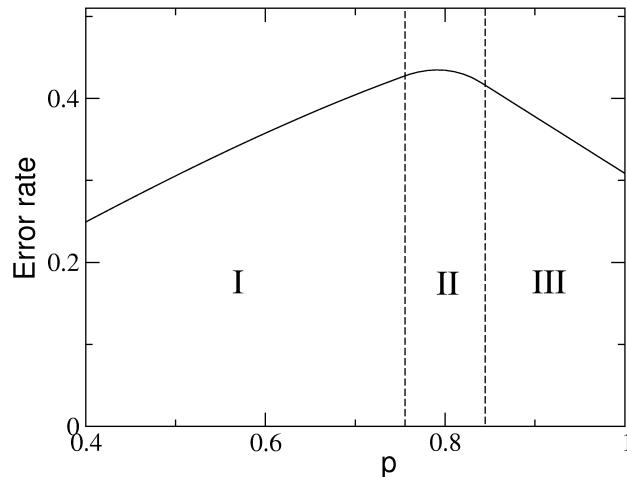
The advantage of the SDP formulation of the quantum-state discrimination problem is that there are strong numerical tools designed for solving SDP problems, for their review see [87]. They make use of the duality of SDP problems. The optimal value is bracketed between the trial maximum of the dual problem and trial minimum of the primal problem. One then hunts the optimal value down by making this interval gradually smaller. SDP tools are much more complicated than the proposed algorithm (3.117)-(3.118) but they are guaranteed to converge to the real solution.

Let us illustrate the utility of our algorithm on a simple example of discriminating between three coplanar pure qubit states. The geometry of this problem is shown in Fig. 3.8.  $\Psi_1$  and  $\Psi_2$  are equal-prior states,  $p_1 = p_2 = p/2$ , symmetrically placed around the  $z$  axis; the third state lies in the direction of  $x$  or  $y$ . A similar configuration (with  $\Psi_3$  lying along  $z$ ) has been investigated in [79]. Exploiting the mirror symmetry of their problem the authors derived analytic expressions for POVMs minimizing the average error rate. For a given angle  $\varphi$  the optimum POVM turned out to have two or three nonzero elements depending on the amount of the prior information  $p$ .



**Fig. 3.8.** A cut through the Bloch sphere showing the states to be discriminated.

Our problem is a bit more complicated one due to the lack of mirror symmetry. Let us see whether the transition from the mirror-symmetric configuration to a non-symmetric one has some influence on the qualitative behavior of the optimal POVMs. Minimal error rates calculated using the proposed iterative procedure [Eqs. (3.117) and (3.118)] for the fixed angle of  $\varphi = \pi/16$  are summarized in Fig. 3.9. For large  $p$  (region III) the optimum strategy



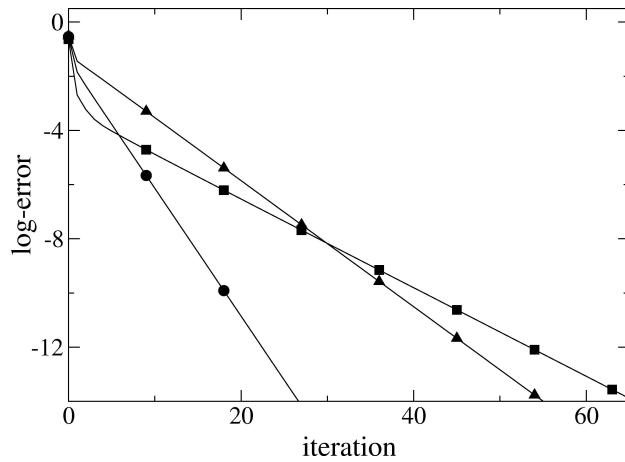
**Fig. 3.9.** Average error rate ( $1 - P_S$ ) in dependence on Bob's prior information  $p$ ;  $\varphi = \pi/16$ . Regions I, II, and III are regions where the optimum discriminating device has two, three, and two output channels, respectively.

consist in the optimal discrimination between states  $\Psi_1$  and  $\Psi_2$ . When  $\xi$  becomes smaller than a certain  $\varphi$ -dependent threshold (region II), state  $\Psi_3$  can no longer be ignored and the optimum POVM has three nonzero elements. Simple calculation yields

$$p_{\text{II,III}} = \frac{1}{1 + \sin \varphi \cos \varphi} \quad (3.122)$$

for the threshold value of the prior. However, when  $p$  becomes still smaller (region I), the optimum POVM will eventually become a two-element POVM again – the optimal strategy now being the optimal discrimination between states  $\Psi_1$  and  $\Psi_3$ . This last regime is absent in the mirror-symmetric case. The transition between regions I and II is governed by a much more complicated expression than Eq. (3.122), and will not be given here.

The convergence properties of the algorithm are shown in Fig. 3.10 for three typical prior probabilities representing regions I, II, and III of Fig. 3.9. After a short transient period an exponentially fast convergence sets in.



**Fig. 3.10.** Accuracy of the calculated error rate of the optimal POVM vs. the number of iterations. Convergence of the proposed algorithm is shown for three different priors:  $p = 0.6$  (squares),  $p = 0.8$  (triangles), and  $p = 0.9$  (circles). Ordinate is labeled by the precision in decimal digits.

Sixteen-digit precision in the resulting error-rate is usually obtained after less than one hundred iterations. Let us close the example noting that already a few iterations are enough to determine the optimum discriminating device to the precision the elements of the realistic experimental setup can be controlled within the laboratory.

### 3.7.2 Minimum-error discrimination involving inconclusive results

The above discussed scenario can be considered as limiting cases of a more general scheme that involves certain fraction of inconclusive results  $P_1$  for which we maximize the success rate. The main feature of this scheme is that

if we allow for inconclusive results then we can improve the relative (or re-normalized) success rate

$$P_{\text{RS}} = \frac{P_{\text{S}}}{1 - P_{\text{I}}}. \quad (3.123)$$

In other words, with probability  $P_{\text{I}}$  Bob fails completely and he cannot say at all which state was sent to him. However, in the rest of the cases when he succeeds he can correctly guess the state with higher probability than if he would not allow for the inconclusive results. Thus, there exists a trade-off between probability  $P_{\text{I}}$  of inconclusive results and a re-normalized success rate. For pure linearly independent states this generalized scenario was discussed in papers [88, 89]. The analysis was extended to general mixed states in [90–92].

Let us assume that the quantum state sent to Bob is drawn from the set of  $N$  mixed states  $\{\rho_j\}_{j=1}^N$  with the a-priori probabilities  $p_j$ . Bob's measurement on the state may yield  $N + 1$  different results and is formally described by the POVM whose  $N + 1$  components satisfy

$$\Pi_j \geq 0, \quad j = 0, \dots, N, \quad \sum_{j=0}^N \Pi_j = 1. \quad (3.124)$$

The outcome  $\Pi_0$  indicates failure and the probability of inconclusive results is thus given by

$$P_{\text{I}} = \sum_{j=1}^N p_j \text{Tr}[\rho_j \Pi_0]. \quad (3.125)$$

For a certain fixed value of  $P_{\text{I}}$  we want to maximize the relative success rate (3.123) that is equivalent to the maximization of the success rate (3.113). To account for the linear constraints (3.124) and (3.125) we introduce Lagrange multipliers  $\lambda$  and  $a$  where  $\lambda$  is hermitian operator and  $a$  is a real number. Taking everything together we should maximize the constrained success rate functional

$$\mathcal{L} = \sum_{j=1}^N p_j \text{Tr}[\rho_j \Pi_j] - \sum_{j=0}^N \text{Tr}[\lambda \Pi_j] + a \sum_{j=1}^N p_j \text{Tr}[\rho_j \Pi_0]. \quad (3.126)$$

With the help of Cholesky decomposition and calculus of variation we arrive at the extremal equations for the optimal POVM,

$$(\lambda - p_j \rho_j) \Pi_j = 0, \quad j = 1, \dots, N, \quad (3.127)$$

$$(\lambda - a\sigma) \Pi_0 = 0, \quad (3.128)$$

where the operator  $\sigma$  introduced for the sake of notational simplicity reads

$$\sigma = \sum_{j=1}^N p_j \rho_j. \quad (3.129)$$

From the constraint  $\text{Tr}[\sigma \Pi_0] = P_I$  we can express  $a$  in terms of  $\lambda$ ,

$$a = P_I^{-1} \text{Tr}[\lambda \Pi_0]. \quad (3.130)$$

Furthermore, if we sum all Eqs. (3.127) and also Eq.(3.128) and use the resolution of the identity (3.124), we obtain the following formula for  $\lambda$ ,

$$\lambda = \sum_{j=1}^N p_j \rho_j \Pi_j + a \sigma \Pi_0. \quad (3.131)$$

The analytical solution of this extremal problem seems to be extremely complicated. Nevertheless, we can solve again the extremal equations numerically [90] as in the case of ambiguous discrimination. In principle, one could iterate directly Eqs. (3.127) and (3.128). However, the POVM elements  $\Pi_j$  should be positive semidefinite hermitian operators. All constraints can be exactly satisfied at each iteration step if the extremal equations are symmetrized. First we express  $\Pi_j = p_j \lambda^{-1} \rho_j \Pi_j$  and combine it with its hermitian conjugate. We proceed similarly also for  $\Pi_0$  and we get

$$\Pi_j = p_j^2 \lambda^{-1} \rho_j \Pi_j \rho_j \lambda^{-1}, \quad j = 1, \dots, N, \quad (3.132)$$

$$\Pi_0 = a^2 \lambda^{-1} \sigma \Pi_0 \sigma \lambda^{-1}. \quad (3.133)$$

The Lagrange multipliers  $\lambda$  and  $a$  must be determined self-consistently so that all the constraints will hold. If we sum Eqs. (3.132) and (3.133) and take into account that  $\sum_{j=0}^N \Pi_j = 1$ , we obtain

$$\lambda = \left[ \sum_{j=1}^N p_j^2 \rho_j \Pi_j \rho_j + a^2 \sigma \Pi_0 \sigma \right]^{1/2}. \quad (3.134)$$

The fraction of inconclusive results calculated for the POVM after the iteration is given by

$$P_I = a^2 \text{Tr}[\sigma \lambda^{-1} \sigma \Pi_0 \sigma \lambda^{-1}]. \quad (3.135)$$

Since the Lagrange multiplier  $\lambda$  is expressed in terms of  $a$ , Eq. (3.135) forms a nonlinear equation for a single real parameter  $a$  (or, more precisely,  $a^2$ ). This nonlinear equation can be very efficiently solved by Newton's method of halving the interval. At each iteration step for the POVM elements, we thus solve the system of coupled nonlinear equations (3.134) and (3.135) for the Lagrange multipliers. These self-consistent iterations typically exhibit an exponentially fast convergence [82, 90].

As the fraction of inconclusive results  $P_I$  is increased the success rate  $P_S$  decreases. However, the relative success rate  $P_{RS}$  grows until it achieves its maximum  $P_{RS,\max}$ . If  $\{\rho_j\}_{j=1}^N$  are linearly independent pure states, then  $P_{RS,\max} = 1$  because exact IDP scheme works and the unambiguous discrimination is possible. Generally, however the maximum is lower than unity. The analytical expression for this maximum can be found [90].

Let us illustrate the trade-off between probability of inconclusive results and relative success rate  $P_{\text{RS}}$  on explicit example. We consider the problem of optimal discrimination between two mixed qubit states  $\rho_1$  and  $\rho_2$ . To simplify the discussion, we shall assume that the purities of these states as well as the a-priori probabilities are equal,  $\mathcal{P}_1 = \mathcal{P}_2 = \mathcal{P}$ ,  $p_1 = p_2 = 1/2$ . The mixed states can be visualized as points inside the Poincaré sphere and the purity determines the distance of the point from the center of that sphere. Without loss of generality, we can assume that both states lie in the  $xz$  plane and are symmetrically located about the  $z$  axis,

$$\rho_{1,2} = \eta\psi_{1,2}(\theta) + \frac{1-\eta}{2}\mathbb{1}, \quad (3.136)$$

where  $\psi_j = |\psi_j\rangle\langle\psi_j|$  denotes a density matrix of a pure state,

$$|\psi_{1,2}(\theta)\rangle = \cos\frac{\theta}{2}|0\rangle \pm \sin\frac{\theta}{2}|1\rangle, \quad (3.137)$$

and  $\theta \in (0, \pi/2)$ . The parameter  $\eta$  determines the purity of the mixed state (3.136),  $\mathcal{P} = (1 + \eta^2)/2$ .

From the symmetry it follows that the elements  $\Pi_1$  and  $\Pi_2$  of the optimal POVM must be proportional to the projectors  $\psi_1(\phi)$  and  $\psi_2(\phi)$ , where the angle  $\phi \in (\pi/2, \pi)$  is related to the fraction of the inconclusive results. The third component  $\Pi_0$  is proportional to the projector onto state  $|0\rangle$ . The normalization of the POVM elements can be determined from the constraint (3.124) and we find

$$\begin{aligned} \Pi_{1,2}(\phi) &= \frac{1}{2\sin^2(\phi/2)}\psi_{1,2}(\phi), \\ \Pi_0(\phi) &= \left(1 - \frac{1}{\tan^2(\phi/2)}\right)|0\rangle\langle 0|. \end{aligned} \quad (3.138)$$

The relative success rate for this POVM reads

$$P_{\text{RS}} = \frac{1 + \eta \cos(\phi - \theta)}{2(1 + \eta \cos \theta \cos \phi)} \quad (3.139)$$

and the fraction of inconclusive results is given by

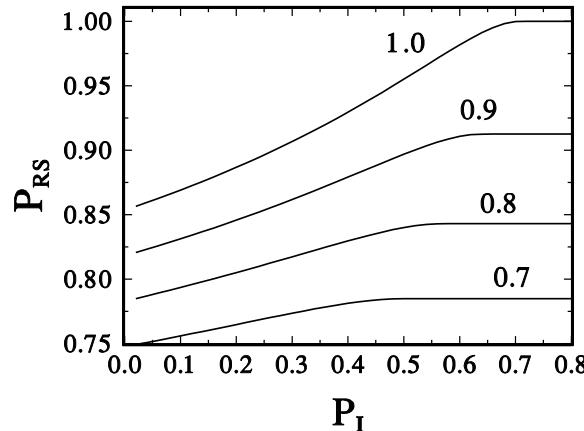
$$P_1 = \frac{1}{2}(1 + \eta \cos \theta) \left(1 - \frac{1}{\tan^2(\phi/2)}\right). \quad (3.140)$$

The formulas (3.139) and (3.140) describe implicitly the dependence of the relative success rate  $P_{\text{RS}}$  on the fraction of the inconclusive results  $P_1$ . From Eqs. (3.130) and (3.131) one can determine the Lagrange multipliers  $\lambda$  and  $a$  for the POVM (3.138) and check that the extremal Eqs. (3.127), (3.128) are satisfied. The maximum  $P_{\text{RS},\text{max}}$  is achieved if the angle  $\phi$  is chosen as follows,

$$\cos \phi_{\max} = -\eta \cos \theta. \quad (3.141)$$

On inserting the optimal  $\phi_{\max}$  back into Eq. (3.139) we get

$$P_{RS,\max} = \frac{1}{2} \left[ 1 + \frac{\eta \sin \theta}{\sqrt{1 - \eta^2 \cos^2 \theta}} \right]. \quad (3.142)$$



**Fig. 3.11.** Relative success rate  $P_{RS}$  versus the fraction of inconclusive results  $P_I$  for the optimal discrimination of two mixed states (3.136) with  $\theta = \pi/4$  and four different parameters  $\eta = 0.7$ ,  $\eta = 0.8$ ,  $\eta = 0.9$ , and  $\eta = 1.0$ .

The optimal POVM (3.138) can be also obtained numerically. We demonstrate the feasibility of iterative solution of the symmetrized extremal equations (3.132), (3.133), (3.134), and (3.135) for mixed quantum states (3.136) with the angle of separation  $\theta = \pi/4$ . The trade-off of the relative success rate and the probability of inconclusive results is shown in Fig. 3.11 for various purities of the states being discriminated. For the given probability  $P_I$  of inconclusive results and the given purity of the states the extremal equations are solved self-consistently by means of repeated iterations. The success rate  $P_S$  is calculated from the obtained optimal POVM and re-normalized according to Eq. (3.123). The numerically obtained dependence of  $P_{RS}$  on  $P_I$  is in excellent agreement with the analytical dependence following from formulas (3.139) and (3.140). Typically, a sixteen digit precision is reached after several tens of iterations. The trade-off curves shown in Fig. 3.11 reveal the monotonous growth of  $P_{RS}$  until the maximal plateau (3.142) is reached.

## Conclusions

We have presented a powerful reconstruction method that is capable of dealing with any experimental data. It stems from mathematical statistics but can

be interpreted equally well in terms of quantum theory. Though the formulations of mathematical statistics and quantum theory are rather independent, they both overlap when describing measurement and information. Maximum likelihood plays a prominent role among other estimation techniques. It exploits the full information potential of registered data, preserves the structure of quantum theory such as completeness or uncertainty relations, and reaches the ultimate resolution asymptotically. This seems to be crucial for the future potential applications in quantum information science for quantifying all the subtle and fragile quantum effects.

## Acknowledgments

This work was supported by Grant No. LN00A015 of the Czech Ministry of Educations and by Czech-Italian project No. 29, “Decoherence and Quantum Measurement. JF also acknowledges support from the EU under projects RESQ (IST-2001-35759) and CHIC (IST-2001-32150).

## References

1. S. Weigert: Phys. Rev. **A 45**, 7688 (1992); *ibid. A 53*, 2078 (1996).
2. K. Vogel, H. Risken: Phys. Rev. A **40**, 2847 (1989).
3. D. T. Smithey, M. Beck, M. G. Raymer, A. Faridani: Phys. Rev. Lett. **70**, 1244 (1993).
4. U. Leonhardt: *Measuring of the Quantum State of Light* (Cambridge Press, Cambridge, 1997).
5. S. Wallentowitz, W. Vogel: Phys. Rev. Lett. **75**, 2932 (1996).
6. P. J. Bardoff, C. Leichtle, G. Schrade, W. P. Schleich: Phys. Rev. Lett. **77**, 2198 (1996).
7. D. Leibfried, D. M. Meekhof, B. E. King, C. Monroe, W. M. Itano, D.J. Wineland: Phys. Rev. Lett. **77**, 4281 (1996).
8. A. Witten: Inverse Problems **7**, L49 (1991).
9. D.-G. Welsch, W. Vogel, T. Opatrný: Homodyne Detection and Quantum State Reconstruction. In: *Progress in Optics*, vol 39, ed by E. Wolf (North Holland, Amsterdam, 1999).
10. J. M. Bernardo, A. F. M. Smith: *Bayesian Theory* (Wiley, Chichester, 1994); S. Sykora: J. Stat. Phys. **11**, 17 (xxx).
11. R. Schack, T. A. Brun, C. M. Caves: Phys. Rev. A **64** 014305 (2001).
12. E. T. Jaynes: Information Theory and Statistical Mechanics. In: *1962 Brandeis Summer Lectures*, vol 3, ed by K. W. Ford (Benjamin, New York, 1963) p 181.
13. V. Buzek, R. Derka: Quantum observations. In *Coherence and Statistics of Photons and Atoms*, ed by J. Perina (Wiley, New York, 2001) pp 198 - 261.
14. C. M. Caves, P. D. Drummond: Rev. Mod. Phys. **66**, 481 (1994).
15. J. J. Sakurai: *Modern Quantum Mechanics* (Addison-Wesley, xxx, 1994).
16. V. Vedral, M. B. Plenio, M. A. Rippin, P. L. Knight: Phys. Rev. Lett. **78**, 2275 (1997).

17. Z. Hradil: Phys. Rev. A **55**, 1561(R) (1997).
18. K. R. W. Jones: Ann. Phys. **207**, 140 (1991).
19. V. Bužek, R. Derka, G. Adam, P. L. Knight: Ann. Phys. **266**, 454 (1998).
20. Z. Hradil, J. Summhammer: J. Phys. A: Math. Gen. **33**, 7607 (2000).
21. S. Kullback, R. A. Leibler: Ann. of Math. Stat. **22**, 79 (1951).
22. J. H. Shapiro, S. R. Shepard, N. C. Wong: Phys. Rev. Lett. **62**, 2377 (1989).
23. A. S. Lane, S. L. Braunstein, C. M. Caves: Phys. Rev. A **47**, 1667 (1993).
24. J. Řeháček, Z. Hradil, M. Zawisky, S. Pascazio, H. Rauch, J. Peřina: Phys. Rev. A **60**, 473 (1999).
25. G. M. D' Ariano, M. G. A. Paris, M. F. Sacchi: Phys. Rev. A **62**, 023815 (2000).
26. Z. Hradil, J. Summhammer, H. Rauch: Phys. Lett. A **261**, 20 (1999).
27. A. P. Dempster, N. M. Laird, D. B. Rubin: J. R. Statist. Soc. B **39**, 1 (1977).
28. Y. Vardi and D. Lee: J. R. Statist. Soc. B **55**, 569 (1993).
29. K. Banaszek, G. M. D' Ariano, M. G. A. Paris, M. F. Sacchi: Phys. Rev. A **61**, 010304(R) (1999).
30. J. Řeháček, Z. Hradil, and M. Ježek: Phys. Rev. A **63**, 040303(R) (2001).
31. C. W. Helstrom: *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
32. A.G.White, D.F.V. James, P.H. Eberhard, and P.G. Kwiat: Phys. Rev. Lett. **83**, 3103 (1999).
33. R. Jozsa: J. Mod. Optics **41**, 2315 (1993).
34. W.K. Wootters: Phys. Rev. D **23**, 357 (1981).
35. D. Dieks: Phys. Lett. A **126**, 303 (1998).
36. A. Peres: Phys. Lett. A **128**, 19 (1998).
37. V. Vedral, M.B.Plenio, K. Jakobs, and P.L. Knight: Phys. Rev. A **56**, 4452 (1997).
38. R.A. Fisher: Proc. Camb. Phi. Soc. **22**, 700 (1925).
39. C.R. Rao: Bull. Calcutta Math. Soc. **37**, 81 (1945).
40. H. Cramér: *Mathematical methods of statistics* (Princeton University Press, Princeton, 1946).
41. D.F.V. James, P.G. Kwiat, W.J. Munro, and A.G. White: Phys. Rev. A **64**, 052312 (2001); e-print arXiv:quant-ph/0103121.
42. M.A. Nielsen and I.L. Chuang: *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
43. I. L. Chuang and M. A. Nielsen: J. Mod. Opt. **44**, 2455 (1997).
44. J. F. Poyatos, J. I. Cirac, and P. Zoller: Phys. Rev. Lett. **78**, 390 (1997).
45. G. M. D'Ariano and L. Maccone: Phys. Rev. Lett. **80**, 5465 (1998).
46. A. Luis and L. L. Sánchez-Soto: Phys. Lett. A **261**, 12 (1999).
47. R. Gutzeit, S. Wallentowitz, and W. Vogel: Phys. Rev. A **61**, 062105 (2000).
48. G.M. D'Ariano and P. Lo Presti: Phys. Rev. Lett. **86**, 4195 (2001).
49. D.G. Fischer, H. Mack, M.A. Cirone, and M. Freyberger: Phys. Rev. A **64**, 022309 (2001).
50. M. F. Sacchi: Phys. Rev. A **63**, 054104 (2001).
51. A. M. Childs, I. L. Chuang, and D. W. Leung: Phys. Rev. A **64**, 012314 (2001).
52. N. Boulant, T. F. Havel, M. A. Pravia, and D. G. Cory: Phys. Rev. A **67**, 042322 (2003).
53. F. De Martini, A. Mazzei, M. Ricci, and G.M. D'Ariano: Phys. Rev. A **67**, 062307 (2003).

54. J.B. Altepeter, D. Branning, E. Jeffrey, T.C. Wei, P.G. Kwiat, R.T. Thew, J.L. O'Brien, M.A. Nielsen, and A.G. White: Phys. Rev. Lett. **90**, 193601 (2003).
55. M.W. Mitchell, C.W. Ellenor, S. Schneider, and A.M. Steinberg: Phys. Rev. Lett. **91**, 120402 (2003).
56. J. Fiurášek and Z. Hradil: Phys. Rev. A **63**, 020101(R) (2001).
57. M. Ježek, J. Fiurášek, and Z. Hradil: Phys. Rev. A **68**, 012305 (2003).
58. A. Jamolkowski: Rep. Math. Phys. **3**, 275 (1972).
59. M. Dakna, J. Clausen, L. Knöll, and D.-G. Welsch: Phys. Rev. A **59**, 1658 (1999);
60. H. Lee, P. Kok, N.J. Cerf, and J.P. Dowling: Phys. Rev. A **65**, 030101 (2002).
61. J. Fiurášek: Phys. Rev. A **65**, 053818 (2002).
62. C.H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J.A. Smolin, and W.K. Wootters: Phys. Rev. Lett. **78**, 2031 (1997).
63. D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera: Phys. Rev. Lett. **77**, 2818 (1996).
64. E. Knill, R. Laflamme, and G.J. Milburn: Nature (London) **409**, 46 (2001).
65. A.S. Holevo: *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
66. A. Luis and L.L. Sanchez-Soto: Phys. Rev. Lett. **83**, 3573 (1999).
67. T. Opatrný, D.-G. Welsch, and W. Vogel: Phys. Rev. A **56**, 1788 (1997).
68. S.M. Tan: J. Mod. Opt. **44**, 2233 (1997).
69. J. Fiurášek: Phys. Rev. A **64**, 024102 (2001).
70. M. Ježek: Phys. Lett. A **299**, 441 (2002),
71. Z. Hradil and J. Summhammer: Phys. Rev. A **62**, 014101 (2000).
72. W.K. Wootters and W.H. Zurek: Nature (London) **299**, 802 (1982).
73. D. Dieks: Phys. Lett. A **92**, 271 (1982).
74. C.H. Bennett and G. Brassard: In: *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, New York, Bangalore, India, 1984) p 175.
75. C.H. Bennett, G. Brassard, and N.D. Mermin: Phys. Rev. Lett. **68**, 557 (1992).
76. C. H. Bennett: Phys. Rev. Lett. **68**, 3121 (1992).
77. M. Sasaki, K. Kato, M. Izutsu, and O. Hirota: Phys. Rev. A **58**, 146 (1998).
78. S.M. Barnett: Phys. Rev. A **64**, 030303(R) (2001).
79. E. Andersson, S.M. Barnett, C.R. Gilson, and K. Hunter: Phys. Rev. A **65**, 052308 (2002).
80. M. Ban, K. Kurokawa, R. Momose, and O. Hirota: Int. J. Theor. Phys. **36**, 1269 (1997).
81. C.W. Helstrom: IEEE Trans. IT-28, 359 (1982).
82. M. Ježek, J. Řeháček, and J. Fiurášek: Phys. Rev. A **65**, 060301(R) (2002).
83. A.S. Holevo: J. Multivar. Anal. **3**, 337 (1973).
84. P. Hausladen and W. K. Wootters: J. Mod. Opt. **41**, 2385 (1994).
85. H.P. Yuen, R.S. Kennedy, and M. Lax: IEEE Trans. IT-21, 125 (1975).
86. Y.C. Eldar, A. Megretski, and G.C. Verghese: IEEE Trans. Inform. Theory **49**, 1007 (2003).
87. L. Vandenberghe and S. Boyd: Semidefinite programming, SIAM Review **38**, 49 (1996).
88. A. Chefles and S. M. Barnett: J. Mod. Opt. **45**, 1295 (1998).
89. C. W. Zhang, C. F. Li, and G. C. Guo: Phys. Lett. A **261**, 25 (1999).
90. J. Fiurášek and M. Ježek: Phys. Rev. A **67**, 012321 (2003).
91. Y. C. Eldar: IEEE Trans. Inform. Theory **49**, 446 (2003).
92. Y. C. Eldar: Phys. Rev. A **67**, 042309 (2003).



## 4 Qubit Quantum State Tomography

Joseph. B. Altepeter<sup>1</sup>, Daniel F. V. James<sup>2</sup>, and Paul G. Kwiat<sup>1</sup>

<sup>1</sup> Dept. of Physics, University of Illinois at Urbana-Champaign, Urbana IL 61801  
[altepete@uiuc.edu](mailto:altepete@uiuc.edu), [kwiat@uiuc.edu](mailto:kwiat@uiuc.edu)

<sup>2</sup> Theoretical Division T-4, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 [dfvj@lanl.gov](mailto:dfvj@lanl.gov)

Much like its classical counterpart, which aims at reconstructing three-dimensional images via a series of two-dimensional projections along various ‘cuts’, quantum tomography characterizes the complete quantum state of a particle or particles through a series of measurements in different bases. While the characterization of a classical object can involve a series of measurements on the same subject, measuring a single quantum particle perturbs its state, often making its further investigation uninformative. For this reason, quantum tomography must be carried out in stages on a number of identical copies of the same state, and can never be successfully applied to a single unknown particle. The etymology of tomography is therefore descriptive (the Greek *tomos* means section) – a series of measurements on identical particle ensembles each allow a glimpse into a distinct aspect of a quantum state’s reality. Each new type of measurement illuminates a new dimension of an unknown state; subjecting more identical copies of that state to a single type of measurement brings that particular observable into sharper relief.

This chapter will present a review of the representation of quantum states, a procedure for and an explanation of the reconstruction of an unknown state from a series of ideal measurements on an ensemble of identical particles, and the adaptation of that method to real systems, i.e., non-ideal measurements. Each section will be presented first in general, without respect to a particular physical qubit implementation, followed by the application of that theory to the specific case of qubits encoded into the polarization states of photons. We chose this system for convenience and availability of clean results. However, the techniques presented here can be applied to determine the quantum state of *any* system of one or more qubits (2-level systems). This includes photons [1–10], spin- $\frac{1}{2}$  particles (as, e.g., are used in NMR quantum computing [11–14]), and (effectively) 2-level atoms [15, 16]. In order to facilitate the use of these techniques by groups and individuals working in any field, a website is available which provides both further details about these techniques and working, documented code for implementing them.<sup>3</sup>

---

<sup>3</sup> <http://www.physics.uiuc.edu/Research/QI/Photonics/Tomography/>

## 4.1 State Representation

Before states can be analyzed, it is necessary to understand their representation. In particular, the reconstruction of an unknown state is often simplified by a specific state parametrization.

### 4.1.1 Representation of Single Qubit States

Rather than begin with a general treatment of tomography for an arbitrary number of qubits, throughout this chapter the single-qubit case will be investigated initially. This provides the opportunity to strengthen an intuitive grasp of the fundamentals of state representation and tomography before moving on to the more complex (and more useful) general case. In pursuance of this goal, we will use several graphical representations only available at the single-qubit level.

#### Pure States, Mixed States, and Diagonal Representations

In general, any single-qubit in a pure state can be represented by

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (4.1)$$

where  $\alpha$  and  $\beta$  are complex and  $|\alpha|^2 + |\beta|^2 = 1$  [19]. If the normalization is written implicitly and the global phase is ignored, this can be rewritten as

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)e^{i\phi}|1\rangle. \quad (4.2)$$

These representations are sufficient to enable the description of the action of any operator (e.g., projectors or unitary rotations) on a pure state, and therefore to carry out tomography on that state. However, as previously discussed, any such tomography would require an ensemble of such states. What if the measured ensemble – being as yet unknown – contained an ensemble of *different* pure states? Or what if the members of the ensemble were themselves not in pure states (perhaps because they are entangled with unobserved degrees of freedom)? In this case the overall state is *mixed*.

In general, these mixed states may be described by a probabilistically weighted incoherent sum of pure states. In other words, it is as if any particle in the ensemble has a specific probability of being in a given pure state, and this state is distinguishably labelled in some way. If it were not distinguishable, its constituent pure states would add coherently (with a definite relative phase), yielding a single pure state. A mixed state can be represented by a density matrix  $\hat{\rho}$ , where

$$\hat{\rho} = \sum_i P_i |\psi_i\rangle \langle \psi_i| = \begin{pmatrix} \langle 0| & \langle 1| \\ |0\rangle & |1\rangle \end{pmatrix} \begin{pmatrix} A & Ce^{i\phi} \\ Ce^{-i\phi} & B \end{pmatrix}. \quad (4.3)$$

$P_i$  is the probabilistic weighting ( $\sum_i P_i = 1$ ),  $A, B$  and  $C$  are all real and non-negative,  $A + B = 1$ , and  $C \leq \sqrt{AB}$  [19].

While any ensemble of pure states can be represented in this way, it is also true that *any* ensemble of single-qubit states can be represented by an ensemble of only *two* orthogonal pure states. (Two pure states  $|\psi_i\rangle$  and  $|\psi_j\rangle$  are orthogonal if  $|\langle\psi_i|\psi_j\rangle| = 0$ ). For example, if the matrix from equation 4.3 were diagonal, then it would clearly be a probabilistic combination of two orthogonal states, as

$$\begin{array}{cc} \langle 0 | & \langle 1 | \\ | 0 \rangle & \left( \begin{array}{cc} A & 0 \\ 0 & B \end{array} \right) = A|0\rangle\langle 0| + B|1\rangle\langle 1|. \end{array} \quad (4.4)$$

However, *any* physical density matrix can be diagonalized, such that

$$\hat{\rho} = \begin{array}{cc} \langle \psi | & \langle \psi^\perp | \\ |\psi\rangle & \left( \begin{array}{cc} E_1 & 0 \\ 0 & E_2 \end{array} \right) = E_1|\psi\rangle\langle\psi| + E_2|\psi^\perp\rangle\langle\psi^\perp|, \end{array} \quad (4.5)$$

where  $\{E_1, E_2\}$  are the eigenvalues of  $\hat{\rho}$ , and  $\{|\psi\rangle, |\psi^\perp\rangle\}$  are the eigenvectors (recall that these eigenvectors can always be made mutually orthogonal, denoted here by the  $\perp$  symbol). Thus the representation of any quantum state, no matter how it is constructed, is identical to that of an ensemble of two orthogonal pure states.<sup>4</sup>

#### Examples in the Photon Case

Throughout this chapter, examples will be provided using qubits encoded into the electric field polarization of photons. For a single photon, this system has two levels, e.g., horizontal ( $|H\rangle \equiv |0\rangle$ ) and vertical ( $|V\rangle \equiv |1\rangle$ ), with all possible pure polarization states constructed from coherent superpositions of these two states. For example, diagonal, antidiagonal, right-circular and left-circular light are respectively represented by

---

<sup>4</sup> It is an interesting question whether all physical states described by a mixed state (e.g., equation 4.5) are indeed completely equivalent. For example, Leonhardt discussed the notion that two types of unpolarized light could be considered, depending on whether the incoherence between polarization components arose purely due to an averaging over rapidly varying phases, or from an entanglement with another quantum system altogether [17]. This line of thought can even be pushed further, by asking whether all mixed states necessarily arise only from tracing over some unobserved degrees of freedom with which the quantum system has become entangled, or if indeed such entanglement may ‘collapse’ when the systems involved approach macroscopic size [18]. If the latter were true, then there would exist mixed states that could *not* be seen as pure in some larger Hilbert space. In any event, to our knowledge, at least insofar as state tomography is concerned, these subtleties of interpretation do not in any way affect experimental results.

$$\begin{aligned}
|D\rangle &\equiv (|H\rangle + |V\rangle)/\sqrt{2}, \\
|A\rangle &\equiv (|H\rangle - |V\rangle)/\sqrt{2}, \\
|R\rangle &\equiv (|H\rangle + i|V\rangle)/\sqrt{2}, \\
\text{and } |L\rangle &\equiv (|H\rangle - i|V\rangle)/\sqrt{2}.
\end{aligned} \tag{4.6}$$

Now consider measuring a source of photons which emits a one-photon wave packet each second, but alternates between horizontal, vertical, and diagonal polarizations. Their emission time labels these states (in principle) as distinguishable, and so if we ignore that timing information when they are measured, we must represent their state as a density matrix  $\hat{\rho}$ :

$$\begin{aligned}
\hat{\rho} &= \frac{1}{3}(|H\rangle\langle H| + |V\rangle\langle V| + |D\rangle\langle D|) \\
&= \frac{1}{3} \left( |H\rangle \begin{pmatrix} \langle H| & \langle V| \\ 1 & 0 \\ 0 & 0 \end{pmatrix} + |V\rangle \begin{pmatrix} \langle H| & \langle V| \\ 0 & 0 \\ 0 & 1 \end{pmatrix} + |D\rangle \begin{pmatrix} \langle H| & \langle V| \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \right) \\
&= \frac{1}{6} \left( |H\rangle \begin{pmatrix} \langle H| & \langle V| \\ 3 & 1 \\ 1 & 3 \end{pmatrix} \right).
\end{aligned} \tag{4.7}$$

When diagonalized,

$$\hat{\rho} = \frac{1}{3} \left( |D\rangle \begin{pmatrix} \langle D| & \langle A| \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \right) = \frac{2}{3}|D\rangle\langle D| + \frac{1}{3}|A\rangle\langle A|, \tag{4.8}$$

which, as predicted in equation 4.5, is a sum of only *two* orthogonal states.

Henceforth, the ‘bra’ and ‘ket’ labels will be suppressed from written density matrices where the basis is  $\{|0\rangle, |1\rangle\}$  or  $\{|H\rangle, |V\rangle\}$ .

### The Stokes Parameters and the Poincaré Sphere

Any single-qubit density matrix  $\hat{\rho}$  can be uniquely represented by three parameters  $\{S_1, S_2, S_3\}$ :

$$\hat{\rho} = \frac{1}{2} \sum_{i=0}^3 S_i \hat{\sigma}_i. \tag{4.9}$$

The  $\hat{\sigma}_i$  matrices are

$$\hat{\sigma}_0 \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \hat{\sigma}_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \hat{\sigma}_2 \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \hat{\sigma}_3 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{4.10}$$

and the  $S_i$  values are given by [19]

$$S_i \equiv \text{Tr} \{ \hat{\sigma}_i \hat{\rho} \}. \quad (4.11)$$

For all pure states,  $\sum_{i=1}^3 S_i^2 = 1$ ; for mixed states,  $\sum_{i=1}^3 S_i^2 < 1$ ; for the completely mixed state,  $\sum_{i=1}^3 S_i^2 = 0$ . Due to normalization,  $S_0$  will always equal one.

Physically, each of these parameters directly corresponds to the outcome of a specific pair of projective measurements:

$$\begin{aligned} S_0 &= P_{|0\rangle} + P_{|1\rangle} \\ S_1 &= P_{\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)} - P_{\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)} \\ S_2 &= P_{\frac{1}{\sqrt{2}}(|0\rangle+i|1\rangle)} - P_{\frac{1}{\sqrt{2}}(|0\rangle-i|1\rangle)} \\ S_3 &= P_{|0\rangle} - P_{|1\rangle}, \end{aligned} \quad (4.12)$$

where  $P_{|\psi\rangle}$  is the probability to measure the state  $|\psi\rangle$ . As we shall see below, these relationships between probabilities and  $S$  parameters are extremely useful in understanding more general operators. Because  $P_{|\psi\rangle} + P_{|\psi^\perp\rangle} = 1$ , these can be simplified in the single-qubit case, and

$$P_{|\psi\rangle} - P_{|\psi^\perp\rangle} = 2P_{|\psi\rangle} - 1, \quad (4.13)$$

where  $|\psi^\perp\rangle$  denotes the state orthogonal to  $|\psi\rangle$ .

The probability of projecting a given state  $\hat{\rho}$  into the state  $|\psi\rangle$  (the probability of measuring  $|\psi\rangle$ ) is given by [20]:

$$\begin{aligned} P_{|\psi\rangle} &= \langle \psi | \hat{\rho} | \psi \rangle \\ &= \text{Tr} \{ |\psi\rangle \langle \psi | \hat{\rho} \}. \end{aligned} \quad (4.14)$$

In 4.12 above, the  $S_i$  are defined with respect to three states,  $|\phi\rangle_i$ :

$$\begin{aligned} |\phi\rangle_1 &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ |\phi\rangle_2 &= \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle) \\ |\phi\rangle_3 &= |0\rangle, \end{aligned} \quad (4.15)$$

and their orthogonal complements,  $|\phi^\perp\rangle$ ; parameters similar to these and serving the same function can be defined with respect to any three arbitrary states,  $|\psi_i\rangle$ . (Arbitrary with one condition: the matrices  $|\psi_i\rangle \langle \psi_i|$  along with the identity must be linearly independent.) Operators analogous to the  $\hat{\sigma}$  operators can be defined relative to these states:

$$\hat{\tau}_i \equiv |\psi_i\rangle \langle \psi_i| - |\psi_i^\perp\rangle \langle \psi_i^\perp|. \quad (4.16)$$

We can further define an ‘S-like’ parameter  $T$ , given by:

$$T_i = \text{Tr} \{ \hat{\tau}_i \hat{\rho} \}. \quad (4.17)$$

Continuing the previous convention and to complete the set, we define  $\hat{\tau}_0 \equiv \hat{\sigma}_0$ , which then requires that  $T_0 = 1$ . Note that the  $S_i$  parameters are simply a special case of the  $T_i$ , for the case when  $\hat{\tau}_i = \hat{\sigma}_i$ .

Unlike the specific case of the  $S$  parameters which describe *orthogonal* measurement bases, for non-orthogonal measurements

$$\hat{\rho} \neq \frac{1}{2} \sum_{i=0}^3 T_i \hat{\tau}_i. \quad (4.18)$$

In order to reconstruct the density matrix, the  $T$  parameters must first be transformed into the  $S$  parameters (see equation 4.21).

### The Photon Case

For photon polarization, the  $S_i$  are the famous Stokes parameters (though normalized), and correspond to measurements of D/A, R/L, and H/V [21]. In terms of the  $\hat{\tau}$  matrices just introduced, we would define a set of basis states  $|\psi_1\rangle \equiv |D\rangle$ ,  $|\psi_2\rangle \equiv |R\rangle$ , and  $|\psi_3\rangle \equiv |H\rangle$ . For these analysis bases,  $\hat{\tau}_1 = \hat{\sigma}_1$ ,  $\hat{\tau}_2 = \hat{\sigma}_2$ , and  $\hat{\tau}_3 = \hat{\sigma}_3$  (and therefore  $T_i = S_i$  for this specific choice of analysis bases).

As the simplest example, consider the input state  $|H\rangle$ . Applying equation (4.11), we find that

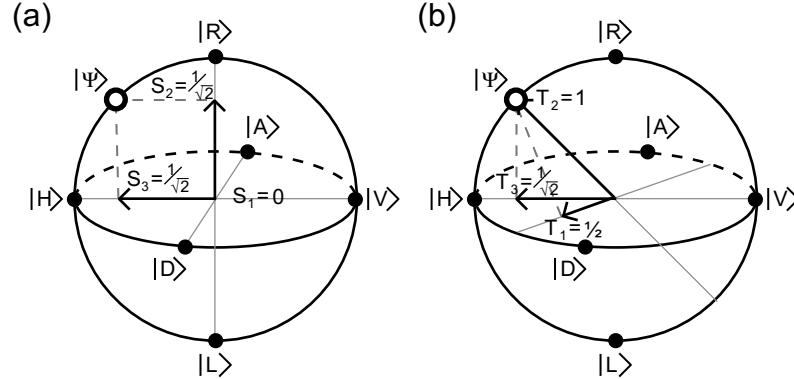
$$\begin{aligned} S_0 &= \text{Tr} \{ \sigma_0 \rho_H \} = 1 \\ S_1 &= \text{Tr} \{ \sigma_1 \rho_H \} = 0 \\ S_2 &= \text{Tr} \{ \sigma_2 \rho_H \} = 0 \\ S_3 &= \text{Tr} \{ \sigma_3 \rho_H \} = 1, \end{aligned} \quad (4.19)$$

which from equation 4.9 implies that

$$\rho_H = (\sigma_0 + \sigma_3) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (4.20)$$

When the Stokes parameters are used as coordinates in 3-space, the space of all legal states falls within a sphere of radius one (the Poincaré sphere for polarization, the Bloch sphere for electron spin or other two-level systems). The pure states are found on the surface, states of linear polarization on the equator, circular states at the poles, mixed states within, and the totally mixed state – corresponding to completely unpolarized photons – at the center of the sphere. This provides a very convenient way to visualize one-qubit states (see figure 4.1). The  $\theta$  and  $\phi$  values from equation 4.2 allow any pure

state to be easily mapped onto the sphere surface. These values are the polar coordinates of the pure state they represent on the Poincaré sphere.<sup>5</sup> In addition to mapping states, the sphere can be used to represent any unitary operation as a rotation about an arbitrary axis. For example, waveplates implement rotations about an axis that passes through the equator.



**Fig. 4.1.** The Bloch Poincaré sphere. Any single-qubit quantum state  $\hat{\rho}$  can be represented by three parameters  $T_i = \text{Tr}\{\hat{\tau}_i \hat{\rho}\}$ , as long as the operators  $\hat{\tau}_i$  in addition to the identity are linearly independent. Physically, the  $T_i$  parameters directly correspond to the outcome of a specific projective measurement:  $T_i = 2P_i - 1$ , where  $P_i$  is the probability of success for the measurement. The  $T_i$  may be used as coordinates in 3-space. Then all 1-qubit quantum states fall on or within a sphere of radius one. The surface of the sphere corresponds to pure states, the interior to mixed states, and the origin to the totally mixed state. Shown is a particular pure state  $|\psi\rangle$ , which is completely specified by its projection onto a set of non-parallel basis vectors. (a) When  $\hat{\tau}_i = \hat{\sigma}_i$  (the Pauli matrices), the basis vectors are orthogonal, and in this particular case the  $T_i$  are equal to the  $S_i$ , the well known parameters, corresponding to measurements of diagonal ( $S_1$ ), right-circular ( $S_2$ ), and horizontal ( $S_3$ ) polarizations. (b) A non-orthogonal coordinate system in Poincaré space. It is possible to represent a state using its projection onto non-orthogonal axes in Poincaré space. This is of particular use when attempting to reconstruct a quantum state from non-orthogonal measurements. Shown here are the axes corresponding to measurements of 22.5° linear ( $T_1$ ), elliptical light rotated 22.5° from H towards R ( $T_2$ ), and horizontal ( $T_3$ ).

Any state  $|\psi\rangle$  and its orthogonal partner,  $|\psi^\perp\rangle$ , are found on opposite points of the Poincaré sphere. The line connecting these two points forms an

<sup>5</sup> These polar coordinates are by convention rotated by 90°, so that  $\theta = 0$  is on the equator corresponding to the state  $|H\rangle$  and  $\theta = 90^\circ, \phi = 90^\circ$  is at the North Pole corresponding to the state  $|R\rangle$ . This 90° rotation is particular to the Poincaré representation of photon polarization [22]; representations of two-level systems on the Bloch sphere do not introduce it.

axis of the sphere, useful for visualizing the outcome of a measurement in the  $|\psi\rangle/|\psi^\perp\rangle$  basis. The projection of any state  $\hat{\rho}$  (through a line perpendicular to the  $|\psi\rangle/|\psi^\perp\rangle$  axis), will lie a distance along this axis corresponding to the relevant Stokes-like parameter ( $T = \langle\psi|\hat{\rho}|\psi\rangle - \langle\psi^\perp|\hat{\rho}|\psi^\perp\rangle$ ).

Thus, just as any point in three-dimensional space can be specified by its projection onto three linearly independent axes, any quantum state can be specified by the three parameters  $T_i = \text{Tr}\{\hat{\tau}_i\hat{\rho}\}$ , where  $\hat{\tau}_{i=1,2,3}$  are linearly independent matrices equal to  $|\psi_i\rangle\langle\psi_i| - |\psi_i^\perp\rangle\langle\psi_i^\perp|$ . The  $\hat{\tau}_i$  correspond to general Stokes-like parameters for any three linearly independent axes on the Poincaré sphere. However, they can differ from the canonical Stokes axes and need not even be orthogonal. See figure 4.1b for an example of state representation using non-orthogonal axes.

In order to use these non-orthogonal Stokes-like parameters, it is necessary to be able to transform a state from a Stokes representation to a new nonorthogonal representation and vice-versa. In general, for any two representations  $S_i = \text{Tr}\{\hat{\sigma}_i\hat{\rho}\}$  and  $T_i = \text{Tr}\{\hat{\tau}_i\hat{\rho}\}$  it is possible to transform between them by using

$$\begin{pmatrix} T_0 \\ T_1 \\ T_2 \\ T_3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \text{Tr}\{\hat{\tau}_0\hat{\sigma}_0\} & \text{Tr}\{\hat{\tau}_0\hat{\sigma}_1\} & \text{Tr}\{\hat{\tau}_0\hat{\sigma}_2\} & \text{Tr}\{\hat{\tau}_0\hat{\sigma}_3\} \\ \text{Tr}\{\hat{\tau}_1\hat{\sigma}_0\} & \text{Tr}\{\hat{\tau}_1\hat{\sigma}_1\} & \text{Tr}\{\hat{\tau}_1\hat{\sigma}_2\} & \text{Tr}\{\hat{\tau}_1\hat{\sigma}_3\} \\ \text{Tr}\{\hat{\tau}_2\hat{\sigma}_0\} & \text{Tr}\{\hat{\tau}_2\hat{\sigma}_1\} & \text{Tr}\{\hat{\tau}_2\hat{\sigma}_2\} & \text{Tr}\{\hat{\tau}_2\hat{\sigma}_3\} \\ \text{Tr}\{\hat{\tau}_3\hat{\sigma}_0\} & \text{Tr}\{\hat{\tau}_3\hat{\sigma}_1\} & \text{Tr}\{\hat{\tau}_3\hat{\sigma}_2\} & \text{Tr}\{\hat{\tau}_3\hat{\sigma}_3\} \end{pmatrix} \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix}. \quad (4.21)$$

This relation allows  $S$  parameters to be transformed into any set of  $T$  parameters. In order to transform from  $T$  to  $S$ , we can invert the 4 by 4 matrix in equation 4.21 and multiply both sides by this new matrix. This inversion is possible because we have chosen the  $\hat{\tau}_i$  operators to be linearly independent, as otherwise the  $T_i$  parameters would not specify a single point in Hilbert space.

#### 4.1.2 Representation of Multiple Qubits

With the extension of these ideas to cover multiple qubits, it becomes possible to investigate non-classical features, including the quintessentially quantum mechanical phenomenon of entanglement.

#### Pure States, Mixed States, and Diagonal Representations

As the name implies, multiple-qubit states are constructed out of individual qubits. As such, the Hilbert space of a many qubit system is spanned by state vectors which are the tensor product of single-qubit state vectors. A general  $n$ -qubit system can be written as

$$|\psi\rangle = \sum_{i_1, i_2, \dots, i_n=0,1} \alpha_{i_1, i_2, \dots, i_n} |i_1\rangle \otimes |i_2\rangle \otimes \dots \otimes |i_n\rangle. \quad (4.22)$$

Here the  $\alpha_i$  are complex,  $\sum_i |\alpha_i|^2 = 1$ , and  $\otimes$  denotes a tensor product, used to join component Hilbert spaces. For example, a general two-qubit pure state can be written

$$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle, \quad (4.23)$$

where  $|00\rangle$  is shorthand for  $|0\rangle_1 \otimes |0\rangle_2$ .

As before, we represent a general mixed state through an incoherent sum of pure states:

$$\hat{\rho} = \sum_i P_i |\psi_i\rangle\langle\psi_i|. \quad (4.24)$$

And, as before, each  $n$ -qubit state can be represented by a  $2^n$  by  $2^n$  density matrix which, when diagonalized, allows any state to be written as

$$\hat{\rho} = \sum_{i=1}^{2^n} P_i |\phi_i\rangle\langle\phi_i|. \quad (4.25)$$

(4.24) differs from (4.25) in that the  $\phi_i$  are necessarily orthogonal ( $\langle\psi_i|\psi_j\rangle = \delta_{ij}$ ), and there are at most  $2^n$  of them (in (4.24) there could be an arbitrary number of  $|\psi_i\rangle$ ).

### The Photon Case

Any two-qubit polarization state can be written as

$$\hat{\rho} = \begin{pmatrix} |HH\rangle & \langle HH| & |HV\rangle & \langle HV| \\ |HV\rangle & A_1 & B_1 e^{i\phi_1} & B_2 e^{i\phi_2} \\ |VH\rangle & B_1 e^{-i\phi_1} & A_2 & B_4 e^{i\phi_4} \\ |VV\rangle & B_2 e^{-i\phi_2} & B_4 e^{-i\phi_4} & A_3 \\ & B_3 e^{-i\phi_3} & B_5 e^{-i\phi_5} & B_6 e^{i\phi_6} \end{pmatrix}, \quad (4.26)$$

where  $\hat{\rho}$  is positive and Hermitian with unit trace. Henceforth, the ‘bra’ and ‘ket’ labels will be omitted from density matrices presented in this standard basis.

Perhaps the most famous examples of pure two-qubit states are the Bell states [23]:

$$\begin{aligned} |\phi^\pm\rangle &= \frac{1}{\sqrt{2}} (|HH\rangle \pm |VV\rangle) \\ |\psi^\pm\rangle &= \frac{1}{\sqrt{2}} (|HV\rangle \pm |VH\rangle). \end{aligned} \quad (4.27)$$

Mixed states of note include the Werner states [24],

$$\hat{\rho}_W = P|\gamma\rangle\langle\gamma| + (1-P)\frac{1}{4}I, \quad (4.28)$$

where  $|\gamma\rangle$  is a maximally entangled state and  $\frac{1}{4}I$  is the totally mixed state, and the maximally entangled mixed states (MEMS), which possess the maximal amount of entanglement for a given amount of mixture [25].

The entanglement and the mixture are two of the many quantities derived from the density matrix used to characterize a quantum state, several of which will be included here for reference:

*Fidelity* Fidelity is a measure of state overlap.  $F(\rho_1, \rho_2) = (\text{Tr} \{ \sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}} \})^2$ , which - for  $\rho_1$  or  $\rho_2$  pure - simplifies to  $\text{Tr} \{ \rho_1\rho_2 \}$  [19].

*Tangle* The concurrence and tangle are measures of the quantum-coherence properties of a quantum state [34]. For two qubits<sup>6</sup>, concurrence is defined as follows: consider the non-Hermitian matrix  $\hat{R} = \hat{\rho}\hat{\Sigma}\hat{\rho}^T\hat{\Sigma}$  where the superscript T denotes transpose and the ‘spin flip matrix’  $\hat{\Sigma}$  is defined by:

$$\hat{\Sigma} \equiv \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}. \quad (4.29)$$

If the eigenvalues of  $\hat{R}$ , arranged in decreasing order, are given by  $r_1 \geq r_2 \geq r_3 \geq r_4$ , then the concurrence is defined by

$$C = \text{Max} \{0, \sqrt{r_1} - \sqrt{r_2} - \sqrt{r_3} - \sqrt{r_4}\}. \quad (4.30)$$

The tangle is calculated directly from the concurrence:

$$T = C^2. \quad (4.31)$$

The tangle (and the concurrence) range from 0 for product states (or, more generally, any incoherent mixture of product states) to a maximum value of 1 for Bell states.

*Entropy and the Linear Entropy* The Von Neuman entropy quantifies the degree of mixture in a quantum state, and is given by

$$S \equiv -\text{Tr} \{ \hat{\rho} \ln [\hat{\rho}] \} = -\sum_i p_i \ln \{ p_i \}, \quad (4.32)$$

where the  $p_i$  are the eigenvalues of  $\rho$ . The linear entropy [1] is a more analytically convenient form of the same quantity. The linear entropy for a two-qubit system is defined by:

---

<sup>6</sup> The analysis in this subsection applies to the two qubit case only. Measures of entanglement for mixed  $n$ -qubit systems are a subject of on-going research: see, for example, [35] for a recent survey. It may be possible to measure entanglement directly, without quantum state tomography; this possibility was investigated in [36].

$$\begin{aligned} S_L &= \frac{4}{3} (1 - \text{Tr} \{ \hat{\rho}^2 \}) \\ &= \frac{4}{3} \left( 1 - \sum_{a=1}^4 p_a^2 \right), \end{aligned} \quad (4.33)$$

where  $p_a$  are the eigenvalues of  $\rho$ .  $S_L$  ranges from 0 for pure states to 1 for the completely mixed state.

### Multiple Qubit Stokes Parameters

Extending the single-qubit density matrix representation (equation 4.9), any  $n$ -qubit state  $\hat{\rho}$  may be represented as

$$\hat{\rho} = \frac{1}{2^n} \sum_{i_1, i_2, \dots, i_n=0}^3 S_{i_1, i_2, \dots, i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \dots \otimes \hat{\sigma}_{i_n}. \quad (4.34)$$

Normalization requires that  $S_{0,0,\dots,0} = 1$ , allowing  $4^n - 1$  real parameters (the multiple-qubit analog of the single-qubit Stokes parameters) to identify any point in Hilbert space, just as three parameters determined the exact position of a one-qubit state in Bloch/Poincaré space. Already for two qubits, the state space is much larger, requiring 15 independent real parameters to describe it. For this reason, there is no convenient graphical picture of this space, as there was in the single-qubit case (see, however, the interesting approaches made by Zyczkowski [31]).

For multiple qubits the link between the multiple-qubit Stokes parameters [28, 30] and measurement probabilities still exists. The formalism of  $\hat{\tau}$  operators also still holds for larger qubit systems, so that

$$T = \text{Tr} \{ \hat{\tau} \hat{\rho} \}. \quad (4.35)$$

For ‘local’ measurements (a local measurement is the tensor product of a number of single-qubit measurements: the first projecting qubit one along  $\hat{\tau}_{i_1}$ , the second qubit two along  $\hat{\tau}_{i_2}$ , etc.),  $\hat{\tau} = \hat{\tau}_{i_1} \otimes \hat{\tau}_{i_2} \otimes \dots \otimes \hat{\tau}_{i_n}$ . Combining equations 4.34 and 4.35,

$$\begin{aligned} T_{i_1, i_2, \dots, i_n} &= \text{Tr} \{ (\hat{\tau}_{i_1} \otimes \hat{\tau}_{i_2} \otimes \dots \otimes \hat{\tau}_{i_n}) \hat{\rho} \} \\ &= \frac{1}{2^n} \sum_{j_1, j_2, \dots, j_n=0}^3 \text{Tr} \{ \hat{\tau}_{i_1} \hat{\sigma}_{j_1} \} \text{Tr} \{ \hat{\tau}_{i_2} \hat{\sigma}_{j_2} \} \dots \text{Tr} \{ \hat{\tau}_{i_n} \hat{\sigma}_{j_n} \} S_{j_1, j_2, \dots, j_n}. \end{aligned} \quad (4.36)$$

Recall that for single qubits,

$$\begin{aligned} T_{i=1,2,3} &= P_{|\psi_i\rangle} - P_{|\psi_i^\perp\rangle} \\ T_0 &= P_{|\psi\rangle} + P_{|\psi^\perp\rangle}, \forall \psi \end{aligned} \quad (4.37)$$

Therefore, for an  $n$ -qubit system,

$$T_{i_1, i_2, \dots, i_n} = \\ (P_{|\psi_{i_1}\rangle} \pm P_{|\psi_{i_1}^\perp\rangle}) \otimes (P_{|\psi_{i_2}\rangle} \pm P_{|\psi_{i_2}^\perp\rangle}) \otimes \dots \otimes (P_{|\psi_{i_n}\rangle} \pm P_{|\psi_{i_n}^\perp\rangle}), \quad (4.38)$$

where the plus sign is used for a 0 index and the minus sign is used for a nonzero index. For a two-qubit system where  $i_1 \neq 0$  and  $i_2 \neq 0$ ,  $T_{i_1, i_2}$  simplifies dramatically, giving

$$T_{i_1, i_2} = (P_{|\psi_{i_1}\rangle} - P_{|\psi_{i_1}^\perp\rangle}) \otimes (P_{|\psi_{i_2}\rangle} - P_{|\psi_{i_2}^\perp\rangle}) \\ = P_{|\psi_{i_1}\rangle|\psi_{i_2}\rangle} - P_{|\psi_{i_1}\rangle|\psi_{i_2}^\perp\rangle} - P_{|\psi_{i_1}^\perp\rangle|\psi_{i_2}\rangle} + P_{|\psi_{i_1}^\perp\rangle|\psi_{i_2}^\perp\rangle}. \quad (4.39)$$

This relation will be crucial for rebuilding a two-qubit state from local measurements.

As before, we are not restricted to multiple-qubit Stokes parameters based on orthogonal operators. Extending equation 4.21 to multiple qubits, and again assuming two representations  $S_{i_1, i_2, \dots, i_n} = \text{Tr} \{(\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \dots \otimes \hat{\sigma}_{i_n}) \hat{\rho}\}$ , and  $T_{i_1, i_2, \dots, i_n} = \text{Tr} \{(\hat{\tau}_{i_1} \otimes \hat{\tau}_{i_2} \otimes \dots \otimes \hat{\tau}_{i_n}) \hat{\rho}\}$ ,

$$T_{i_1, i_2, \dots, i_n} = \\ \frac{1}{2^n} \sum_{j_1, j_2, \dots, j_n=0}^3 \text{Tr} \{(\hat{\tau}_{i_1} \otimes \hat{\tau}_{i_2} \otimes \dots \otimes \hat{\tau}_{i_n}) (\hat{\sigma}_{j_1} \otimes \hat{\sigma}_{j_2} \otimes \dots \otimes \hat{\sigma}_{j_n})\} S_{j_1, j_2, \dots, j_n}. \quad (4.40)$$

#### *Example: Two-Qubit Polarization States*

Consider the state  $|HH\rangle$ . Following the example in equation 4.20,

$$\hat{\rho}_{HH} = |HH\rangle\langle HH| \\ = \frac{1}{2}(\hat{\sigma}_0 + \hat{\sigma}_3) \otimes \frac{1}{2}(\hat{\sigma}_0 + \hat{\sigma}_3) \\ = \frac{1}{4}(\hat{\sigma}_0 \otimes \hat{\sigma}_0 + \hat{\sigma}_3 \otimes \hat{\sigma}_0 + \hat{\sigma}_0 \otimes \hat{\sigma}_3 + \hat{\sigma}_3 \otimes \hat{\sigma}_3). \quad (4.41)$$

This implies that there are exactly four non-zero two-qubit Stokes parameters:  $S_{0,0}, S_{0,3}, S_{3,0}$ , and  $S_{3,3}$  – all of which are equal to one. (As earlier, for the special case when  $\hat{\tau}_{i,j} = \hat{\sigma}_{i,j}$ , we relabel the  $T_{i,j}$  as  $S_{i,j}$ , the two-qubit Stokes parameters [28, 30].) The separable nature of this state makes it easy to calculate the two-qubit Stokes decomposition.

If instead we investigate an entangled state,  $|\psi^-\rangle$ , it will be necessary to calculate each two-qubit Stokes parameter from the  $\hat{\sigma}$  matrices. As an example, consider  $\hat{\sigma}_{3,3} \equiv \hat{\sigma}_3 \otimes \hat{\sigma}_3$ , for which

$$S_{3,3} = \text{Tr} \{\hat{\sigma}_{3,3} |\psi^-\rangle\langle \psi^-|\} = -1. \quad (4.42)$$

We could instead calculate  $S_{3,3}$  directly from probability outcomes of measurements on  $|\psi^-\rangle$ :

$$\begin{aligned}
S_{3,3} &= (P_H - P_V) \otimes (P_H - P_V) \\
&= P_{HH} - P_{HV} - P_{VH} + P_{VV} \\
&= 0 - \frac{1}{2} - \frac{1}{2} + 0 \\
&= -1.
\end{aligned} \tag{4.43}$$

In general, a given  $\hat{\tau}$  operator is not uniquely mapped to a single pair of analysis states. For example, if the analysis states are  $|\psi_1\rangle \equiv |H\rangle$  and  $|\psi_2\rangle \equiv |V\rangle$ ,  $\hat{\tau}_{1,1} \equiv \hat{\sigma}_3 \otimes \hat{\sigma}_3 = -\hat{\sigma}_3 \otimes -\hat{\sigma}_3 \equiv \hat{\tau}_{2,2}$ . However, this would be an unacceptable choice for  $|\psi_1\rangle$  and  $|\psi_2\rangle$ , as their respective density matrices are not linearly independent.

Continuing on, we measure  $S_{0,3}$ :

$$\begin{aligned}
S_{0,3} &= (P_H + P_V) \otimes (P_H - P_V) \\
&= P_{HH} - P_{HV} + P_{VH} - P_{VV} \\
&= 0 - \frac{1}{2} + \frac{1}{2} - 0 \\
&= 0.
\end{aligned} \tag{4.44}$$

Here the signs of the probabilities changed due to the zero index in  $S_{0,3}$ . These results would have been the same even if the analysis bases of the first qubit had been shifted to any other orthogonal basis, i.e.,  $S_{0,3} = (P_\psi + P_{\psi^\perp}) \otimes (P_H - P_V)$ .

If the method above is continued for all the Stokes parameters, one concludes that

$$\begin{aligned}
\hat{\rho}_{\psi^-} &= \frac{1}{2}(|HV\rangle - |VH\rangle)(\langle HV| - \langle VH|) \\
&= \frac{1}{4}(\hat{\sigma}_0 \otimes \hat{\sigma}_0 - \hat{\sigma}_1 \otimes \hat{\sigma}_1 - \hat{\sigma}_2 \otimes \hat{\sigma}_2 - \hat{\sigma}_3 \otimes \hat{\sigma}_3).
\end{aligned} \tag{4.45}$$

## 4.2 Exact Tomography

The goal of tomography is to reconstruct the density matrix of an unknown ensemble of particles through a series of measurements. In practice, this can never be performed exactly, as an infinite number of particles would be required to eliminate statistical error. If exact measurements were taken on infinite ensembles, each measurement would yield an exact probability of success, which could then be used to reconstruct a density matrix. Though unrealistic, it is highly illustrative to examine this exact tomography before seeing the more general treatment. Hence, this section will treat all measurements as yielding exact probabilities, and ignore all sources of error in those measurements.

#### 4.2.1 Single Qubit Tomography

Although reconstructive tomography of any size system follows the same general procedure, beginning with tomography of a single qubit allows the visualization of each step using the Poincaré sphere, in addition to providing a simpler mathematical introduction.

##### Visualization of Single Qubit Tomography

Exact single-qubit tomography requires a sequence of three linearly independent measurements. Each measurement exactly specifies one degree of freedom for the measured state, reducing the free parameters of the unknown state's possible Hilbert space by one.

As an example, consider measuring R, D, and H on the mixed state

$$\hat{\rho} = \begin{pmatrix} \frac{5}{8} & \frac{-i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{3}{8} \end{pmatrix} \quad (4.46)$$

Rewriting the state (using equation 4.9) as

$$\hat{\rho} = \frac{1}{2} \left( \hat{\sigma}_0 + \frac{1}{\sqrt{2}} \hat{\sigma}_2 + \frac{1}{4} \hat{\sigma}_3 \right) \quad (4.47)$$

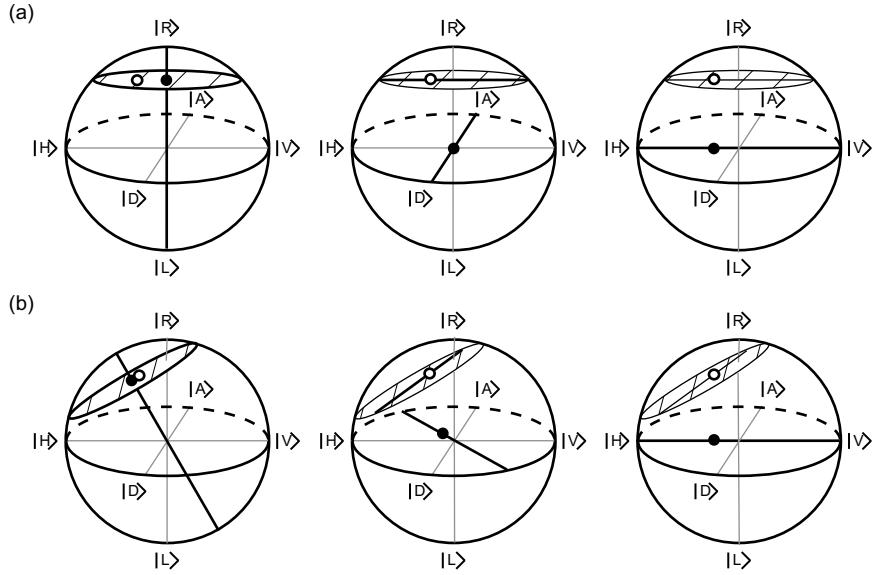
allows us to read off the normalized Stokes parameters corresponding to these measurements:

$$S_1 = 0, S_2 = \frac{1}{\sqrt{2}}, \text{ and } S_3 = \frac{1}{4}. \quad (4.48)$$

As always,  $S_0 = 1$  due to normalization. Measuring R first, and looking to the Poincaré sphere, we see that the unknown state must lie in the  $z = \frac{1}{\sqrt{2}}$  plane (as  $S_2 = \frac{1}{\sqrt{2}}$ ). A measurement in the D basis further constrains the state to the  $y = 0$  plane, resulting in a total confinement to a line parallel to and directly above the  $x$  axis. The final measurement of H pinpoints the state. This process is illustrated in figure 4.2a. Obviously the order of the measurements is irrelevant: it is the intersection point of three orthogonal planes that defines the location of the state.

If instead measurements are made along *non*-orthogonal axes, a very similar picture develops, as indicated in figure 4.2b. The first measurement always isolates the unknown state to a plane, the second to a line, and the third to a point.

Of course, in practice, the experimenter has no knowledge of the unknown state before a tomography. The set of the measured probabilities, transformed into the Stokes parameters as above, allow a state to be directly reconstructed.



**Fig. 4.2.** A sequence of three linearly independent measurements isolates a single quantum state in Hilbert space (shown here as an open circle in the Poincaré sphere representation). The first measurement isolates the unknown state to a plane perpendicular to the measurement basis. Further measurements isolate the state to the intersections of non-parallel planes, which for the second and third measurements correspond to a line and finally a point. The black dots shown correspond to the projection of the unknown state onto the measurement axes, which determines the position of the aforementioned planes. (a) A sequence of measurements along the right-circular, diagonal, and horizontal axes. (b) A sequence of measurements on the same state taken using non-orthogonal projections: elliptical light rotated 30° from H towards R, 22.5° linear, and horizontal.

### A Mathematical Look at Single Qubit Tomography

Using the tools developed in the first section of this chapter, single-qubit tomography is relatively straightforward. Recall equation 4.9,  $\hat{\rho} = \frac{1}{2} \sum_{i=0}^3 S_i \hat{\sigma}_i$ . Considering that  $S_1$ ,  $S_2$ , and  $S_3$  completely determine the state, we need only measure them to complete the tomography. As  $S_{j>0} = 2P_{|\psi\rangle} - 1$  (equation 4.13), three measurements respectively in the  $|0\rangle$ ,  $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ , and  $\frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$  bases will completely specify the unknown state. If instead measurements are made in another basis, even a non-orthogonal one, they can be easily related back to the  $S_i$  parameters, and therefore the density matrix, by means of equation 4.21.

While this procedure is straightforward, there is one subtlety which will become important in the multiple-qubit case. Projective measurements generally refer to the measurement of a single basis state and return a single value

between zero and one. This corresponds, for example, to an electron beam passing through a Stern-Gerlach apparatus with a detector placed at one output. While a single detector and knowledge of the input particle intensity will – in the one-qubit case – completely determine a single Stokes parameter, one could collect data from both outputs of the Stern-Gerlach device. This would measure the probability of projecting not only onto the state  $|\psi\rangle$ , but also onto  $|\psi^\perp\rangle$ , and without needing to know the input intensity. All physical measurements on single qubits, regardless of implementation, can in principle be measured this way (though in practice measurements of some qubit systems may typically detect a population in only *one* of the states [32]). We will see below that although one detector functions as well as two in the single-qubit case, this situation will not persist into higher dimensions.

#### *Arbitrary Measurements Using Waveplates and Polarizers*

An arbitrary polarization measurement and its orthogonal compliment can be realized using, in order, a quarter-wave plate, a half-waveplate, and a polarizing beam splitter. Waveplates implement unitary operations, and in the Poincaré sphere picture, act as rotations about an axis lying within the linear polarization plane (the equator) [26]. Specifically, a waveplate whose optic axis is oriented at angle  $\theta$  with respect to the horizontal induces a rotation on the Poincaré sphere about an axis  $2\theta$  from horizontal, in the linear plane. The magnitude of this rotation is equal to the waveplate's retardance ( $90^\circ$  for quarter-wave plates and  $180^\circ$  for half-wave plates). For the remainder of this chapter we adopt the convention that polarizing beam splitters transmit horizontally polarized light and reflect vertically polarized light.

This analysis, while framed in terms of waveplates acting on photon polarization, is directly applicable to other systems, e.g., spin- $\frac{1}{2}$  particles [11–14] or two-level atoms [15, 16]. In these systems, measurements in arbitrary bases are obtained using suitably phased  $\pi$ - and  $\frac{\pi}{2}$ -pulses (externally applied electromagnetic fields) to rotate the state to be measured into the desired analysis basis.

To derive the settings for these waveplates as a function of the projection state desired, we use the Poincaré sphere (see figure 4.3). For any state on the surface of the sphere, a  $90^\circ$  rotation about a linear axis directly below it will rotate that state into a linear polarization (see figure 4.3b). Assume the desired projection state is

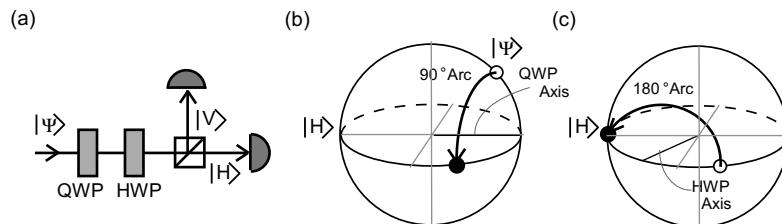
$$|\psi_P\rangle = \cos\left(\frac{\theta}{2}\right)|H\rangle + \sin\left(\frac{\theta}{2}\right)e^{i\phi}|V\rangle. \quad (4.49)$$

Simple coordinate transforms from spherical to cartesian coordinates reveal that a quarter-waveplate at  $\theta_{QWP} = \frac{1}{2}\arccos\{\sin(\theta)\tan(\phi)\}$  will rotate the projection state (4.49) into a linear state

$$|\psi'_P\rangle = \cos\left(\frac{\theta'}{2}\right)|H\rangle + \sin\left(\frac{\theta'}{2}\right)|V\rangle. \quad (4.50)$$

A half-waveplate at  $\frac{1}{4}\theta'$  will rotate this state to  $|H\rangle$ .<sup>7</sup> The PBS will then transmit the projected state and reflect its orthogonal compliment.

Of course, these calculations assume that waveplates with retardances equal to exactly  $\pi$  or  $\frac{\pi}{2}$  are used (or Rabi pulses producing perfect phase differences). Imperfect yet well characterized waveplates will lead to measurements in slightly different yet known bases. This will still yield an accurate tomography, but first these results must be transformed from a non-orthogonal basis into the canonical Stokes parameters using equation 4.21. Later, the maximum likelihood technique (see section 3) will provide a different but equally effective way to accomodate for imperfect measurements.



**Fig. 4.3.** A quarter-waveplate (QWP), half-waveplate (HWP), and polarizing beam splitter (PBS) are used to make an arbitrary polarization measurement. Both a diagram of the experimental apparatus (a) and the step-by-step evolution of the state on the Poincaré sphere are shown. (b) The quarter-waveplate rotates the projection state (the state we are projecting into, *not* the incoming unknown state) into the linear polarization plane (the equator). (c) The half-waveplate rotates this linear state to horizontal. The PBS transmits the projection state (now  $|H\rangle$ ) and reflects its orthogonal compliment (now  $|V\rangle$ ), which can then both be measured.

### Exact Tomography of Multiple Qubits Using $2^n$ Detectors

Tomography of multiple qubits, though an extension of the single-qubit technique, becomes more complicated and depends on the experimental apparatus used. The simplest, fastest, and most intuitive version of this tomography uses an array of  $2^n$  detectors, which project every incoming  $n$ -qubit state into one of  $2^n$  basis states. This is the generalization of simultaneously measuring both outputs in the single qubit case. These detectors must measure in  $n$ -fold coincidence, and for the purposes of exact tomography it is assumed they have no errors and operate on an infinite ensemble of states. It should be emphasized that these additional detectors are not some ‘trick’, effectively

<sup>7</sup>  $\theta' = \text{acos} \{ \sin(\theta)\tan(\phi) \} - \text{acos} \{ \cot(\theta)\cot(\phi) \}$ . In practice, care must be taken that consistent conventions are used (e.g., right vs. left circular polarization), and it may be easier to calculate this angle directly from waveplate operators and the initial state.

masking a number of sequential settings of  $n$  detectors. If only  $n$  detectors are used, then over the course of a tomography most states within the input ensemble will never be measured. For example, consider measuring the projection of an unknown state into the  $|00\rangle$  basis using two detectors. While this will give a number of counts, unmeasured coincidences will be routed into the  $|01\rangle$ ,  $|10\rangle$ , and  $|11\rangle$  modes. The information of how many coincidences are routed to which mode will be lost, unless another two detectors are in place in the ‘1’ modes to measure it.

The primary advantage to using  $2^n$  detectors is that every setting of the analysis system (every group of the projector and its orthogonal compliments) generates exactly enough information to define a single multiple-qubit Stokes vector. Expanding out the probabilities that a multiple-qubit Stokes vector (which for now we will limit to those with only non-zero indices) is based on,

$$\begin{aligned} S_{i_1, i_2, \dots, i_n} &= (P_{\psi_1} - P_{\psi_1^\perp}) \otimes (P_{\psi_2} - P_{\psi_2^\perp}) \otimes \dots \otimes (P_{\psi_n} - P_{\psi_n^\perp}) \\ &= P_{\psi_1, \psi_2, \dots, \psi_n} - P_{\psi_1, \psi_2, \dots, \psi_n^\perp} - \dots \pm P_{\psi_1^\perp, \psi_2^\perp, \dots, \psi_n^\perp}, \end{aligned} \quad (4.51)$$

where the sign of each term on the last line is determined by the parity of the number of orthogonal ( $\perp$ ) terms. (If instead we had included zero indices, each zero index in  $S$  would correspond to a plus sign in the first line of 4.51.)

These probabilities are precisely those measured by a single setting of the entire analysis system followed by a  $2^n$  detector array. Returning to our primary decomposition of the density matrix (equation 4.34),

$$\hat{\rho} = \frac{1}{2^n} \sum_{i_1, i_2, \dots, i_n=0}^3 S_{i_1, i_2, \dots, i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \dots \otimes \hat{\sigma}_{i_n},$$

we once again need only determine all of the multiple-qubit Stokes parameters to exactly characterize the density matrix. At first glance this might seem to imply that we need to use  $4^n - 1$  settings of the analysis system, in order to find all of the multiple-qubit Stokes parameters save  $S_{0,0,\dots,0}$ , which is always one.

While this is certainly sufficient to solve for  $\hat{\rho}$ , many of these measurements are redundant. In order to choose the smallest possible number of settings, note that the probabilities that constitute some multiple-qubit Stokes parameters overlap exactly with the probabilities for other multiple-qubit Stokes parameters. Specifically, any multiple-qubit Stokes parameter with at least one 0 subscript is derived from a set of probabilities that at least one other multiple-qubit Stokes vector (with no 0 subscripts) is also derived from. As an example, consider that

$$S_{0,3} = P_{|00\rangle} - P_{|01\rangle} + P_{|10\rangle} - P_{|11\rangle}, \quad (4.52)$$

while

$$S_{3,3} = P_{|00\rangle} - P_{|01\rangle} - P_{|10\rangle} + P_{|11\rangle}. \quad (4.53)$$

The same analysis settings will provide enough information to determine both values. This dependent relationship between multiple-qubit Stokes vectors is true in general, as can be seen by returning to equation 4.51. Each non-zero subscript for  $S$  contributes a term to the tensor product on the right that looks like  $(P_{\psi_i} - P_{\psi_i^\perp})$ . Had there been zero subscripts, however, they each would have contributed a  $(P_{\psi_i} + P_{\psi_i^\perp})$  term, which would have been totally redundant with any other term that has a non-zero term in the same index. This reduces the minimum number of analysis settings to  $3^n$ , a huge improvement in multiple qubit systems (e.g., 9 vs. 15 settings for 2-qubit tomography, 81 vs. 255 for 4-qubit tomography, etc.). Note that, as discussed earlier, this benefit is only possible if one employs  $2^n$  detectors, leading to a total of  $6^n$  measurements.<sup>8</sup>

Because equation 4.41 can be used to transform any set of non-orthogonal multiple-qubit Stokes parameters into the canonical form, orthogonal measurement sets need not be used. One advantage of the option to use non-orthogonal measurement sets is that an orthogonal set may not be experimentally achievable, for instance, due to waveplate imperfections.

*Example: A Complete Ideal 2-Qubit Tomography of Photon Pairs*

Consider measuring a state with nine settings of the apparatus and four detectors, for a total of 36 measurement results. The results for this example are compiled below, with each row representing a single setting of the apparatus, and therefore a single two-qubit Stokes parameter. All measured probabilities are non-negative, but the results below have been given a minus sign if they are negated when summed into their respective two-qubit Stokes parameter.

---

<sup>8</sup> These measurements, even though they result from the minimum number of analysis settings for  $2^n$  detectors, are overcomplete. A density matrix has only  $4^n - 1$  free parameters, which implies that only  $4^n - 1$  measurements are necessary to specify it (see  $n$ -detector tomography). Because the overcomplete set of  $6^n$  measurements is not linearly independent, they can be reduced to a  $4^n - 1$  element subset and still completely specify an unknown state.

$$\begin{aligned}
S_{1,1} &= \frac{+P_{DD}}{\frac{1}{3}} - \frac{P_{DA}}{\frac{1}{6}} - \frac{P_{AD}}{\frac{1}{6}} + \frac{P_{AA}}{\frac{1}{3}} = \frac{1}{3} \\
S_{1,2} &= \frac{+P_{DR}}{\frac{1}{4}} - \frac{P_{DL}}{\frac{1}{4}} - \frac{P_{AR}}{\frac{1}{4}} + \frac{P_{AL}}{\frac{1}{4}} = 0 \\
S_{1,3} &= \frac{+P_{DH}}{\frac{1}{4}} - \frac{P_{DV}}{\frac{1}{4}} - \frac{P_{AH}}{\frac{1}{4}} + \frac{P_{AV}}{\frac{1}{4}} = 0 \\
S_{2,1} &= \frac{+P_{RD}}{\frac{1}{4}} - \frac{P_{RA}}{\frac{1}{4}} - \frac{P_{LD}}{\frac{1}{4}} + \frac{P_{LA}}{\frac{1}{4}} = 0 \\
S_{2,2} &= \frac{+P_{RR}}{\frac{1}{6}} - \frac{P_{RL}}{\frac{1}{3}} - \frac{P_{LR}}{\frac{1}{3}} + \frac{P_{LL}}{\frac{1}{6}} = -\frac{1}{3} \\
S_{2,3} &= \frac{+P_{RH}}{\frac{1}{4}} - \frac{P_{RV}}{\frac{1}{4}} - \frac{P_{LH}}{\frac{1}{4}} + \frac{P_{LV}}{\frac{1}{4}} = 0 \\
S_{3,1} &= \frac{+P_{HD}}{\frac{1}{4}} - \frac{P_{HA}}{\frac{1}{4}} - \frac{P_{VD}}{\frac{1}{4}} + \frac{P_{VA}}{\frac{1}{4}} = 0 \\
S_{3,2} &= \frac{+P_{HR}}{\frac{1}{4}} - \frac{P_{HL}}{\frac{1}{4}} - \frac{P_{VR}}{\frac{1}{4}} + \frac{P_{VL}}{\frac{1}{4}} = 0 \\
S_{3,3} &= \frac{+P_{HH}}{\frac{1}{3}} - \frac{P_{HV}}{\frac{1}{6}} - \frac{P_{VH}}{\frac{1}{6}} + \frac{P_{VV}}{\frac{1}{3}} = \frac{1}{3}
\end{aligned} \tag{4.54}$$

Measurements are taken at each of these nine settings, directly determining the above nine two-qubit Stokes Parameters. The six remaining required parameters, listed below, are dependent upon the same measurements.

$$\begin{aligned}
S_{0,1} &= \frac{+P_{DD}}{\frac{1}{3}} - \frac{P_{DA}}{\frac{1}{6}} + \frac{P_{AD}}{\frac{1}{6}} - \frac{P_{AA}}{\frac{1}{3}} = 0 \\
S_{0,2} &= \frac{+P_{RR}}{\frac{1}{6}} - \frac{P_{LR}}{\frac{1}{3}} + \frac{P_{RL}}{\frac{1}{3}} - \frac{P_{LL}}{\frac{1}{6}} = 0 \\
S_{0,3} &= \frac{+P_{HH}}{\frac{1}{3}} - \frac{P_{HV}}{\frac{1}{6}} + \frac{P_{VH}}{\frac{1}{6}} - \frac{P_{VV}}{\frac{1}{3}} = 0 \\
S_{1,0} &= \frac{+P_{DD}}{\frac{1}{3}} + \frac{P_{DA}}{\frac{1}{6}} - \frac{P_{AD}}{\frac{1}{6}} - \frac{P_{AA}}{\frac{1}{3}} = 0 \\
S_{2,0} &= \frac{+P_{RR}}{\frac{1}{6}} + \frac{P_{LR}}{\frac{1}{3}} - \frac{P_{RL}}{\frac{1}{3}} - \frac{P_{LL}}{\frac{1}{6}} = 0 \\
S_{3,0} &= \frac{+P_{HH}}{\frac{1}{3}} + \frac{P_{HV}}{\frac{1}{6}} - \frac{P_{VH}}{\frac{1}{6}} - \frac{P_{VV}}{\frac{1}{3}} = 0
\end{aligned} \tag{4.55}$$

These terms will not in general be zero. Recall (c.f. equation 4.41) that for  $|HH\rangle$ ,  $S_{0,3} = S_{3,0} = 1$ . Of course,  $S_{0,0} = 1$ . Taken together, these two-qubit Stokes parameters determine the density matrix.

$$\begin{aligned}
\hat{\rho} &= \frac{1}{4} \left( \hat{\sigma}_0 \otimes \hat{\sigma}_0 + \frac{1}{3} \hat{\sigma}_1 \otimes \hat{\sigma}_1 - \frac{1}{3} \hat{\sigma}_2 \otimes \hat{\sigma}_2 + \frac{1}{3} \hat{\sigma}_3 \otimes \hat{\sigma}_3 \right) \\
&= \frac{1}{6} \begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} + \frac{1}{6} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
\end{aligned} \tag{4.56}$$

This is the final density matrix, a Werner State, as defined in equation 4.28.

### Exact Tomography of Multiple Qubits Using $n$ Detectors

While the scheme outlined above is most efficient in the sense that it requires only  $3^n$  analysis settings for the experimental apparatus, in practice it may only be experimentally possible to make a single projective measurement at a time (e.g., because the analyzer can monitor only a single outcome). As discussed earlier, this will require  $4^n - 1$  probabilities in order to define a complete set of  $T_i$  parameters. In practice, this will mean that  $4^n$  measurements are necessary in order to normalize counts to probabilities. By making a set of single projective measurements on each qubit and only taking into account those results where a definite result is obtained (e.g., the photon was transmitted by the polarizer), it is possible to reconstruct a state using only  $n$  detectors.

First define a set of  $2^n \times 2^n$  matrices which have the following properties:

$$\begin{aligned} \text{Tr} \left\{ \hat{\Gamma}_\nu \cdot \hat{\Gamma}_\mu \right\} &= \delta_{\nu,\mu} \\ \hat{A} &= \sum_\nu \hat{\Gamma}_\nu \text{Tr} \left\{ \hat{\Gamma}_\nu \cdot \hat{A} \right\} \quad \forall \hat{A}, \end{aligned} \quad (4.57)$$

where  $\hat{A}$  is an arbitrary  $2^n \times 2^n$  matrix. A convenient set of  $\hat{\Gamma}$  matrices to use are the tensor product of  $\hat{\sigma}$  matrices used throughout this paper:

$$\hat{\Gamma}_\nu = \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \dots \otimes \hat{\sigma}_{i_n}, \quad (4.58)$$

where  $\nu$  is simply a short-handed index by which to label the  $\Gamma$  matrices, which is simpler than writing out  $i_1, i_2, \dots, i_n$ . Given this notation, we substitute into equation 4.34 and find that

$$\hat{\rho} = \frac{1}{2^n} \sum_\nu \hat{\Gamma}_\nu S_\nu. \quad (4.59)$$

Now that the density matrix is represented in a useful form, it is necessary to consider exactly which measurements to use. In particular, we now wish to determine the necessary and sufficient conditions on the  $4^n$  measurements.<sup>9</sup> Let  $|\psi_\mu\rangle$  ( $\mu = 1$  to  $4^n$ ) be the measurement bases. We can then define the probability of the  $\mu^{\text{th}}$  measurement as  $P_\mu = \langle \psi_\mu | \hat{\rho} | \psi_\mu \rangle$ .

---

<sup>9</sup> While only  $4^n - 1$  measurements are necessary, in practice it is often the case that exact probabilities are not known, only numbers of counts (successful measurements), with no information about the number of counts which would have been measured by detectors in orthogonal bases. In this case an extra measurement is necessary to normalize the inferred probabilities. If instead exact probabilities are known, the final result need not actually be measured, but instead can be calculated from three of the other measurements (choose this last measurement to be the  $n^{\text{th}}$  in a complete basis of measurements, and use the normalization constraint that all probabilities in a complete basis must sum to one).

Combining this with equation 4.59,

$$P_\mu = \langle \psi_\mu | \frac{1}{2^n} \sum_{\nu=1}^{4^n} \hat{\Gamma}_\nu S_\nu | \psi_\mu \rangle = \frac{1}{2^n} \sum_{\nu=1}^{4^n} B_{\mu,\nu} S_\nu, \quad (4.60)$$

where the  $4^n \times 4^n$  matrix  $B_{\mu,\nu}$  is given by

$$B_{\mu,\nu} = \langle \psi_\mu | \hat{\Gamma}_\nu | \psi_\mu \rangle. \quad (4.61)$$

Immediately we find a necessary and sufficient condition for the completeness of the set of tomographic states  $\{|\psi_\mu\rangle\}$ : if the matrix  $B_{\mu,\nu}$  is nonsingular, then eq.(4.60) can be inverted to give

$$S_\nu = 2^n \sum_{\mu=1}^{4^n} (B^{-1})_{\mu,\nu} P_\mu. \quad (4.62)$$

As before, the density matrix can be calculated directly from the  $S_\nu$  values. Obviously, this reconstruction technique would also work with  $2^n$  detectors, as the measurements gained from  $3^n$  analysis settings are more than sufficient to define the  $P_\mu$  above.

### 4.3 Real Tomography: Errors and the Maximum Likelihood Technique

The tools developed so far allow the perfect reconstruction of a density matrix from an infinite set of ideal data. When applying this technique to any set of real measurements, the assumption of ideal probabilities must be discarded. In fact, the probabilities predicted by real results can in practice be contradictory or even physically impossible. It is therefore necessary to implement a procedure that takes these errors into account yet always returns a legitimate density matrix.

#### 4.3.1 Real Tomography of a Single Qubit

Once again, the single-qubit example precedes the general treatment in order to provide intuition and a graphical picture.

##### Types of Errors

Errors in the measurement of a density matrix fall into three main categories: errors in the measurement basis, errors from counting statistics, and errors from experimental stability. The first problem can be addressed by increasing the accuracy of the measurement apparatus (e.g., obtaining higher tolerance

waveplates, better controlling the Rabi pulses, etc.) while the second problem is reduced by performing each measurement on a larger ensemble (counting for a longer time). The final difficulty is drift which occurs over the course of the tomography.<sup>10</sup> This drift occurs either in the state produced or the efficiency of the detection system, and can constrain the data-collection time.

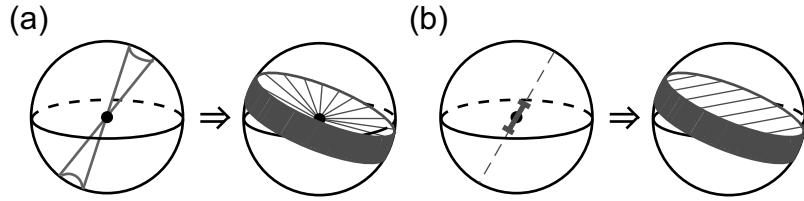
Figure 4.4a shows what a basis error looks like on the Poincaré sphere and how that error affects the ability to isolate a state in Poincaré space. Accidentally measuring in a different basis, slightly different than the intended one, introduces a different amount of error depending on the state being measured. All possible measurement axes must pass through the center of the Poincaré sphere, corresponding to the fact that the totally mixed state has a probability of  $\frac{1}{2}$  of being measured in any basis. Some Gaussian distribution of error around the intended axis looks like two cones meeting point to point at the center of the sphere, traced out by all of the nearby measurement axes. When translated into knowledge of the state, this transforms what would have been a single plane (perpendicular to the measurement axis) into a disk, thick at the edges of the sphere and thinning to a single point at the origin. This picture indicates that a basis error is more pronounced when measuring a pure state, but actually has no effect when measuring a totally mixed state (because all bases give the same answer).

Figure 4.4b shows the same analysis of errors in counting statistics. Any real measurement can only be carried out on a limited size ensemble. Though the details of the statistics will be dealt with later, the detection events are accurately described by a Poissonian distribution, which for large numbers of counts is well approximated by a Gaussian distribution. This will cause the resultant knowledge about the unknown state to change from a plane (in the exact case) to a thick disk (uniformly thick for pure and mixed states), a one-dimensional Gaussian distribution plotted in three-dimensional space.

The final source of error, that of drift, can occur on three distinct timescales. Quickly varying systems experience drift which cycles many times within the span of a single measurement. Rather than producing errors, these types of drift result in the measurement of a mixed state (which describes a probabilistic mixture of several states, exactly the situation when a state drifts). Very slowly varying systems also do not introduce errors, as the timescale of the drift exceeds the timescale of the measurements (during the measurement process, the state is essentially constant). The final source of drift occurs on the timescale of the measurement, allowing a totally differ-

---

<sup>10</sup> These are the main sources of error that are likely to be present to some degree in *any* qubit implementation. In addition, each implementation may have its own unique errors. For example, if the waveplates used to analyze optical polarization qubits are slightly wedged, and the detectors slightly non-uniform, this can lead to a troubling systematic error whereby the detector efficiency depends on the analysis settings. Other errors can be compensated, such as when accidental coincidence counts due to background light are subtracted from measured count rates. Here we neglect such system-specific difficulties.



**Fig. 4.4.** Graphical representation of errors in a single-qubit tomography. (a) Basis errors. Errors in the setting of measurement apparatus can result in an accurate measurement being taken in an unintended basis. Shown graphically is the effect that an uncertainty in the measurement basis can have on the reconstruction of a state. Instead of a single axis on the Poincaré sphere, the possible measurement axes form uncertainty cones touching at the center, since all possible measurement axes pass through the origin. This uncertainty in axis is then translated into an uncertainty in the state (shown on the right). Instead of isolating the state to a plane, all possible measurement axes trace out a volume with large uncertainty near the surface of the sphere and low uncertainty near the center. (b) Counting errors. Even if the measurement basis is exactly known, only a limited number of qubits can be measured to gain an estimate of a state's projection onto this axis (taken directly from the probability of a successful measurement). This uncertainty results in an unknown state being isolated to a one-dimensional gaussian (approximately) in three-dimensional space, rather than to a plane.

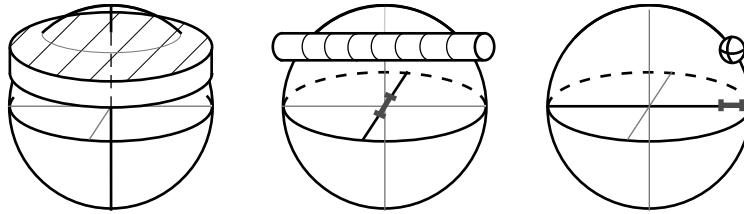
ent state to be measured from one analysis setting to the next. This error is difficult to characterize, as it is difficult to know what the correct answer for a tomography would be, much less how far the predicted value strayed from it.

Drift in state *intensity* (e.g., the rate of photons produced) can be compensated for by employing to a  $2^n$  detector system. Because every member of the ensemble is measured in a complete basis, this gives an exact intensity for each set of  $2^n$  measurements, eliminating the need for the assumption that each measurement run is over the same size ensemble. (Recall that a *measurement* gives the probability to project the unknown  $\rho$  into a single state  $|\psi\rangle$ . When using  $2^n$  detectors, therefore, each analysis setting will yield  $2^n$  measurement results.)

Detector efficiency drift also presents a problem, with similar effects to state drift. Special care must be taken to address this problem when using a  $2^n$  detector scheme, as measurements taken with each pair of detectors may appear to give different values, when in fact a particular detector pair is simply more efficient than another. In order to compensate for this, the relative efficiencies of each detector pair must be measured, and all counts taken renormalized using these efficiencies.

### Combining Real Measurements: State Estimation

After all sources of error are taken into account, a single measurement results in a distribution over all possible states describing the experimenter's knowledge of the unknown state. This distribution represents the likelihood that a particular state would give the measured results, relative to another state. When independent measurements are combined, these distributions are multiplied, and ideally the knowledge of the unknown state is restricted to a small ball in Poincaré space, approximately equal to a three-dimensional Gaussian. This type of state isolation occurs regardless of which measurements are taken, as long as they are linearly independent. State isolation is shown graphically in figure 4.5 for a set of orthogonal measurements.



**Fig. 4.5.** Isolation of a quantum state through inexact measurements. Although a series of real measurements (those with uncertainties) will never be able to exactly isolate an unknown quantum state, they can isolate it to a region of Hilbert space that is far more likely than any other region to contain the unknown state. Consider a series of three measurements, each containing counting errors, along orthogonal axes. From left to right, the area of Hilbert space containing the unknown state is truncated from a one-dimensional Gaussian probability distribution (the disk in the left figure) to a two-dimensional Gaussian (the cylinder in the middle figure) and finally to a three-dimensional Gaussian (the ball in the right figure). This results in an ‘error ball’ which approximates the position of the unknown state. The global maximum, however, can often be outside allowed Hilbert space (outside the Poincaré sphere), which is one reason a maximum likelihood technique must be used to search over only allowed quantum states.

In contrast to the ideal case in the previous section, for which the accuracy of a reconstructed state did not depend on whether orthogonal or non-orthogonal measurements were made, with real measurements the advantage of orthogonal measurements becomes clear. Unlike in figure 4.5, non-orthogonal measurements result in a non-symmetric error ball, increasing the error in state estimation in one direction in Hilbert space. In practice, making the right measurements for the right amount of time to make this error ball spherical can be achieved by adaptive tomography [37], but this is beyond the scope of this work.

Even after tomography returns a distribution of likelihood over Poincaré space, one final problem remains. It is very possible, especially with low

counts or with the measurement of very pure states, that state estimation will return an illegal state. For example, in figure 4.5, the measurements seem to place the error ball just on the edge of the sphere and slightly outside it. As all legal states have a radius of less than or equal to one in Poincaré space, it is necessary to find a way to return the most likely *legitimate* state reconstructed from a set of measurements.

### The Maximum Likelihood Technique

The problem of reconstructing illegal density matrices is resolved by selecting the legitimate state most likely to have returned the measured counts [28,29]. In practice, analytically calculating this maximally likely state is prohibitively difficult, and a numerical search is necessary. Three elements are required: a manifestly legal parametrization of a density matrix, a likelihood function which can be maximized, and a technique for numerically finding this maximum over a search of the density matrix's parameters.

The Stokes parameters are an unacceptable parametrization for this search, as there are clearly legitimate combinations of these parameters which result in an illegal state. In this context, a legitimate state refers to a non-negative definite Hermitian density matrix of trace one. The property of non-negative definiteness for any matrix  $\hat{\mathcal{G}}$  is written mathematically as

$$\langle \psi | \hat{\mathcal{G}} | \psi \rangle \geq 0 \quad \forall |\psi\rangle. \quad (4.63)$$

Any matrix that can be written in the form  $\hat{\mathcal{G}} = \hat{T}^\dagger \hat{T}$  must be non-negative definite. To see that this is the case, substitute into eq.(4.63):

$$\langle \psi | \hat{T}^\dagger \hat{T} | \psi \rangle = \langle \psi' | \psi' \rangle \geq 0, \quad (4.64)$$

where we have defined  $|\psi'\rangle = \hat{T}|\psi\rangle$ . Furthermore  $(\hat{T}^\dagger \hat{T})^\dagger = \hat{T}^\dagger (\hat{T}^\dagger)^\dagger = \hat{T}^\dagger \hat{T}$ , i.e.,  $\hat{\mathcal{G}} = \hat{T}^\dagger \hat{T}$  must be Hermitian. To ensure normalization, one can simply divide by the trace. Thus the matrix  $\hat{g}$  given by the formula

$$\hat{g} = \hat{T}^\dagger \hat{T} / \text{Tr}\{\hat{T}^\dagger \hat{T}\} \quad (4.65)$$

has all three of the mathematical properties required for density matrices.

For the one-qubit system, we have a  $2 \times 2$  density matrix with 3 independent real parameters (although we will search over 4 in order to fit the intensity of the data). Since it will be useful to be able to invert relation (4.65), it is convenient to choose a tri-diagonal form for  $\hat{T}$ :

$$\hat{T}(t) = \begin{pmatrix} t_1 & 0 \\ t_3 + it_4 & t_2 \end{pmatrix}. \quad (4.66)$$

The manifestly ‘physical’ density matrix  $\hat{\rho}_p$  is then given by the formula

$$\hat{\rho}_p(t) = \hat{T}^\dagger(t) \hat{T}(t) / \text{Tr}\{\hat{T}^\dagger(t) \hat{T}(t)\}, \quad (4.67)$$

where  $t$  is shorthand for all  $t_i$ .

This satisfies the first criterion for a successful maximum likelihood search, by providing an explicitly physical parametrization for  $\hat{\rho}$ . The second criterion, a likelihood function, will in general depend on the specific measurement apparatus used and the physical implementation of the qubit (as these will determine the statistical distributions of counts, and therefore their relative weightings).

If we assume both Gaussian counting statistics and that each of our measurements is taken for the same amount of time, then we can provide a suitable likelihood function. (Basis errors are neglected on the assumption that they are symmetric about a known central basis, and so will not affect the outcome, though they *will* affect the error on that outcome.)

Let  $n_\nu$  be the result for the  $\nu^{\text{th}}$  measurement, out of a total of  $\Xi$  measurements. For the two techniques presented in the previous section applied to measurement of a single-qubit,  $\Xi$  would equal four for one detector and six for two detectors (with two detectors there are *three* analysis settings but *six* measurements). The expected values for these measurements on an unknown state  $\hat{\rho}$  are given by  $\bar{n}_\nu = \mathcal{N}\langle\psi_\nu|\hat{\rho}|\psi_\nu\rangle$ . Here  $\mathcal{N}$  is an unknown normalization parameter defined by  $\frac{n_\nu}{\mathcal{N}} = P_\nu$ , corresponding to the total size per measurement of the ensemble. (It is not always possible to know the size of a measured ensemble, and so the counts rather than the probabilities are used in the likelihood function. In addition, the Gaussian statistical distribution is over counts, not probabilities.) Given these definitions, the probability of obtaining the observed experimental counts  $n_\nu$  from the density matrix  $\hat{\rho}$  is

$$P(n_1, n_2, \dots, n_\Xi) = \frac{1}{\text{Norm}} \prod_\nu \exp \left[ -\frac{(\bar{n}_\nu - n_\nu)^2}{2\hat{\sigma}_\nu^2} \right], \quad (4.68)$$

where  $\hat{\sigma}_\nu$  is the standard deviation of the  $\nu^{\text{th}}$  measurement (given approximately by  $\sqrt{\bar{n}_\nu}$ ) and  $\text{Norm}$  is the normalization constant. For our candidate physical density matrix  $\hat{\rho}_p$  the number of counts expected for the  $\nu$ -th measurement is

$$\bar{n}_\nu(t_1, t_2, \dots, t_{n^2}) = \mathcal{N}\langle\psi_\nu|\hat{\rho}_p(t_1, t_2, \dots, t_{n^2})|\psi_\nu\rangle. \quad (4.69)$$

Thus the likelihood that the matrix  $\hat{\rho}_p(t_1, t_2, \dots, t_{n^2})$  could produce the measured data  $\{n_1, n_2, \dots, n_\Xi\}$  is

$$P(n_1, n_2, \dots, n_\Xi) = \frac{1}{\text{Norm}} \prod_\nu \exp \left[ -\frac{(\mathcal{N}\langle\psi_\nu|\hat{\rho}_p(t_1, t_2, \dots, t_{n^2})|\psi_\nu\rangle - n_\nu)^2}{2\mathcal{N}\langle\psi_\nu|\hat{\rho}_p(t_1, t_2, \dots, t_{n^2})|\psi_\nu\rangle} \right]. \quad (4.70)$$

Note that here we assume that  $\mathcal{N}$  is the same for each measurement. In practice this may not necessarily be the case due to drift (in either source intensity or detector efficiencies) or differing measurement conditions.

Rather than find the maximum value of  $P(t_1, t_2, \dots, t_{n^2})$ , it is somewhat simpler to find the maximum of its logarithm (which is mathematically equivalent). In addition, because  $\mathcal{N}$  is unknown, we absorb it into the  $\hat{T}$  matrix, by setting

$$t'_i = \mathcal{N}t_i. \quad (4.71)$$

Thus the optimization problem reduces to finding the *minimum* of the following function:

$$\mathcal{L}(t'_1, t'_2, \dots, t'_{n^2}) = \sum_{\nu} \frac{(\langle \psi_{\nu} | \hat{\rho}_p(t'_1, t'_2, \dots, t'_{n^2}) | \psi_{\nu} \rangle - n_{\nu})^2}{2\langle \psi_{\nu} | \hat{\rho}_p(t'_1, t'_2, \dots, t'_{n^2}) | \psi_{\nu} \rangle}. \quad (4.72)$$

The final piece in the maximum likelihood technique is an optimization routine, of which there are many available. The authors used the **Matlab** function **fminunc** for the examples given in this chapter.<sup>11</sup> After a minimum is found,  $\hat{\rho}$  can be reconstructed from the values of  $t'$ .

Note that the maximum likelihood technique easily adapts to measurements in non-orthogonal bases (e.g., due to imperfect yet well characterized waveplates) and overcomplete measurements (taking more measurements than is necessary). In the first case the set of  $|\psi\rangle$  is simply non-orthogonal with no effect on the above equations; in the second case the sum in equation 4.72 is extended beyond the minimum  $\Xi$ .

#### *An Example Using Photons*

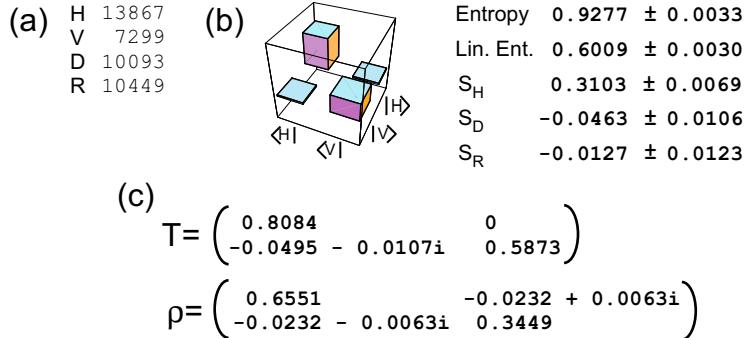
Photon pairs generated via spontaneous parametric downconversion from a nonlinear crystal can be used to generate single photon states. Measuring a photon in one arm collapses the state of its partner to a single qubit Fock state [33]. An ensemble of these photons can be used to implement the maximum likelihood technique. See figure 4.6 for the measured counts and the final density matrix. While these counts would have resulted in an illegal density matrix using the simple linear reconstruction of section 2, the maximum likelihood technique returns the most likely legal state to have given this data.

#### **4.3.2 The Multiple Qubit Maximum Likelihood Technique**

Extending the maximum likelihood technique into the multiple-qubit regime is surprisingly straightforward. Virtually all of the mathematics from the previous section is still applicable, with minor changes. In fact, after rewriting equation 4.67 to be more general:

---

<sup>11</sup> For freely available code and further examples, see:  
<http://www.physics.uiuc.edu/Research/QI/Photonics/Tomography/>



**Fig. 4.6.** Example results for the maximum likelihood technique, taken from single-qubit data. The data was taken using downconverted photons from a nonlinear crystal, using one half of a photon pair as a trigger. (a) The measured counts in four linear independent bases. (b) The likelihood function is minimized for these counts, yielding the listed density matrix, shown graphically. The entropy, linear entropy, and the three normalized Stokes parameters are shown. The errors on these values were calculated using a Monte Carlo simulation of the data in (a). (c) Numerical results. Both the  $\hat{T}$  matrix, calculated using the maximum likelihood technique, and the resulting density matrix  $\hat{\rho}$  are shown,

$$\hat{T}(t) = \begin{pmatrix} t_1 & 0 & \dots & 0 \\ t_{2^n+1} + it_{2^n+2} & t_2 & \dots & 0 \\ \dots & \dots & \dots & 0 \\ t_{4^n-1} + it_{4^n} & t_{4^n-3} + it_{4^n-2} & t_{4^n-5} + it_{4^n-4} & t_{2^n} \end{pmatrix}, \quad (4.73)$$

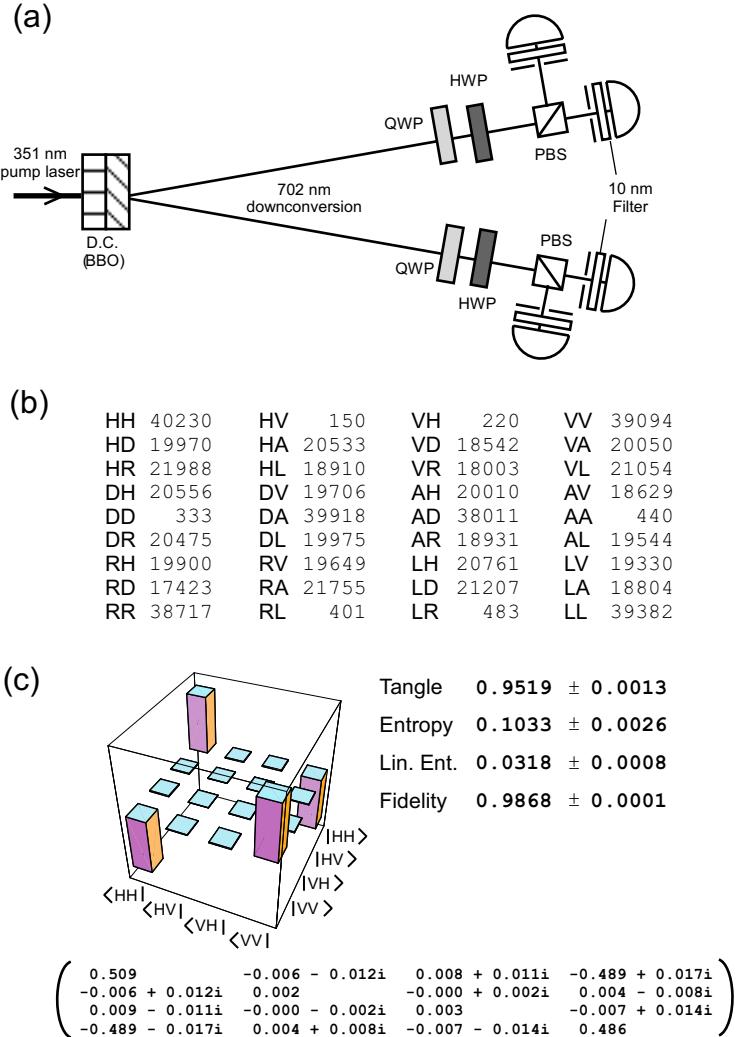
all other equations in the previous section are correct, and the procedure for isolating the reconstructed density matrix is exactly the same.

#### A 2-Qubit Example in Photons

In the 2-qubit case, equation 4.73 becomes

$$\hat{T}(t) = \begin{pmatrix} t_1 & 0 & 0 & 0 \\ t_5 + it_6 & t_2 & 0 & 0 \\ t_{11} + it_{12} & t_7 + it_8 & t_3 & 0 \\ t_{15} + it_{16} & t_{13} + it_{14} & t_9 + it_{10} & t_4 \end{pmatrix}. \quad (4.74)$$

Figure 4.7 shows a set of real data and its analysis. This data represents two-qubit polarization states produced via spontaneous parametric downconversion from two adjacent nonlinear crystals. The resulting uncertainty as to which crystal generated the pair leads to a nearly ideal maximally entangled state [27]. The data shown was measured using the four-detector technique; 9 analysis settings were required, for a total of 36 measurement results.



**Fig. 4.7.** (a) Experimental setup for producing two-qubit polarization states from spontaneous parametric downconversion. (b) Experimental data for a near-Bell state ( $\tilde{\phi}^- = \frac{1}{\sqrt{2}}(|HH\rangle - |VV\rangle)$ ). Shown are the single photon counts from a complete tomography using four detectors. Nine analysis settings yields the 36 measurement results shown. These counts are adjusted for accidentals (dark counts) and renormalized for differences in detector efficiencies (see text for details). (c) The density matrix is shown in both numeric and graphical form, along with several quantities derived from that matrix. The errors shown were calculated using a Monte Carlo simulation of the data from (b).

#### 4.4 Error Analysis

Error analysis of reconstructed density matrices is in practice a non-trivial process. The traditional method of error analysis involves analytically solving for the error in each measurement due to each source of error, then propagating these errors through a calculation of any derived quantity. In the photon case, for example, errors in counting statistics and waveplate settings are analyzed in some detail in reference [28], giving errors in both density matrices and commonly derived quantities, such as the tangle and the linear entropy. In practice, however, these errors appear to be too large: We have experimentally repeated some of our measurements many times, and observed a spread in the value of derived quantities which is approximately an order of magnitude smaller than the spread predicted from an analytic calculation of the uncertainty. Thus it is worthwhile to discuss alternate methods of error analysis.

One promising numerical method is the ‘Monte Carlo’ technique, whereby additional numerically simulated data is used to provide a statistical distribution over any derived quantity. Once an error distribution is understood over a single measurement (e.g., Gaussian for waveplate setting errors or Poissonian over count statistics), a set of ‘simulated’ results can be generated. These results are simulated using the known error distributions in such a way as to produce a full set of numerically generated data which could feasibly have come from the same system. Many of these sets of data are numerically generated (at the measured counts level), and each set is used to calculate a density matrix via the maximum likelihood technique. This set of density matrices is used to calculate the standard error on any quantity implicit in or derived from the density matrix.

As an example, consider the application of the Monte Carlo technique to the downconversion results from figure 4.7. Two polarization encoded qubits are generated within ensembles that obey Poissonian statistics, and these ensembles are used to generate a density matrix using the maximum likelihood technique. In order to find the error on a quantity derived from this density matrix (e.g., the tangle), 36 new measurement results are numerically generated, each drawn randomly from a Poissonian distribution with mean equal to the original number of counts. These 36 numerically generated results are then fed into the maximum likelihood technique, in order to generate a new density matrix, from which, e.g., the tangle may be calculated. This process is repeated many times, generating both many density matrices and a distribution of tangle values, from which the error in the initial tangle may be determined. In practice, additional sets of simulated data must be generated until the error on the quantity of interest converges to a single value. For the data in figures 4.6 and 4.7, a total of 100 simulations were used.

Clearly, the problem of error analysis in state tomography is an area of continuing research – one of many. The development of adaptive tomography techniques could allow both specific measurements and the data collection

times to be tailored in order to optimize for each state to be measured [37]. In addition, because the number of measurements necessary to perform tomography grows exponentially with the number of qubits, it will eventually be necessary to partially characterize states with fewer measurements. Finally, each distinct qubit implementation provides a myriad of unique challenges. Nevertheless, we hope the discussions presented here will be useful for characterizing quantum systems in a broad spectrum of qubit realizations.

The authors would like to thank Evan Jeffrey, Nick Peters, Aaron Vandeven- der, and Tzu-Cheih Wei for helpful conversations and invaluable assistance in the preparation of this manuscript. The development for this work was supported over several years by grants from the National Science Foundation, the National Security Agency, the Advanced Research and Development Activity, the DCI Postdoctoral Program, and most recently, the Army Research Office-sponsored MURI: ‘Center for Photonic Quantum Information Systems’.

## References

1. A. G. White, D. F. V. James, P. H. Eberhard and P. G. Kwiat: Phys. Rev. Lett. **83**, 3103 (1999)
2. K. Sanaka, K Kawahara, and T. Kuga: Physical Review Letters **86**, 5620 (2001)
3. A. Mair, A. Vaziri, G. Weihs, A. Zeilinger: Nature **412**, 3123 (2001)
4. Y. Nambu, K. Usami, Y. Tsuda, K. Matsumoto, and K. Nakamura: Phys. Rev. A **66**, 033816 (2002)
5. G. Giorgi, G. Di Nepi, P. Mataloni, and F De Martini: Laser Phys. **13**, 350 (2003)
6. T. Yamamoto, M. Koashi, S.K. Özdemir, and N. Imoto: Nature **421**, 343 (2003)
7. A.V. Sergienko, G. Di Giuseppe, M. Atatüre, et al: Entangled-Photon State Engineering. In: *Proceedings of the Sixth International Conference on Quantum Communication, Measurement and Computing (QCMC)*, ed by J.H. Shapiro and O. Hirota (Rinton, Princeton, 2003) pp 147–152
8. T. B. Pittman, M. J. Fitch, B. C. Jacobs, and J. D. Franson: arXiv:quant-ph/0303095 (2003)
9. J. L. O’Brien, G. J. Pryde, A. G. White, T. C. Ralph, and D. Branning: Demonstration of an all-optical controlled-NOT gate, to appear in Nature, (2003)
10. I. Marcikic, H. de Riedmatten, W. Tittel, H. Zbinden, N. Gisin: Nature **421**, 509 (2003)
11. D. G. Cory, A. F. Fahmy, and T. F. Havel: Proc Natl. Acad. Sci. U.S.A. **94**, 1634 (1997)
12. J. A. Jones, M. Mosca, and R. H. Hansen: Nature **392**, 344 (1997)
13. Y. S. Weinstein, M. A. Pravia, E. M. Fortunato, S. Lloyd, and D. G. Cory: Physical Review Letters **86**, 1889 (2001)
14. R. Laflamme et al.: arXiv:quant-ph/0207172
15. C. Monroe: Nature **416**, 238 (2002)
16. R. Blatt et al.: Nature **422**, 408 (2003)
17. J. Lehner, U. Leonhardt, and H. Paul: Phys. Rev. A **53**, 2727 (1996)

18. P. G. Kwiat and B. Englert: *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity* (Cambridge Univ. Press, 2004), Chapter 15
19. M. A. Nielsen and I. L. Chuang: *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000)
20. S. Gasiorowitz: *Quantum Physics* (John Wiley and Sons Inc., NY, 1996)
21. G. C. Stokes: *Trans. Cambr. Phil. Soc.* **9**, 399 (1852)
22. N. Peters, J. Altepeter, E. Jeffrey, D. Branning, and P. G. Kwiat: *Quantum Information and Computation* **3**, 503 (2003)
23. J. S. Bell: *Physics* **1**, 195 (1965)
24. R. Werner: *Phys. Rev. A* **40**, 4277 (1989)
25. W. J. Munro, D. F. V. James, A. G. White, P. G. Kwiat: *Phys. Rev. A* **64**, 030302 (2001)
26. M. Born and E. Wolf: *Principles of Optics* (Pergamon Press, New York, 1987)
27. P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P. H. Eberhard: *Phys. Rev. A* **60**, R773 (1999)
28. D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White: *Phys. Rev. A* **64**, 052312 (2001)
29. Z. Hradil and J. Rehacek: *Fortschr. Phys.* **49**, 1083 (2001)
30. A. F. Abouraddy, A. V. Sergienko, B. E. A. Saleh, and M. C. Teich: *Optics Communications* **201**, 93 (2002)
31. K. Zyczkowski: arXiv:quant-ph/0006068, arXiv:quant-ph/0108064
32. D. Kielpinski, V. Meyer, M. A. Rowe, C. A. Sackett, W. M. Itano, C. Monroe, D. J. Wineland: *Science* **291**, 1013 (2001)
33. C. K. Hong and L. Mandel: *Phys. Rev. Lett.* **56**, 58 (1986)
34. W. K. Wootters: *Phys. Rev. Lett.* **80**, 2245 (1998); V. Coffman, J. Kundu, W. K. Wootters: *Phys. Rev. A* **61**, 052306 (2000)
35. B. M. Terhal: arXiv:quant-ph/0101032
36. J. M. G. Sancho and S. F. Huelga: *Phys. Rev. A* **61**, 042303 (2000)
37. G. M. D'Ariano, M. G. A. Paris, M. F. Sacchi: arXiv:quant-ph/0302028



# 5 Unknown Quantum States and Operations, a Bayesian View

Christopher A. Fuchs<sup>1</sup> and Rüdiger Schack<sup>2</sup>

<sup>1</sup> Quantum Information and Optics Research, Bell Labs, Lucent Technologies,  
600–700 Mountain Avenue, Murray Hill, New Jersey 07974, USA

<sup>2</sup> Department of Mathematics, Royal Holloway, University of London, Egham,  
Surrey TW20 0EX, UK

**Summary.** The classical de Finetti theorem provides an operational definition of the concept of an unknown probability in Bayesian probability theory, where probabilities are taken to be degrees of belief instead of objective states of nature. In this chapter, we motivate and review two results that generalize de Finetti's theorem to the quantum mechanical setting: Namely a de Finetti theorem for quantum states and a de Finetti theorem for quantum operations. The quantum-state theorem, in a closely analogous fashion to the original de Finetti theorem, deals with exchangeable density-operator assignments and provides an operational definition of the concept of an “unknown quantum state” in quantum-state tomography. Similarly, the quantum-operation theorem gives an operational definition of an “unknown quantum operation” in quantum-process tomography. These results are especially important for a Bayesian interpretation of quantum mechanics, where quantum states and (at least some) quantum operations are taken to be states of belief rather than states of nature.

## 5.1 Introduction

What is a quantum state?<sup>3</sup> Since the earliest days of quantum theory, it has been understood that the quantum state can be used (through the Born rule) to derive probability distributions for the outcomes of all measurements that can be performed on a quantum system. But is it more than that? Is a quantum state an actual property of the system it describes? The Bayesian view of quantum states [1–6, 13–22] is that it is not: The quantum state is not something the system itself possesses. Rather it is solely a function of the observer

---

<sup>3</sup> This chapter represents predominantly a culling of the material in Refs. [1–3]. Everything, however, has been updated to accommodate the major shift in our thinking represented in Refs. [4–6]. In particular, it reflects a change in our views of quantum probabilities from that of an objective Bayesianism of the type promoted by E. T. Jaynes [7] to a subjective or personalistic Bayesianism of the type promoted by B. de Finetti, L. J. Savage, J. M. Bernardo and A. F. M. Smith, and R. Jeffrey [8–12]. This makes all the difference in the world with regard to the meaning of quantum states, operations, and their usage within statistical theory.

(or, better, agent) who contemplates the predictions, gambles, decisions, or actions he might make with regard to those quantum measurements.

What distinguishes this view from a more traditional “Copenhagen-interpretation style” view—for instance the view expressed so clearly and carefully in Ref. [23]—is that there is no pretense that a quantum state represents a physical fact. Quantum states come logically before that: They represent the temporary and provisional *beliefs* a physicist holds as he travels down the road of inquiry. It is the outcomes of quantum measurements that represent physical facts within quantum theory, not the quantum states. In particular, there is no fact of nature to prohibit two different agents from using distinct pure states  $|\psi\rangle$  and  $|\phi\rangle$  for a single quantum system.<sup>4</sup> Difficult though this may be to accept for someone trained in the traditional presentation of quantum mechanics, the only thing it demonstrates is a careful distinction between the terms *belief* and *fact*.

Quantum states are not facts.<sup>5</sup> But if so, then what is an “unknown quantum state”? There is hardly a paper in the field of quantum information that does not make use of the phrase. Unknown quantum states are teleported [28, 29], protected with quantum error correcting codes [30, 31], and used to check for quantum eavesdropping [32, 33]. The list of uses grows each day. Are all these papers nonsense? In a Bayesian view of quantum states, the phrase is an oxymoron, a contradiction in terms: If quantum states are states of belief rather than states of nature, then a state is *known* by someone—at the very least, by the agent who holds it.

Thus for a quantum Bayesian, if a phenomenon ostensibly invokes the concept of an unknown state in its formulation, the unknown state must be a kind of shorthand for a more involved story. In other words, the usage should be viewed a call to arms, an opportunity for further analysis. For any phenomenon using the idea of an unknown quantum state, the quantum Bayesian should demand that either:

1. The owner of the unknown state—some further agent—be explicitly identified. (In this case, the unknown state is merely a stand-in for the unknown state of belief of an essential player who went unrecognized in the original formulation.) Or,
2. If there is clearly no further agent upon the scene, then a way must be found to reexpress the phenomenon with the term “unknown state” banished from the formulation. (In this case, the end-product will be a

---

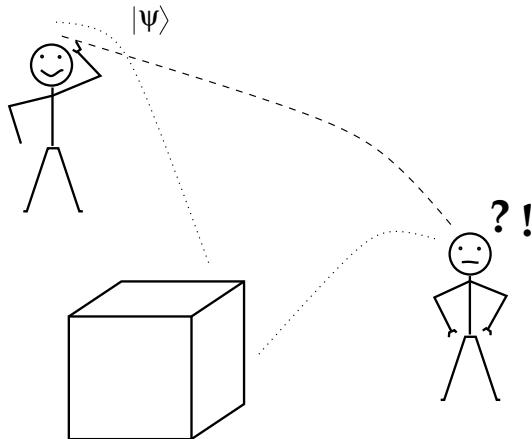
<sup>4</sup> Contrast this to the treatment of Refs. [24–26]. In any case, the present point does not imply that a single agent can believe willy-nilly anything he wishes. To quote D. M. Appleby, “You know, it is *really* hard to believe something you don’t actually believe.”

<sup>5</sup> For a selection of papers that we believe help shore up this statement in various ways—though most of them are not explicitly Bayesian in their view of probability—see Ref. [27].

single quantum state used for describing the phenomenon—namely, the state that actually captures the initial agent's overall beliefs throughout.)

In this chapter, we will analyze the particular use of unknown states that comes from *quantum-state tomography* [34–36]. Beyond that, we will also argue for the necessity of (and carry out) a similar analysis for *quantum-process tomography* [37–39].

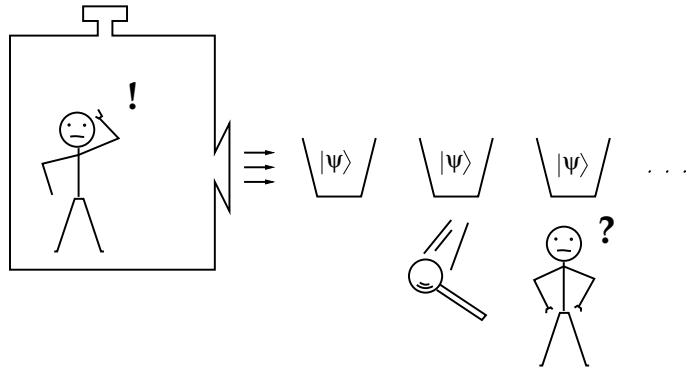
The usual, non-Bayesian description of quantum-state tomography is this. A device of some sort repeatedly prepares many instances of a quantum system in a fixed quantum state  $\rho$ , pure or mixed. An experimentalist who wishes to characterize the operation of the device or to calibrate it for future use may be able to perform measurements on the systems it prepares even if he cannot get at the device itself. This can be useful if the experimenter has some prior knowledge of the device's operation that can be translated into a probability distribution over states. Then learning about the state will also be learning about the device. Most importantly, though, this description of tomography assumes the state  $\rho$  is unknown. The goal of the experimenter is to perform enough measurements, and enough kinds of measurements (on a large enough sample), to estimate the identity of  $\rho$ .



**Fig. 5.1.** What can the term “unknown state” mean if quantum states are taken to be solely compendia of Bayesian expectations rather than states of nature? When we say that a system has an unknown state, must we always imagine a further agent whose state of belief is symbolized by some  $|\psi\rangle$ , and it is the identity of that belief which we are ignorant of?

This is clearly an example where there is no further player on whom to pin the unknown state as a state of belief or judgment. Any attempt to find

such a missing player would be entirely artificial: Where would the player be placed? On the inside of the device the tomographer is trying to characterize? The only available course is the second strategy above—to banish the idea of the unknown state from the formulation of tomography.



**Fig. 5.2.** To make sense of quantum tomography, must we resort to imagining a “man in the box” who has a better description of the systems than we do? How contrived would the Bayesian story be if this were so!

To do this, we take a cue from the field of Bayesian probability theory itself [8–11, 40, 41]. In Bayesian theory, probabilities are not objective states of nature, but measures of personalistic belief. The overarching Bayesian theme is to identify the conditions under which a set of decision-making agents can come to a common belief or probability assignment for a random variable even though their initial beliefs differ. Bernardo and Smith [11] make the point forcefully:

Individual degrees of belief, expressed as probabilities, are inescapably the starting point for descriptions of uncertainty. There can be no theories without theoreticians; no learning without learners; in general, no science without scientists. It follows that learning processes, whatever their particular concerns and fashions at any given point in time, are necessarily reasoning processes which take place in the minds of individuals. To be sure, the object of attention and interest may well be an assumed external, objective reality: but the actuality of the learning process consists in the evolution of individual, subjective beliefs about that reality. However, it is important to emphasize ... that the primitive and fundamental notions of *individual* preference and belief will typically provide the starting point for *interpersonal* communication and reporting processes. ... [W]e shall

therefore often be concerned to identify and examine features of the individual learning process which relate to interpersonal issues, such as the conditions under which an approximate consensus of beliefs might occur in a population of individuals.

Following that theme is the key to understanding quantum tomography from a Bayesian point of view.

The offending classical concept is an “unknown probability,” an oxymoron for the precisely same reason as an unknown quantum state.<sup>6</sup> The procedure analogous to quantum-state tomography is the estimation of an unknown probability from the results of repeated trials on “identically prepared systems,” all of which are said to be described by the same, but unknown probability.

Let us first consider a trivial example to clinch the idea of subjective probabilities before moving to a full tomographic setting. Suppose a die is thrown 10 times. Let  $k \in \{1, \dots, 6\}$  represent the outcome of a single throw of the die, and suppose the results in the 10 throws were these:

- $k = 1$  appeared 1 times,
- $k = 2$  appeared 4 times,
- $k = 3$  appeared 2 times,
- $k = 4$  appeared 2 times,
- $k = 5$  appeared 1 times,
- $k = 6$  appeared 0 times.

A typical inference problem is to assign a probability,  $p$ , to the outcome  $k = 6$  in the next throw of the die, given these data. But clearly the answer cannot be determined by the data alone. It depends on a prior probability assignment. Possibilities include:

1. The assumption that the die is fair. In this case  $p$  is independent of the data and given by  $p = 1/6$ .
2. A totally uninformative prior. In this case, a generalization of the Laplace rule of succession [7] gives  $p = 1/16$ .
3. The die came from a box that contains only trick dice of two types: The first type never comes up 1, and the other type never comes up 6. In this case,  $p = 0$ .

The point of this trivial example cannot be stressed enough. Data alone is never enough to specify a probability distribution. The only way to “derive” a probability distribution from data is to cheat—to make use of an implicit probability assignment made prior to the collection of the data. This holds as much for a problem like the present one, whose setting is a finite number of trials, as for a problem containing a potentially infinite number of trials.

---

<sup>6</sup> If a quantum state is nothing more than a compendium of probabilities—Bayesian probabilities—then, of course, it is for precisely the same reason.

The way to eliminate unknown probabilities from the discussion of infinite numbers of trials was introduced by Bruno de Finetti in the early 1930s [9,41]. The idea is to make explicit the implicit class of priors generally used in such problems. He did this by focusing on the meaning of the *equivalence* of repeated trials. What could equivalent trials mean—de Finetti asked—but that a probability assignment for multiple trials should be symmetric under the permutation of those trials? With his *classical representation theorem*, de Finetti [9] showed that a multi-trial probability assignment that is permutation-symmetric for an arbitrarily large number of trials—he called such multi-trial probabilities *exchangeable*—is equivalent to a probability for the “unknown probabilities.” Thus the unsatisfactory concept of an unknown probability vanishes from the description in favor of the fundamental idea of assigning an exchangeable probability distribution to multiple trials.

This cue in hand, it is easy to see how to reword the description of quantum-state tomography to meet our goals. What is relevant is simply a judgment on the part of the experimenter—notice the essential subjective character of this “judgment”—that there is no distinction between the systems the device is preparing. In operational terms, this is the judgment that *all the systems are and will be the same as far as observational predictions are concerned*. At first glance this statement might seem to be contentless, but the important point is this: To make this statement, one need never use the notion of an unknown state—a completely operational description is good enough. Putting it into technical terms, the statement is that if the experimenter judges a collection of  $N$  of the device’s outputs to have an overall quantum state  $\rho^{(N)}$ , he will also judge any permutation of those outputs to have the same quantum state  $\rho^{(N)}$ . Moreover, he will do this no matter how large the number  $N$  is. This, complemented only by the consistency condition that for any  $N$  the state  $\rho^{(N)}$  be derivable from  $\rho^{(N+1)}$ , makes for the complete story.

The words “quantum state” appear in this formulation, just as in the original formulation of tomography, but there is no longer any mention of *unknown* quantum states. The state  $\rho^{(N)}$  is known by the experimenter (if no one else), for it represents his state of belief. More importantly, the experimenter is in a position to make an unambiguous statement about the structure of the whole sequence of states  $\rho^{(N)}$ : Each of the states  $\rho^{(N)}$  has a kind of permutation invariance over its factors. The content of the *quantum de Finetti representation theorem* [42, 43]—which we will demonstrate in a later section—is that a sequence of states  $\rho^{(N)}$  can have these properties, which are said to make it an *exchangeable* sequence, if and only if each term in it can also be written in the form

$$\rho^{(N)} = \int P(\rho) \rho^{\otimes N} d\rho , \quad (5.1)$$

where  $\rho^{\otimes N} = \rho \otimes \rho \otimes \cdots \otimes \rho$  is an  $N$ -fold tensor product and  $P(\rho)$  is a fixed probability distribution over density operators.<sup>7</sup>

The interpretive import of this theorem is paramount. It alone gives a mandate to the term unknown state in the usual description of tomography. It says that the experimenter can act *as if* his state of belief  $\rho^{(N)}$  comes about because he knows there is a “man in the box,” hidden from view, repeatedly preparing the same state  $\rho$ . He does not know which such state, and the best he can say about the unknown state is captured in the probability distribution  $P(\rho)$ .

The quantum de Finetti theorem furthermore makes a connection to the overarching theme of Bayesianism stressed above. It guarantees for two independent observers—as long as they have a rather minimal agreement in their initial beliefs—that the outcomes of a sufficiently informative set of measurements will force a convergence in their state assignments for the remaining systems [2]. This “minimal” agreement is characterized by a judgment on the part of both parties that the sequence of systems is exchangeable, as described above, and a promise that the observers are not absolutely inflexible in their opinions. Quantitatively, the latter means that though  $P(\rho)$  might be arbitrarily close to zero, it can never vanish.

This coming to agreement works because an exchangeable density operator sequence can be updated to reflect information gathered from measurements by a quantum version of Bayes’s rule for updating probabilities. Specifically, suppose the starting point of a quantum tomography experiment is a prior state

$$\rho^{(N+M)} = \int P(\rho) \rho^{\otimes(N+M)} d\rho \quad (5.2)$$

for  $N + M$  copies of the system. Then the first  $N$  systems are measured. Say the measurement outcomes are represented by a vector  $\alpha = (\alpha_1, \dots, \alpha_N)$ . It can be shown that the post-measurement state of the remaining  $M$  copies conditioned on the outcome  $\alpha$  is of the form

$$\rho^{(M)} = \int P(\rho|\alpha) \rho^{\otimes M} d\rho, \quad (5.3)$$

where  $P(\rho|\alpha)$  is given by a quantum Bayes rule [2]. In the special case that the same measurement,  $\{E_\alpha\}$ , is measured on all  $N$  copies, the quantum Bayes rule takes the simple form

$$P(\rho|\alpha) = \frac{P(\rho)P(\alpha|\rho)}{P(\alpha)}, \quad (5.4)$$

where

$$P(\alpha|\rho) = \text{tr}(\rho^{\otimes N} E_{\alpha_1} \otimes \cdots \otimes E_{\alpha_N}) \quad (5.5)$$

---

<sup>7</sup> For further technical elaborations on the quantum de Finetti theorem, see Ref. [44].

and

$$P(\alpha) = \int P(\rho)P(\alpha|\rho) d\rho. \quad (5.6)$$

For a sufficiently informative set of measurements, as  $N$  becomes large, the updated probability  $P(\rho|\alpha)$  becomes highly peaked on a particular state  $\rho_\alpha$  dictated by the measurement results, regardless of the prior probability  $P(\rho)$ , as long as  $P(\rho)$  is nonzero in a neighborhood of  $\rho_\alpha$ . Suppose the two observers have different initial beliefs, encapsulated in different priors  $P_i(\rho)$ ,  $i = 1, 2$ . The measurement results force them with high probability to a common state of belief in which any number  $M$  of additional systems are assigned the product state  $\rho_\alpha^{\otimes M}$ , i.e.,

$$\int P_i(\rho|\alpha) \rho^{\otimes M} d\rho \longrightarrow \rho_\alpha^{\otimes M} \quad (5.7)$$

for  $N$  sufficiently large.

This shifts the perspective on the purpose of quantum-state tomography: It is not about uncovering some “unknown state of nature,” but rather about the various observers’ coming to agreement over future probabilistic predictions.<sup>8</sup> In this connection, it is interesting to note that the quantum de Finetti theorem and the conclusions just drawn from it work only within the framework of complex vector-space quantum mechanics. For quantum mechanics based on real and quaternionic Hilbert spaces, the connection between exchangeable density operators and unknown quantum states does not hold [1].

The plan of the remainder of the chapter is as follows. In Sec. 5.2, we discuss the classical de Finetti representation theorem [9, 46] in the context of Bayesian probability theory. In Sec. 5.3, we introduce the Bayesian formulation of tomography in terms of exchangeable multi-system density operators, accompanied by a critical discussion of objectivist formulations of tomography. Furthermore, we state the quantum-state de Finetti representation theorem. Section 5.4 presents an elementary proof of the quantum de Finetti theorem. There, also, we introduce a novel measurement technique for tomography based upon generalized quantum measurements. In Sec. 5.5, we come to an intermezzo, mentioning possible extensions of the main theorem. In Sec. 5.6, we change course to consider the issue of quantum operations. In particular, we argue that (at least some) quantum operations should be considered subjective states of belief, just as quantum states themselves. This brings to the fore the issue of “unknown quantum operations” within a Bayesian formulation of quantum mechanics. In Sec. 5.7, we pose the need for a version of a quantum de Finetti theorem for quantum operations in order to make sense of quantum-process tomography from a Bayesian point of view. In Sec. 5.8, we make the statement of the theorem precise. And in

---

<sup>8</sup> For an emphasis of this point in the setting of quantum cryptography, see Ref. [45].

Sec. 5.9, we run through the proof of this quantum-process de Finetti theorem. Finally in Sec. 5.10, we conclude with a discussion of where this research program is going. In particular, we defend ourselves against the (glib) “shot gun” reaction that all of this amounts to a rejection of realism altogether: It simply does not.

## 5.2 The Classical de Finetti Theorem

The tension between the objectivist and Bayesian points of view is not new with quantum mechanics. It arises already in classical probability theory in the form of the war between “objective” and “subjective” interpretations [47]. According to the subjective or Bayesian interpretation, probabilities are measures of personal belief, reflecting how an agent would behave or bet in a certain situation. On the other hand, the objective interpretations—in all their varied forms, from frequency interpretations to propensity interpretations—attempt to view probabilities as real states of affairs or “states of nature” that have nothing to do with an agent at all. Following our discussion in Sec. 5.1, it will come as no surprise to the reader that the authors wholeheartedly adopt the Bayesian approach. For us, the reason is simply our experience with this question, part of which is an appreciation that objective interpretations inevitably run into insurmountable difficulties. (See Refs. [8, 10, 11, 40] for a sampling of criticisms of the objectivist approach.)

We will note briefly, however, that the game of roulette provides an illuminating example. In the European version of the game, the possible outcomes are the numbers  $0, 1, \dots, 36$ . For a player without any privileged information, all 37 outcomes have the same probability  $p = 1/37$ . But suppose that shortly after the ball is launched by the croupier, another player obtains information about the ball’s position and velocity relative to the wheel. Using the information obtained, this other player can make more accurate predictions than the first.<sup>9</sup> His probability is peaked around some group of numbers. The probabilities are thus different for two players with different states of belief.

Whose probability is the true probability? From the Bayesian viewpoint, this question is meaningless: There is no such thing as a true probability. All probability assignments are subjective assignments based specifically upon one’s prior data and beliefs.

For sufficiently precise data—including precise initial data on positions and velocities and probably also including other details such as surface properties of the wheel—Newtonian mechanics assures us that the outcome can be predicted with certainty. This is an important point: The determinism of classical physics provides a strong reason for adopting the subjectivist view

---

<sup>9</sup> An entertaining account of a serious attempt to make money from this idea can be found in Ref. [48].

of probabilities [49]. If the conditions of a trial are exactly specified, the outcomes are predictable with certainty, and all probabilities are 0 or 1. In a deterministic theory, all probabilities strictly greater than 0 and less than 1 arise as a consequence of incomplete information and depend upon their assigner's state of belief.

Of course, we should keep in mind that our ultimate goal is to consider the status of quantum states and, by way of them, quantum probabilities. One can ask, “Does this not change the flavor of these considerations?” Quantum mechanics is avowedly *not* a theory of one’s ignorance of a set of hidden variables [50]:<sup>10</sup> So how can its probabilities be subjective? In Sec. 5.3 we argue that despite the intrinsic indeterminism of quantum mechanics, the essence of the above discussion carries over to the quantum setting intact. Furthermore, there are specifically quantum-motivated arguments for a Bayesian interpretation of quantum probabilities.

For the present, though, let us consider in some detail the general problem of a repeated experiment—spinning a roulette wheel  $N$  times is an example. As discussed briefly in Sec. 5.1, this allows us to make a conceptual connection to quantum-state tomography. Here the individual trials are described by discrete random variables  $x_n \in \{1, 2, \dots, k\}$ ,  $n = 1, \dots, N$ ; that is to say, there are  $N$  random variables, each of which can assume  $k$  discrete values. In an objectivist theory, such an experiment has a standard formulation in which the probability in the multi-trial hypothesis space is given by an independent, identically distributed distribution

$$p(x_1, x_2, \dots, x_N) = p_{x_1} p_{x_2} \cdots p_{x_N} = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}, \quad (5.8)$$

where  $n_j$  is the number of times outcome  $j$  is listed in the vector  $(x_1, x_2, \dots, x_N)$ , so that  $\sum_j n_j = N$ . The number  $p_j$  ( $j = 1, \dots, k$ ) describes the objective, “true” probability that the result of a single experiment will be  $j$  ( $j = 1, \dots, k$ ). This simple description—for the objectivist—only describes the situation from a kind of “God’s eye” point of view. To the experimentalist, the “true” probabilities  $p_1, \dots, p_k$  will very often be *unknown* at the outset. Thus, his burden is to estimate the unknown probabilities by a statistical analysis of the experiment’s outcomes.

In the Bayesian approach, it does not make sense to talk about estimating a true probability. Instead, a Bayesian assigns a prior probability distribution  $p(x_1, x_2, \dots, x_N)$  on the multi-trial hypothesis space and then uses Bayes’s theorem to update the distribution in the light of measurement results. A common criticism from the objectivist camp is that the choice of distribution  $p(x_1, x_2, \dots, x_N)$  with which to start the process seems overly arbitrary to them. On what can it be grounded, they would ask? From the Bayesian viewpoint, the subjectivity of the prior is a strength rather than a weakness, because assigning a prior amounts to laying bare the necessarily subjective

---

<sup>10</sup> Perhaps, more carefully, we should have added, “without a stretch of the imagination.” For a stretch of the imagination, see Ref. [51].

assumptions behind *any* probabilistic argument, be it Bayesian or objectivist. Choosing a prior among all possible distributions on the multi-trial hypothesis space is, however, a daunting task. As we will now see, this task becomes tractable by the de Finetti representation theorem.

It is very often the case that one or more features of a problem stand out so clearly that there is no question about how to incorporate them into an initial assignment. In the present case, the key feature is contained in the assumption that an arbitrary number of repeated trials are equivalent. This means that one has no reason to believe there will be a difference between one trial and the next. In this case, the prior distribution is judged to have the sort of permutation symmetry discussed briefly in Sec. 5.1, which de Finetti [41] called *exchangeability*. The rigorous definition of exchangeability proceeds in two stages.

A probability distribution  $p(x_1, x_2, \dots, x_N)$  is said to be *symmetric* (or finitely exchangeable) if it is invariant under permutations of its arguments, i.e., if

$$p(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(N)}) = p(x_1, x_2, \dots, x_N) \quad (5.9)$$

for any permutation  $\pi$  of the set  $\{1, \dots, N\}$ . The distribution  $p(x_1, x_2, \dots, x_N)$  is called *exchangeable* (or infinitely exchangeable) if it is symmetric and if for any integer  $M > 0$ , there is a symmetric distribution  $p_{N+M}(x_1, x_2, \dots, x_{N+M})$  such that

$$p(x_1, x_2, \dots, x_N) = \sum_{x_{N+1}, \dots, x_{N+M}} p_{N+M}(x_1, \dots, x_N, x_{N+1}, \dots, x_{N+M}). \quad (5.10)$$

This last statement means that the distribution  $p$  can be extended to a symmetric distribution of arbitrarily many random variables. Expressed informally, an exchangeable distribution can be thought of as arising from an infinite sequence of random variables whose order is irrelevant.

We now come to the main statement of this section: If a probability distribution  $p(x_1, x_2, \dots, x_N)$  is exchangeable, then it can be written uniquely in the form

$$p(x_1, x_2, \dots, x_N) = \int_{\mathcal{S}_k} P(\mathbf{p}) p_{x_1} p_{x_2} \cdots p_{x_N} d\mathbf{p} = \int_{\mathcal{S}_k} P(\mathbf{p}) p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k} d\mathbf{p}, \quad (5.11)$$

where  $\mathbf{p} = (p_1, p_2, \dots, p_k)$ , and the integral is taken over the probability simplex

$$\mathcal{S}_k = \left\{ \mathbf{p} : p_j \geq 0 \text{ for all } j \text{ and } \sum_{j=1}^k p_j = 1 \right\}. \quad (5.12)$$

Furthermore, the function  $P(\mathbf{p}) \geq 0$  is required to be a probability density function on the simplex:

$$\int_{\mathcal{S}_k} P(\mathbf{p}) d\mathbf{p} = 1. \quad (5.13)$$

Equation (5.11) comprises the classical de Finetti representation theorem for discrete random variables. (A simple proof of this theorem in the case of binary random variables can be found in Refs. [1, 46].)

Let us reiterate the importance of this result for the present considerations. It says that an agent, making solely the judgment of exchangeability for a sequence of random variables  $x_j$ , can proceed *as if* his state of belief had instead come about through ignorance of an *unknown*, but objectively existent set of probabilities  $\mathbf{p}$ . His precise ignorance of  $\mathbf{p}$  is captured by the “probability on probabilities”  $P(\mathbf{p})$ . This is in direct analogy to what we desire of a solution to the problem of the unknown quantum state in quantum-state tomography.

As a final note before finally addressing the quantum problem in Sec. 5.3, we point out that both conditions in the definition of exchangeability are crucial for the proof of the de Finetti theorem. In particular, there are probability distributions  $p(x_1, x_2, \dots, x_N)$  that are symmetric, but not exchangeable. A simple example is the distribution  $p(x_1, x_2)$  of two binary random variables  $x_1, x_2 \in \{0, 1\}$ ,

$$p(0, 0) = p(1, 1) = 0 , \quad (5.14)$$

$$p(0, 1) = p(1, 0) = \frac{1}{2} . \quad (5.15)$$

One can easily check that  $p(x_1, x_2)$  cannot be written as the marginal of a symmetric distribution of three variables, as in Eq. (5.10). Therefore it can have no representation along the lines of Eq. (5.11). (For an extended discussion of this, see Ref. [52].) Indeed, Eqs. (5.14) and (5.15) characterize a perfect “anticorrelation” of the two variables, in contrast to the positive correlation implied by distributions of de Finetti form.

### 5.3 The quantum de Finetti representation

Let us now return to the problem of quantum-state tomography described in Sec. 5.1. In the objectivist formulation of the problem, a device repeatedly prepares copies of a system in the same quantum state  $\rho$ . This is generally a mixed-state density operator on a Hilbert space  $\mathcal{H}_d$  of  $d$  dimensions. We call the totality of such density operators  $\mathcal{D}_d$ . The joint quantum state of the  $N$  systems prepared by the device is then given by

$$\rho^{\otimes N} = \rho \otimes \rho \otimes \cdots \otimes \rho , \quad (5.16)$$

the  $N$ -fold tensor product of  $\rho$  with itself. This, of course, is a very restricted example of a density operator on the tensor-product Hilbert space  $\mathcal{H}_d^{\otimes N} \equiv \mathcal{H}_d \otimes \cdots \otimes \mathcal{H}_d$ . The experimenter, who performs quantum-state tomography, tries to determine  $\rho$  as precisely as possible. Depending upon the version of

the argument,  $\rho$  is interpreted as the “true” state of each of the systems or as a description of the “true” preparation procedure.

We have already articulated our dissatisfaction with this way of stating the problem, but we give here a further sense of why both interpretations above are untenable. Let us deal first with the version where  $\rho$  is regarded as the true, objective state of each of the systems. In this discussion it is useful to consider separately the cases of mixed and pure states  $\rho$ . The arguments against regarding mixed states as objective properties of a quantum system are essentially the same as those against regarding probabilities as objective. In analogy to the roulette example given in the previous section, we can say that, whenever an observer assigns a mixed state to a physical system, one can think of another observer who assigns a different state based on privileged information.

The quantum argument becomes yet more compelling if the apparently nonlocal nature of quantum states is taken into consideration. Consider two parties,  $A$  and  $B$ , who are far apart in space, say several light years apart. Each party possesses a spin-1/2 particle. Initially the joint state of the two particles is the maximally entangled pure state  $\frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$ . Consequently,  $A$  assigns the totally mixed state  $\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$  to her own particle. Now  $B$  makes a measurement on his particle, finds the result 0, and assigns to  $A$ 's particle the pure state  $|0\rangle$ . Is this now the “true,” objective state of  $A$ 's particle? At what precise time does the objective state of  $A$ 's particle change from totally mixed to pure? If the answer is “simultaneously with  $B$ 's measurement,” then what frame of reference should be used to determine simultaneity? These questions and potential paradoxes are avoided if states are interpreted as states of belief. In our example,  $A$  and  $B$  have different states of belief and therefore assign different states. For a detailed analysis of this example, see Ref. [53]; for an experimental investigation see Ref. [54].

If one admits that mixed states cannot be objective properties, because another observer, possessing privileged information, can know which pure state underlies the mixed state, then it becomes very tempting to regard the pure states as giving the “true” state of a system. Probabilities that come from pure states would then be regarded as objective, and the probabilities for pure states within an ensemble decomposition of mixed states would be regarded as subjective, expressing our ignorance of which pure state is the “true” state of the system. An immediate and, in our view, irremediable problem with this idea is that a mixed state has infinitely many ensemble decompositions into pure states [55–57], so the distinction between subjective and objective becomes hopelessly blurred.

This problem can be made concrete by the example of a spin- $\frac{1}{2}$  particle. Any pure state of the particle can be written in terms of the Pauli matrices as

$$|\mathbf{n}\rangle\langle\mathbf{n}| = \frac{1}{2}(I + \mathbf{n} \cdot \boldsymbol{\sigma}) = \frac{1}{2}(I + n_1\sigma_1 + n_2\sigma_2 + n_3\sigma_3), \quad (5.17)$$

where the unit vector  $\mathbf{n} = n_1\mathbf{e}_1 + n_2\mathbf{e}_2 + n_3\mathbf{e}_3$  labels the pure state, and  $I$  denotes the unit operator. An arbitrary state  $\rho$ , mixed or pure, of the particle can be expressed as

$$\rho = \frac{1}{2}(I + \mathbf{S} \cdot \boldsymbol{\sigma}), \quad (5.18)$$

where  $0 \leq |\mathbf{S}| \leq 1$ . If  $|\mathbf{S}| < 1$ , there is an infinite number of ways in which  $\mathbf{S}$  can be written in the form  $\mathbf{S} = \sum_j p_j \mathbf{n}_j$ ,  $|\mathbf{n}_j| = 1$ , with the numbers  $p_j$  comprising a probability distribution, and hence an infinite number of ensemble decompositions of  $\rho$ :

$$\rho = \sum_j p_j \frac{1}{2}(I + \mathbf{n}_j \cdot \boldsymbol{\sigma}) = \sum_j p_j |\mathbf{n}_j\rangle\langle\mathbf{n}_j|. \quad (5.19)$$

Suppose for specificity that the particle's state is a mixed state with  $\mathbf{S} = \frac{1}{2}\mathbf{e}_3$ . Writing  $\mathbf{S} = \frac{3}{4}\mathbf{e}_3 + \frac{1}{4}(-\mathbf{e}_3)$  gives the eigendecomposition,

$$\rho = \frac{3}{4}|\mathbf{e}_3\rangle\langle\mathbf{e}_3| + \frac{1}{4}|-\mathbf{e}_3\rangle\langle-\mathbf{e}_3|, \quad (5.20)$$

where we are to regard the probabilities 3/4 and 1/4 as subjective expressions of ignorance about which eigenstate is the “true” state of the particle. Writing  $\mathbf{S} = \frac{1}{2}\mathbf{n}_+ + \frac{1}{2}\mathbf{n}_-$ , where  $\mathbf{n}_\pm = \frac{1}{2}\mathbf{e}_3 \pm \frac{\sqrt{3}}{2}\mathbf{e}_1$ , gives another ensemble decomposition,

$$\rho = \frac{1}{2}|\mathbf{n}_+\rangle\langle\mathbf{n}_+| + \frac{1}{2}|\mathbf{n}_-\rangle\langle\mathbf{n}_-|, \quad (5.21)$$

where we are now to regard the two probabilities of 1/2 as expressing ignorance of whether the “true” state is  $|\mathbf{n}_+\rangle$  or  $|\mathbf{n}_-\rangle$ .

The problem becomes acute when we ask for the probability that a measurement of the  $z$  component of spin yields spin up; this probability is given by  $\langle\mathbf{e}_3|\rho|\mathbf{e}_3\rangle = \frac{1}{2}(1 + \frac{1}{2}\langle\mathbf{e}_3|\sigma_3|\mathbf{e}_3\rangle) = 3/4$ . The eigendecomposition gets this probability by the route

$$\langle\mathbf{e}_3|\rho|\mathbf{e}_3\rangle = \underbrace{\frac{3}{4}|\langle\mathbf{e}_3|\mathbf{e}_3\rangle|^2}_{1} + \underbrace{\frac{1}{4}|\langle\mathbf{e}_3|-\mathbf{e}_3\rangle|^2}_{0}. \quad (5.22)$$

Here the objective quantum probabilities, calculated from the eigenstates, report that the particle definitely has spin up or definitely has spin down; the overall probability of 3/4 comes from mixing these objective probabilities with the subjective probabilities for the eigenstates. The decomposition (5.21) gets the same overall probability by a different route,

$$\langle\mathbf{e}_3|\rho|\mathbf{e}_3\rangle = \underbrace{\frac{1}{2}|\langle\mathbf{e}_3|\mathbf{n}_+\rangle|^2}_{3/4} + \underbrace{\frac{1}{2}|\langle\mathbf{e}_3|\mathbf{n}_-\rangle|^2}_{3/4}. \quad (5.23)$$

Now the quantum probabilities tell us that the objective probability for the particle to have spin up is 3/4. This simple example illustrates the folly of

trying to have two kinds of probabilities in quantum mechanics. The lesson is that if a density operator is even partially a reflection of one's state of belief, the multiplicity of ensemble decomposition means that a pure state must also be a state of belief.

Return now to the second version of the objectivist formulation of tomography, in which the experimenter is said to be using quantum-state tomography to determine an unknown preparation procedure. Imagine that the tomographic reconstruction results in the mixed state  $\rho$ , rather than a pure state, as in fact all actual laboratory procedures do. Now there is a serious problem, because a mixed state does not correspond to a well-defined procedure, but is itself a probabilistic mixture of well-defined procedures, i.e., pure states. The experimenter is thus trying to determine an unknown procedure that has no unique decomposition into well defined procedures. Thus he cannot be said to be determining an unknown procedure at all. This problem does not arise in a Bayesian interpretation, according to which all quantum states, pure or mixed, are states of belief. In analogy to the classical case, the quantum de Finetti representation provides an operational definition for the idea of an unknown quantum state in this case.

Let us therefore turn to the Bayesian formulation of the quantum-state tomography problem. Before the tomographic measurements, the Bayesian experimenter assigns a prior quantum state to the joint system composed of the  $N$  systems, reflecting his prior state of belief. Just as in the classical case, this is a daunting task unless the assumption of exchangeability is justified.

The definition of the quantum version of exchangeability is closely analogous to the classical definition. Again, the definition proceeds in two stages. First, a joint state  $\rho^{(N)}$  of  $N$  systems is said to be *symmetric* (or finitely exchangeable) if it is invariant under any permutation of the systems. To see what this means formally, first write out  $\rho^{(N)}$  with respect to any orthonormal tensor-product basis on  $\mathcal{H}_d^{\otimes N}$ , say  $|i_1\rangle|i_2\rangle\cdots|i_N\rangle$ , where  $i_k \in \{1, 2, \dots, d\}$  for all  $k$ . The joint state takes the form

$$\rho^{(N)} = \sum_{i_1, \dots, i_N; j_1, \dots, j_N} R_{i_1, \dots, i_N; j_1, \dots, j_N}^{(N)} |i_1\rangle \cdots |i_N\rangle \langle j_1| \cdots \langle j_N|, \quad (5.24)$$

where  $R_{i_1, \dots, i_N; j_1, \dots, j_N}^{(N)}$  is the density matrix in this representation. What we demand is that for any permutation  $\pi$  of the set  $\{1, \dots, N\}$ ,

$$\begin{aligned} \rho^{(N)} &= \sum_{i_1, \dots, i_N; j_1, \dots, j_N} R_{i_1, \dots, i_N; j_1, \dots, j_N}^{(N)} |i_{\pi^{-1}(1)}\rangle \cdots |i_{\pi^{-1}(N)}\rangle \langle j_{\pi^{-1}(1)}| \cdots \langle j_{\pi^{-1}(N)}| \\ &= \sum_{i_1, \dots, i_N; j_1, \dots, j_N} R_{i_{\pi(1)}, \dots, i_{\pi(N)}; j_{\pi(1)}, \dots, j_{\pi(N)}}^{(N)} |i_1\rangle \cdots |i_N\rangle \langle j_1| \cdots \langle j_N|, \end{aligned} \quad (5.25)$$

which is equivalent to

$$R_{i_{\pi(1)}, \dots, i_{\pi(N)}; j_{\pi(1)}, \dots, j_{\pi(N)}}^{(N)} = R_{i_1, \dots, i_N; j_1, \dots, j_N}^{(N)}. \quad (5.26)$$

The state  $\rho^{(N)}$  is said to be *exchangeable* (or infinitely exchangeable) if it is symmetric and if, for any  $M > 0$ , there is a symmetric state  $\rho^{(N+M)}$  of  $N + M$  systems such that the marginal density operator for  $N$  systems is  $\rho^{(N)}$ , i.e.,

$$\rho^{(N)} = \text{tr}_M \rho^{(N+M)}, \quad (5.27)$$

where the trace is taken over the additional  $M$  systems. In explicit basis-dependent notation, this requirement is

$$\begin{aligned} \rho^{(N)} = & \sum_{i_1, \dots, i_N; j_1, \dots, j_N} \left( \sum_{i_{N+1}, \dots, i_{N+M}} R_{i_1, \dots, i_N, i_{N+1}, \dots, i_{N+M}; j_1, \dots, j_N, i_{N+1}, \dots, i_{N+M}}^{(N+M)} \right) \\ & \times |i_1\rangle \cdots |i_N\rangle \langle j_1| \cdots \langle j_N|. \end{aligned} \quad (5.28)$$

In analogy to the classical case, an exchangeable density operator can be thought of informally as the description of a subsystem of an infinite sequence of systems whose order is irrelevant.

The precise statement of the quantum de Finetti representation theorem [42, 58] is that any exchangeable state of  $N$  systems can be written uniquely in the form

$$\rho^{(N)} = \int_{\mathcal{D}_d} P(\rho) \rho^{\otimes N} d\rho. \quad (5.29)$$

Here  $P(\rho) \geq 0$  is normalized by

$$\int_{\mathcal{D}_d} P(\rho) d\rho = 1, \quad (5.30)$$

with  $d\rho$  being a suitable measure on density operator space  $\mathcal{D}_d$  [e.g., one could choose  $d\rho = dS d\Omega$  in the parameterization (5.18) for a spin-1/2 particle]. The upshot of the theorem, as already advertised, is that it makes it possible to think of an exchangeable quantum-state assignment *as if* it were a probabilistic mixture characterized by a probability density  $P(\rho)$  for the product states  $\rho^{\otimes N}$ .

Just as in the classical case, both components of the definition of exchangeability are crucial for arriving at the representation theorem of Eq. (5.29). The reason now, however, is much more interesting than it was previously. In the classical case, extendibility was used solely to exclude anticorrelated probability distributions. Here extendibility is necessary to exclude the possibility of Bell inequality violations for measurements on the separate systems. This is because the assumption of symmetry alone for an  $N$ -party quantum system does not exclude the possibility of quantum entanglement, and all states that can be written as a mixture of product states—of which Eq. (5.29) is an example—have no entanglement [59]. A simple example for a state that is symmetric but not exchangeable is the Greenberger-Horne-Zeilinger (GHZ) state of three spin- $\frac{1}{2}$  particles [60],

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle|0\rangle + |1\rangle|1\rangle|1\rangle) , \quad (5.31)$$

which is not extendible to a symmetric state on four systems. This follows because the only states of four particles that marginalize to a three-particle pure state, like the GHZ state, are product states of the form  $|\text{GHZ}\rangle\langle\text{GHZ}| \otimes \rho$ , where  $\rho$  is the state of the fourth particle; such states clearly cannot be symmetric. These considerations show that in order for the proposed theorem to be valid, it must be the case that as  $M$  increases in Eq. (5.27), the possibilities for entanglement in the separate systems compensatingly decrease [61].

#### 5.4 Proof of the quantum de Finetti theorem

To prove the quantum version of the de Finetti theorem, we rely on the classical theorem as much as possible. We start from an exchangeable density operator  $\rho^{(N)}$  defined on  $N$  copies of a system. We bring the classical theorem to our aid by imagining a sequence of identical quantum measurements on the separate systems and considering the outcome probabilities they would produce. Because  $\rho^{(N)}$  is assumed exchangeable, such identical measurements give rise to an exchangeable probability distribution for the outcomes. The trick is to recover enough information from the exchangeable statistics of these measurements to characterize the exchangeable density operators.

With this in mind, the proof is expedited by making use of the theory of generalized quantum measurements or positive operator-valued measures (POVMs) [62–64]. POVMs generalize the textbook notion of measurement by distilling the essential properties that make the Born rule work. The generalized notion of measurement is this: *any* set  $\mathcal{E} = \{E_\alpha\}$  of positive-semidefinite operators on  $\mathcal{H}_d$  that forms a resolution of the identity, i.e., that satisfies

$$\langle\psi|E_\alpha|\psi\rangle \geq 0 , \quad \text{for all } |\psi\rangle \in \mathcal{H}_d \quad (5.32)$$

and

$$\sum_\alpha E_\alpha = I , \quad (5.33)$$

corresponds to at least one laboratory procedure counting as a measurement. The outcomes of the measurement are identified with the indices  $\alpha$ , and the probabilities of those outcomes are computed according to the generalized Born rule,

$$p_\alpha = \text{tr}(\rho E_\alpha) . \quad (5.34)$$

The set  $\mathcal{E}$  is called a POVM, and the operators  $E_\alpha$  are called POVM elements. Unlike standard or von Neumann measurements, there is no limitation on the number of values  $\alpha$  can take, the operators  $E_\alpha$  need not be rank-1, and there is no requirement that the  $E_\alpha$  be idempotent and mutually orthogonal. This definition has important content because the older notion of measurement

is simply too restrictive: there are laboratory procedures that clearly should be called “measurements,” but that cannot be expressed in terms of the von Neumann measurement process alone.

One might wonder whether the existence of POVMs contradicts everything taught about standard measurements in the traditional graduate textbooks [65] and the well-known classics [66]. It does not. The reason is that any POVM can be represented formally as a standard measurement on an ancillary system that has interacted in the past with the system of main interest. Thus in a certain sense, von Neumann measurements capture everything that can be said about quantum measurements [63]. A way to think about this is that by learning something about the ancillary system through a standard measurement, one in turn learns something about the system of real interest. Indirect though this might seem, it can be a very powerful technique, sometimes revealing information that could not have been revealed otherwise [67].

For instance, by considering POVMs, one can consider measurements with an outcome cardinality that exceeds the dimensionality of the Hilbert space. What this means is that whereas the statistics of a von Neumann measurement can only reveal information about the  $d$  diagonal elements of a density operator  $\rho$ , through the probabilities  $\text{tr}(\rho \Pi_i)$ , the statistics of a POVM generally can reveal things about the off-diagonal elements, too. It is precisely this property that we take advantage of in our proof of the quantum de Finetti theorem.

Our problem hinges on finding a special kind of POVM, one for which any set of outcome probabilities specifies a unique operator. This boils down to a problem of pure linear algebra. The space of operators on  $\mathcal{H}_d$  is itself a linear vector space of dimension  $d^2$ . The quantity  $\text{tr}(A^\dagger B)$  serves as an inner product on that space. If the POVM elements  $E_\alpha$  span the space of operators—there must be at least  $d^2$  POVM elements in the set—the measurement probabilities  $p_\alpha = \text{tr}(\rho E_\alpha)$ —now thought of as *projections* in the directions  $E_\alpha$ —are sufficient to specify a unique operator  $\rho$ . Two distinct density operators  $\rho$  and  $\sigma$  must give rise to different measurement statistics. Such measurements, which might be called *informationally complete*, have been studied for some time [68].

For our proof we need a slightly refined notion—that of a *minimal* informationally complete measurement. If an informationally complete POVM has more than  $d^2$  operators  $E_\alpha$ , these operators form an overcomplete set. This means that given a set of outcome probabilities  $p_\alpha$ , there is generally *no* operator  $A$  that generates them according to  $p_\alpha = \text{tr}(AE_\alpha)$ . Our proof requires the existence of such an operator, so we need a POVM that has precisely  $d^2$  linearly independent POVM elements  $E_\alpha$ . Such a POVM has the minimal number of POVM elements to be informationally complete. Given a set of outcome probabilities  $p_\alpha$ , there is a unique operator  $A$  such that

$p_\alpha = \text{tr}(AE_\alpha)$ , even though, as we discuss below,  $A$  is not guaranteed to be a density operator.

Do minimal informationally complete POVMs exist? The answer is yes. We give here a simple way to produce one, though there are surely more elegant ways with greater symmetry [69, 70]. Start with a complete orthonormal basis  $|e_j\rangle$  on  $\mathcal{H}_d$ , and let  $\Gamma_{jk} = |e_j\rangle\langle e_k|$ . It is easy to check that the following  $d^2$  rank-1 projectors  $\Pi_\alpha$  form a linearly independent set.

1. For  $\alpha = 1, \dots, d$ , let

$$\Pi_\alpha \equiv \Gamma_{jj}, \quad (5.35)$$

where  $j$ , too, runs over the values  $1, \dots, d$ .

2. For  $\alpha = d+1, \dots, \frac{1}{2}d(d+1)$ , let

$$\Pi_\alpha \equiv \Gamma_{jk}^{(1)} = \frac{1}{2}(|e_j\rangle + |e_k\rangle)(\langle e_j| + \langle e_k|) = \frac{1}{2}(\Gamma_{jj} + \Gamma_{kk} + \Gamma_{jk} + \Gamma_{kj}), \quad (5.36)$$

where  $j < k$ .

3. Finally, for  $\alpha = \frac{1}{2}d(d+1) + 1, \dots, d^2$ , let

$$\Pi_\alpha \equiv \Gamma_{jk}^{(2)} = \frac{1}{2}(|e_j\rangle + i|e_k\rangle)(\langle e_j| - i\langle e_k|) = \frac{1}{2}(\Gamma_{jj} + \Gamma_{kk} - i\Gamma_{jk} + i\Gamma_{kj}), \quad (5.37)$$

where again  $j < k$ .

All that remains is to transform these (positive-semidefinite) linearly independent operators  $\Pi_\alpha$  into a proper POVM. This can be done by considering the positive semidefinite operator  $G$  defined by

$$G = \sum_{\alpha=1}^{d^2} \Pi_\alpha. \quad (5.38)$$

It is straightforward to show that  $\langle\psi|G|\psi\rangle > 0$  for all  $|\psi\rangle \neq 0$ , thus establishing that  $G$  is positive definite and hence invertible. Applying the (invertible) linear transformation  $X \rightarrow G^{-1/2}XG^{-1/2}$  to Eq. (5.38), we find a valid decomposition of the identity,

$$I = \sum_{\alpha=1}^{d^2} G^{-1/2} \Pi_\alpha G^{-1/2}. \quad (5.39)$$

The operators

$$E_\alpha = G^{-1/2} \Pi_\alpha G^{-1/2} \quad (5.40)$$

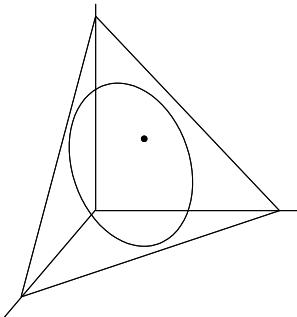
satisfy the conditions of a POVM, Eqs. (5.32) and (5.33), and moreover, they retain the rank and linear independence of the original  $\Pi_\alpha$ .

It is worthwhile noting a special property of all minimal informationally complete POVMs  $\{E_\alpha\}$ : For *no* quantum state  $\rho$  is it ever the case that

$\text{tr } \rho E_\alpha = 1$ . Let us show this for the case where all the  $E_\alpha = k_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|$  are rank one. Suppose it were the case that for some  $\rho$ ,  $\text{tr } \rho E_0 = 1$ . Then it would also have to be the case that  $E_0 = \rho = |\psi\rangle\langle\psi|$  for some vector  $|\psi\rangle$ . But then, because  $\sum E_\alpha = I$ , it follows that  $\langle\psi_\alpha|\psi\rangle = 0$  for all  $\alpha \neq 0$ . That is, all the  $|\psi_\alpha\rangle$  must lie in a  $(d - 1)$ -dimensional subspace. But then it follows that at most  $(d - 1)^2$  of the operators  $E_\alpha$ ,  $\alpha \neq 0$ , can be linearly independent. Since  $(d - 1)^2 + 1 \neq d^2$ , we have a contradiction with the assumption that  $\{E_\alpha\}$  is an informationally complete POVM. (In particular, for the case of the particular, minimal informationally complete POVM in Eq. (5.40), it can be shown that [6]

$$P(h) \leq \left[ d - \frac{1}{2} \left( 1 + \cot \frac{3\pi}{4d} \right) \right]^{-1} < 1. \quad (5.41)$$

For large  $d$ , this bound asymptotes to roughly  $(0.79d)^{-1}$ .) What this means generically is that for no informationally complete POVM can the vectors  $p_\alpha$  completely fill the probability simplex.



**Fig. 5.3.** The planar surface represents the space of all probability distributions over  $d^2$  outcomes. If the probability distributions are imagined to be generated by a minimal informationally complete measurement, the set of valid quantum states in this representation represents a convex region strictly smaller than the whole simplex.

With this generalized measurement (or any other one like it), we can return to the main line of proof. Recall we assumed that we captured our state of belief by an exchangeable density operator  $\rho^{(N)}$ . Consequently, repeated application of the (imagined) measurement  $\mathcal{E}$  must give rise to an exchangeable probability distribution over the  $N$  random variables  $\alpha_n \in \{1, 2, \dots, d^2\}$ ,  $n = 1, \dots, N$ . We now analyze these probabilities.

Quantum mechanically, it is valid to think of the  $N$  repeated measurements of  $\mathcal{E}$  as a single measurement on the Hilbert space  $\mathcal{H}_d^{\otimes N} \equiv \mathcal{H}_d \otimes \dots \otimes \mathcal{H}_d$ . This measurement, which we denote  $\mathcal{E}^{\otimes N}$ , consists of  $d^{2N}$  POVM elements

of the form  $E_{\alpha_1} \otimes \cdots \otimes E_{\alpha_N}$ . The probability of any particular outcome sequence of length  $N$ , namely  $\alpha \equiv (\alpha_1, \dots, \alpha_N)$ , is given by the standard quantum rule,

$$p^{(N)}(\alpha) = \text{tr}(\rho^{(N)} E_{\alpha_1} \otimes \cdots \otimes E_{\alpha_N}) . \quad (5.42)$$

Because the distribution  $p^{(N)}(\alpha)$  is exchangeable, we have by the classical de Finetti theorem [see Eq. (5.11)] that there exists a unique probability density  $P(\mathbf{p})$  on  $\mathcal{S}_{d^2}$  such that

$$p^{(N)}(\alpha) = \int_{\mathcal{S}_{d^2}} P(\mathbf{p}) p_{\alpha_1} p_{\alpha_2} \cdots p_{\alpha_N} d\mathbf{p} . \quad (5.43)$$

It should now begin to be apparent why we chose to imagine a measurement  $\mathcal{E}$  consisting of precisely  $d^2$  linearly independent elements. This allows us to assert the existence of a *unique* operator  $A_{\mathbf{p}}$  on  $\mathcal{H}_d$  corresponding to each point  $\mathbf{p}$  in the domain of the integral. The ultimate goal here is to turn Eqs. (5.42) and (5.43) into a single operator equation.

With that in mind, let us define  $A_{\mathbf{p}}$  as the unique operator satisfying the following  $d^2$  linear equations:

$$\text{tr}(A_{\mathbf{p}} E_{\alpha}) = p_{\alpha} , \quad \alpha = 1, \dots, d^2 . \quad (5.44)$$

Inserting this definition into Eq. (5.43) and manipulating it according to the algebraic rules of tensor products—namely  $(A \otimes B)(C \otimes D) = AC \otimes BD$  and  $\text{tr}(A \otimes B) = (\text{tr}A)(\text{tr}B)$ —we see that

$$\begin{aligned} p^{(N)}(\alpha) &= \int_{\mathcal{S}_{d^2}} P(\mathbf{p}) \text{tr}(A_{\mathbf{p}} E_{\alpha_1}) \cdots \text{tr}(A_{\mathbf{p}} E_{\alpha_N}) d\mathbf{p} \\ &= \int_{\mathcal{S}_{d^2}} P(\mathbf{p}) \text{tr}(A_{\mathbf{p}} E_{\alpha_1} \otimes \cdots \otimes A_{\mathbf{p}} E_{\alpha_N}) d\mathbf{p} \\ &= \int_{\mathcal{S}_{d^2}} P(\mathbf{p}) \text{tr}[A_{\mathbf{p}}^{\otimes N} (E_{\alpha_1} \otimes \cdots \otimes E_{\alpha_N})] d\mathbf{p} . \end{aligned} \quad (5.45)$$

If we further use the linearity of the trace, we can write the same expression as

$$p^{(N)}(\alpha) = \text{tr} \left[ \left( \int_{\mathcal{S}_{d^2}} P(\mathbf{p}) A_{\mathbf{p}}^{\otimes n} d\mathbf{p} \right) E_{\alpha_1} \otimes \cdots \otimes E_{\alpha_N} \right] . \quad (5.46)$$

The identity between Eqs. (5.42) and (5.46) must hold for all sequences  $\alpha$ . It follows that

$$\rho^{(N)} = \int_{\mathcal{S}_{d^2}} P(\mathbf{p}) A_{\mathbf{p}}^{\otimes N} d\mathbf{p} . \quad (5.47)$$

This is because the operators  $E_{\alpha_1} \otimes \cdots \otimes E_{\alpha_N}$  form a complete basis for the vector space of operators on  $\mathcal{H}_d^{\otimes N}$ .

Equation (5.47) already looks very much like our sought after goal, but we are not there quite yet. At this stage one has no right to assume that the

$A_{\mathbf{p}}$  are density operators. Indeed they generally are not: the integral (5.43) ranges over some points  $\mathbf{p}$  in  $\mathcal{S}_{d^2}$  that cannot be generated by applying the measurement  $\mathcal{E}$  to *any* quantum state. Hence some of the  $A_{\mathbf{p}}$  in the integral representation are ostensibly nonphysical.

The solution to this conundrum is provided by the overall requirement that  $\rho^{(N)}$  be a valid density operator. This requirement places a significantly more stringent constraint on the distribution  $P(\mathbf{p})$  than was the case in the classical representation theorem. In particular, it must be the case that  $P(\mathbf{p})$  vanishes whenever the corresponding  $A_{\mathbf{p}}$  is not a proper density operator. Let us move toward showing that.

We first need to delineate two properties of the operators  $A_{\mathbf{p}}$ . One is that they are Hermitian. The argument is simply

$$\text{tr}(E_{\alpha}A_{\mathbf{p}}^{\dagger}) = \text{tr}\left[(A_{\mathbf{p}}E_{\alpha})^{\dagger}\right] = [\text{tr}(A_{\mathbf{p}}E_{\alpha})]^* = \text{tr}(A_{\mathbf{p}}E_{\alpha}), \quad (5.48)$$

where the last step follows from Eq. (5.44). Because the  $E_{\alpha}$  are a complete set of linearly independent operators, it follows that  $A_{\mathbf{p}}^{\dagger} = A_{\mathbf{p}}$ . The second property tells us something about the eigenvalues of  $A_{\mathbf{p}}$ :

$$1 = \sum_{\alpha} p_{\alpha} = \text{tr}\left(A_{\mathbf{p}} \sum_{\alpha} E_{\alpha}\right) = \text{tr}A_{\mathbf{p}}. \quad (5.49)$$

In other words the (real) eigenvalues of  $A_{\mathbf{p}}$  must sum to unity.

We now show that these two facts go together to imply that if there are any nonphysical  $A_{\mathbf{p}}$  with positive weight  $P(\mathbf{p})$  in Eq. (5.47), then one can find a measurement for which  $\rho^{(N)}$  produces illegal “probabilities” for sufficiently large  $N$ . For instance, take a particular  $A_{\mathbf{q}}$  in Eq. (5.47) that has at least one negative eigenvalue  $-\lambda < 0$ . Let  $|\psi\rangle$  be a normalized eigenvector corresponding to that eigenvalue and consider the binary-valued POVM consisting of the elements  $\Pi = |\psi\rangle\langle\psi|$  and  $\tilde{\Pi} = I - \Pi$ . Since  $\text{tr}(A_{\mathbf{q}}\tilde{\Pi}) = -\lambda < 0$ , it is true by Eq. (5.49) that  $\text{tr}(A_{\mathbf{q}}\Pi) = 1 + \lambda > 1$ . Consider repeating this measurement over and over. In particular, let us tabulate the probability of getting outcome  $\Pi$  for every single trial to the exclusion of all other outcomes.

The gist of the contradiction is most easily seen by *imagining* that Eq. (5.47) is really a discrete sum:

$$\rho^{(N)} = P(\mathbf{q}) A_{\mathbf{q}}^{\otimes N} + \sum_{\mathbf{p} \neq \mathbf{q}} P(\mathbf{p}) A_{\mathbf{p}}^{\otimes N}. \quad (5.50)$$

The probability of  $N$  occurrences of the outcome  $\Pi$  is thus

$$\begin{aligned} \text{tr}(\rho^{(N)}\Pi^{\otimes N}) &= P(\mathbf{q}) \text{tr}(A_{\mathbf{q}}^{\otimes N}\Pi^{\otimes N}) + \sum_{\mathbf{p} \neq \mathbf{q}} P(\mathbf{p}) \text{tr}(A_{\mathbf{p}}^{\otimes N}\Pi^{\otimes N}) \\ &= P(\mathbf{q}) [\text{tr}(A_{\mathbf{q}}\Pi)]^N + \sum_{\mathbf{p} \neq \mathbf{q}} P(\mathbf{p}) [\text{tr}(A_{\mathbf{p}}\Pi)]^N \\ &= P(\mathbf{q})(1 + \lambda)^N + \sum_{\mathbf{p} \neq \mathbf{q}} P(\mathbf{p}) [\text{tr}(A_{\mathbf{p}}\Pi)]^N. \end{aligned} \quad (5.51)$$

There are no assurances in general that the right-hand term in Eq. (5.51) is positive, but if  $N$  is an even number it must be. It follows that if  $P(\mathbf{q}) \geq 0$ , for sufficiently large *even*  $N$ ,

$$\mathrm{tr}(\rho^{(N)} \Pi^{\otimes N}) > 1, \quad (5.52)$$

contradicting the assumption that it should always be a probability.

All we need to do now is transcribe the argument leading to Eq. (5.52) to the general integral case of Eq. (5.47). Note that by Eq. (5.44), the quantity  $\mathrm{tr}(A_{\mathbf{p}} \Pi)$  is a (linear) continuous function of the parameter  $\mathbf{p}$ . Therefore, for any  $\epsilon > 0$ , there exists a  $\delta > 0$  such that  $|\mathrm{tr}(A_{\mathbf{p}} \Pi) - \mathrm{tr}(A_{\mathbf{q}} \Pi)| \leq \epsilon$  whenever  $|\mathbf{p} - \mathbf{q}| \leq \delta$ , i.e., whenever  $\mathbf{p}$  is contained within an open ball  $B_{\delta}(\mathbf{q})$  centered at  $\mathbf{q}$ . Choose  $\epsilon < \lambda$ , and define  $\overline{B}_{\delta}$  to be the intersection of  $B_{\delta}(\mathbf{q})$  with the probability simplex. For  $\mathbf{p} \in \overline{B}_{\delta}$ , it follows that

$$\mathrm{tr}(A_{\mathbf{p}} \Pi) \geq 1 + \lambda - \epsilon > 1. \quad (5.53)$$

If we consider an  $N$  that is even,  $[\mathrm{tr}(A_{\mathbf{p}} \Pi)]^N$  is nonnegative in all of  $\mathcal{S}_{d^2}$ , and we have that the probability of the outcome  $\Pi^{\otimes N}$  satisfies

$$\begin{aligned} \mathrm{tr}(\rho^{(N)} \Pi^{\otimes N}) &= \int_{\mathcal{S}_{d^2}} P(\mathbf{p}) [\mathrm{tr}(A_{\mathbf{p}} \Pi)]^N d\mathbf{p} \\ &= \int_{\mathcal{S}_{d^2} - \overline{B}_{\delta}} P(\mathbf{p}) [\mathrm{tr}(A_{\mathbf{p}} \Pi)]^N d\mathbf{p} + \int_{\overline{B}_{\delta}} P(\mathbf{p}) [\mathrm{tr}(A_{\mathbf{p}} \Pi)]^N d\mathbf{p} \\ &\geq \int_{\overline{B}_{\delta}} P(\mathbf{p}) [\mathrm{tr}(A_{\mathbf{p}} \Pi)]^N d\mathbf{p} \\ &\geq (1 + \lambda - \epsilon)^N \int_{\overline{B}_{\delta}} P(\mathbf{p}) d\mathbf{p}. \end{aligned} \quad (5.54)$$

Unless

$$\int_{\overline{B}_{\delta}} P(\mathbf{p}) d\mathbf{p} = 0, \quad (5.55)$$

the lower bound (5.54) for the probability of the outcome  $\Pi^{\otimes N}$  becomes arbitrarily large as  $N \rightarrow \infty$ . Thus we conclude that the requirement that  $\rho^{(N)}$  be a proper density operator constrains  $P(\mathbf{p})$  to vanish almost everywhere in  $\overline{B}_{\delta}$  and, consequently, to vanish almost everywhere that  $A_{\mathbf{p}}$  is not a physical state.

Using Eq. (5.44), we can trivially transform the integral representation (5.47) to one directly over the convex set of density operators  $\mathcal{D}_d$  and be left with the following statement. Under the sole assumption that the density operator  $\rho^{(N)}$  is exchangeable, there exists a unique probability density  $P(\rho)$  such that

$$\rho^{(N)} = \int_{\mathcal{D}_d} P(\rho) \rho^{\otimes N} d\rho. \quad (5.56)$$

This concludes the proof of the quantum de Finetti representation theorem.

## 5.5 Intermezzo

In classical probability theory, exchangeability characterizes those situations where the only data relevant for updating a probability distribution are frequency data, i.e., the numbers  $n_j$  in Eq. (5.11) which tell how often the result  $j$  occurred. The quantum de Finetti representation shows that the same is true in quantum mechanics: Frequency data (with respect to a sufficiently robust measurement) are sufficient for updating an exchangeable state to the point where nothing more can be learned from sequential measurements; that is, one obtains a convergence of the form (5.7), so that ultimately any further measurements on the individual systems are statistically independent.

Beyond the aesthetic point of showing the consistency of the Bayesian conception of quantum state, we also believe the technical methods exhibited here will be of interest in the practical arena. Recently there has been a large literature on which classes of measurements have various advantages for tomographic purposes [71, 72]. To our knowledge, the work in Ref. [1] reiterated here was the first to consider tomographic reconstruction based upon minimal informationally complete POVMs. One can imagine several advantages to this approach via the fact that such POVMs with rank-one elements are automatically extreme points in the convex set of all measurements [73].

Furthermore, the classical de Finetti theorem is only the beginning with respect to general questions in classical statistics to do with exchangeability and generalizations of the concept [74]. One should expect no less of quantum exchangeability studies. In particular here, we are thinking of things like the question of representation theorems for finitely exchangeable distributions [52, 75]. Just as our method for proving the quantum de Finetti theorem was able to rely heavily on the classical theorem, so one might expect similar benefits from the classical results in the case of quantum finite exchangeability—although, there will certainly be new aspects to the quantum case due to the possibility of entanglement in finite exchangeable states [76]. Finally, a practical application of such representation theorems could be their potential to contribute to the solution of some outstanding problems in constructing security proofs for various quantum key distribution schemes [77, 78].

## 5.6 Subjectivity of Quantum Operations

At this point we turn our attention to quantum operations. So far, we have made much to-do of the subjectivity of quantum states. But what of quantum operations? Should this structural element within quantum theory be considered of the nature of a fact—like the outcome of a quantum measurement—or, like the quantum state, should it be recognized as a statement of an agent’s belief or expectation [4]?

The usual presentation of the concept of a quantum operation is that it is the most general quantum-state evolution allowed by the laws of quantum mechanics [63, 79]. Upon the process of measurement of some POVM  $\{E_\alpha\}$ , depending upon the particular details of a measurement interaction and the outcome  $\alpha$ , an initial quantum state  $\rho$  can change to any new state of the form

$$\Phi_\alpha(\rho) = \frac{1}{\text{tr } \rho E_\alpha} \sum_i A_{\alpha i} \rho A_{\alpha i}^\dagger, \quad (5.57)$$

where  $A_{\alpha i}$  can be any operators whatsoever, as long as they satisfy

$$E_\alpha = \sum_i A_{\alpha i}^\dagger A_{\alpha i}. \quad (5.58)$$

In the case where the POVM is the completely uninformative one,  $\{I\}$ , one recovers the general expression for all possible quantum mechanical time evolutions in terms of trace-preserving completely positive maps:

$$\rho \longrightarrow \Phi(\rho) = \frac{1}{\text{tr } \rho E_\alpha} \sum_\alpha A_\alpha \rho A_\alpha^\dagger, \quad (5.59)$$

where the  $A_\alpha$  are any operators that satisfy

$$I = \sum_\alpha A_\alpha^\dagger A_\alpha. \quad (5.60)$$

Let us first focus on the latter type of evolution. A quantum Bayesian should become suspicious that it contains a subjective component because of another representation theorem of Kraus [63]. For any trace preserving completely positive map  $\Phi$  on a system, one can always imagine an ancillary system  $A$ , a quantum state  $\sigma$  for that ancillary system, and unitary interaction  $U$  between the system and the ancilla, such that

$$\Phi(\rho) = \text{tr}_A (U(\rho \otimes \sigma) U^\dagger), \quad (5.61)$$

where  $\text{tr}_A$  represents a partial trace over the ancilla's Hilbert space. The Bayesian should ask, "Whose state of belief is  $\sigma$ ?"

But worse than that, the representation in Eq. (5.61) is not unique. There can well be distinct density operators  $\tau$  for the ancilla and distinct unitary interactions  $V$  such that

$$\Phi(\rho) = \text{tr}_A (V(\rho \otimes \tau) V^\dagger). \quad (5.62)$$

What is to be made of this? If one accepts that a quantum state is a subjective judgment, then it would seem one should also be compelled to accept that the map  $\Phi$  can be thought of as having (at least) a subjective component

through the quantum state  $\sigma$ , and that the subjective judgment even leaks into unitary operation  $U$ , nominally describing the interaction.

Without belaboring that particular point, let us now change focus to the state-change rule associated with the von Neumann collapse postulate. In that context, the measurement operators  $E_\alpha = |\alpha\rangle\langle\alpha|$  correspond to projectors onto an orthonormal basis  $\{|\alpha\rangle\}$  and

$$\Phi_\alpha(\rho) = \frac{1}{\text{tr } \rho E_\alpha} E_\alpha \rho E_\alpha = |\alpha\rangle\langle\alpha|. \quad (5.63)$$

The salient point is, conditioned on the measurement outcome, the final quantum state for the system is uniquely determined after such a measurement.

Note what this implies. If the quantum operation associated with a measurement device is an objective fact (i.e., of the same nature as the measurement outcome), then so too must be the posterior quantum state  $E_\alpha = |\alpha\rangle\langle\alpha|$ . But the whole foundation of this chapter is that quantum states are not objective facts. It follows that the quantum operation  $\Phi_\alpha$  cannot be an objective fact either. Instead, the von Neumann collapse rule associated with a projective measurement must be as much of a subjective judgment as a quantum state is in the first place.

The moral: (At least some) quantum operations are not facts, and general quantum time evolutions have a subjective component.<sup>11</sup> But, if so, then what are the “unknown quantum operations” that experimentalists routinely measure in the laboratory?

## 5.7 Quantum Process Tomography

In quantum process tomography [37–39], an experimenter lets an incompletely specified device act on a quantum system prepared in an input state of his choice, and then performs a measurement (also of his choice) on the output system. This procedure is repeated many times over, with possibly different input states and different measurements, in order to accumulate enough statistics to assign a quantum operation to the device. Here and throughout this section, by a quantum operation we mean a trace-preserving completely positive linear map. Quantum process tomography has been demonstrated experimentally in liquid state nuclear magnetic resonance [80, 81], and recently a number of optical experiments [82–84] have implemented entanglement-assisted quantum process tomography. The latter is a procedure that exploits the fact that quantum process tomography is equivalent to quantum state tomography in a larger state space [85–88].

---

<sup>11</sup> Ref. [4], in fact, tries to argue for more: Namely, that *all* quantum operations are subjective states of belief, just like all quantum states are subjective states of belief. The present work, however, refrains from attempting that larger task.

In the usual description of process tomography, it is assumed that the device performs the same *unknown* quantum operation  $\Phi$  every time it is used, and an experimenter's prior information about the device is expressed via a probability density  $p(\Phi)$  over all possible operations. What, however, is the operational meaning of an unknown quantum operation? When does the action of a device leave off from an initial input so that the next input can be sent through? In particular, what gives the right to suppose that a device does not have memory or, for instance, does not entangle the successive inputs passing through it? These questions boil down to the need to explore a single issue: What essential assumptions must be made so that quantum process tomography is a logically coherent notion?

What is called for is a method of posing quantum process tomography that never requires the invocation of the concept of an unknown quantum operation. This can be done by focusing upon the action of a single *known* quantum operation  $\Phi^{(N)}$ , which acts upon  $N$  nominal inputs. In particular, we identify conditions under which  $\Phi^{(N)}$ , ( $N = 1, 2, \dots$ ), can be represented as

$$\Phi^{(N)} = \int p(\Phi) \Phi^{\otimes N} d\Phi, \quad (5.64)$$

for some probability density  $p(\Phi)$ , and where the integration extends over all single-system quantum operations  $\Phi$ . With this theorem established, the conditions under which an experimenter can act *as if* his prior  $\Phi^{(N)}$  corresponds to *ignorance* of a “true” but unknown quantum operation are made precise.

Our starting point is the closely aligned and similarly motivated de Finetti representation theorem for quantum states of the previous sections. Here, we make use of the correspondence between quantum process tomography and quantum-state tomography mentioned above to derive a de Finetti representation theorem for sequences of quantum operations.

## 5.8 The Process-Tomography Theorem

In this section and the next, we restrict our attention to devices for which the input and output have the same Hilbert space dimension,  $D$ . In the following,  $\mathcal{H}_D$  denotes a  $D$ -dimensional Hilbert space,  $\mathcal{H}_D^{\otimes N} = \mathcal{H}_D \otimes \dots \otimes \mathcal{H}_D$  denotes its  $N$ -fold tensor product, and  $\mathcal{L}(\mathcal{V})$  denotes the space of linear operators on a linear space  $\mathcal{V}$ . The set of density operators for a  $D$ -dimensional quantum system is a convex subset of  $\mathcal{L}(\mathcal{H}_D)$ .

The action of a device on  $N$  nominal inputs systems is then described by a trace-preserving completely positive map

$$\Phi^{(N)} : \mathcal{L}(\mathcal{H}_D^{\otimes N}) \longrightarrow \mathcal{L}(\mathcal{H}_D^{\otimes N}), \quad (5.65)$$

which maps the state of the  $N$  input systems to the state of the  $N$  output systems. We will say, in analogy to the definition of exchangeability for quan-

tum states, that a quantum operation  $\Phi^{(N)}$  is *exchangeable* if it is a member of an exchangeable sequence of quantum operations.

To define exchangeability for a sequence of quantum operations in a natural way, we reduce the properties of symmetry and extendibility for sequences of operations to the corresponding properties for sequences of states. In the following, we will use bold letters to denote vectors of indices, e.g.  $\mathbf{j} = (j_1, \dots, j_N)$ . We will use  $\pi$  to denote a permutation of the set  $\{1, \dots, N\}$ , where the cardinality  $N$  will depend on the context. The action of the permutation  $\pi$  on the vector  $\mathbf{j}$  is defined by  $\pi\mathbf{j} = (j_{\pi(1)}, \dots, j_{\pi(N)})$ .

Any  $N$ -system density operator  $\rho^{(N)}$  can be expanded in the form

$$\rho^{(N)} = \sum_{\mathbf{j}, \mathbf{l}} r_{\mathbf{j}, \mathbf{l}}^{(N)} \bigotimes_{i=1}^N |j_i^{Q_i}\rangle \langle l_i^{Q_i}| \equiv \sum_{\mathbf{j}, \mathbf{l}} r_{\mathbf{j}, \mathbf{l}}^{(N)} |j_1^{Q_1}\rangle \langle l_1^{Q_1}| \otimes \cdots \otimes |j_N^{Q_N}\rangle \langle l_N^{Q_N}|, \quad (5.66)$$

where  $\{|1^{Q_i}\rangle, \dots, |D^{Q_i}\rangle\}$  denotes an orthonormal basis for the Hilbert space  $\mathcal{H}_D$  of the  $i$ th system, and  $r_{\mathbf{j}, \mathbf{l}}^{(N)}$  are the matrix elements of  $\rho^{(N)}$  in the tensor product basis. We define the action of the permutation  $\pi$  on the state  $\rho^{(N)}$  by

$$\pi\rho^{(N)} = \sum_{\mathbf{j}, \mathbf{l}} r_{\pi\mathbf{j}, \pi\mathbf{l}}^{(N)} \bigotimes_{i=1}^N |j_i^{Q_i}\rangle \langle l_i^{Q_i}| = \sum_{\mathbf{j}, \mathbf{l}} r_{\mathbf{j}, \mathbf{l}}^{(N)} \bigotimes_{i=1}^N |j_{\pi^{-1}(i)}^{Q_i}\rangle \langle l_{\pi^{-1}(i)}^{Q_i}|. \quad (5.67)$$

With this notation, we can make the following definition. A sequence of quantum operations,  $\Phi^{(k)} : \mathcal{L}(\mathcal{H}_D^{\otimes k}) \rightarrow \mathcal{L}(\mathcal{H}_D^{\otimes k})$ , is called *exchangeable* if, for  $k = 1, 2, \dots$ ,

1.  $\Phi^{(k)}$  is symmetric, i.e.,

$$\Phi^{(k)}(\rho^{(k)}) = \pi\left(\Phi^{(k)}(\pi^{-1}\rho^{(k)})\right) \quad (5.68)$$

for any permutation  $\pi$  of the set  $\{1, \dots, k\}$  and for any density operator  $\rho^{(k)} \in \mathcal{L}(\mathcal{H}_D^{\otimes k})$ , and

2.  $\Phi^{(k)}$  is extendible, i.e.,

$$\Phi^{(k)}(\text{tr}_{k+1} \rho^{(k+1)}) = \text{tr}_{k+1} \left( \Phi^{(k+1)}(\rho^{(k+1)}) \right) \quad (5.69)$$

for any state  $\rho^{(k+1)}$ .

In words, these conditions amount to the following. Condition (1) is equivalent to the requirement that the quantum operation  $\Phi^{(k)}$  commutes with any permutation operator  $\pi$  acting on the states  $\rho^{(k)}$ : It does not matter what order we send our systems through the device; as long as we rearrange them at the end into the original order, the resulting evolution will be the same. Condition (2) says that it does not matter if we consider a larger map

$\Phi^{(N+1)}$  acting on a larger collection of systems (possibly entangled), or a smaller  $\Phi^{(N)}$  on some subset of those systems: The upshot of the evolution will be the same for the relevant systems.

We are now in a position to formulate the de Finetti representation theorem for quantum operations. A quantum operation  $\Phi^{(N)} : \mathcal{L}(\mathcal{H}_D^{\otimes N}) \rightarrow \mathcal{L}(\mathcal{H}_D^{\otimes N})$  is an element of an exchangeable sequence if and only if it can be written in the form

$$\Phi^{(N)} = \int p(\Phi) \Phi^{\otimes N} d\Phi \quad \text{for all } N, \quad (5.70)$$

where the integral ranges over all single-shot quantum operations  $\Phi : \mathcal{L}(\mathcal{H}_D) \rightarrow \mathcal{L}(\mathcal{H}_D)$ ,  $d\Phi$  is a suitable measure on the space of quantum operations, and the probability density  $p(\Phi) \geq 0$  is unique. The tensor product  $\Phi^{\otimes N}$  is defined by  $\Phi^{\otimes N}(\rho_1 \otimes \cdots \otimes \rho_N) = \Phi(\rho_1) \otimes \cdots \otimes \Phi(\rho_N)$  for all  $\rho_1, \dots, \rho_N$  and by linear extension for arbitrary arguments.

Just as with the original quantum de Finetti theorem [1,42], this result allows a certain latitude in how quantum process tomography can be described. One is free to use the language of an unknown quantum operation if the condition of exchangeability is met by one's prior  $\Phi^{(N)}$  but it is not required: For the quantum Bayesian in particular, the *known* quantum operation  $\Phi^{(N)}$  is the only meaningful quantum operation in the problem.

## 5.9 Proof of the Process-Tomography Theorem

Let  $\Phi^{(N)}$ ,  $N = 1, 2, \dots$ , be an exchangeable sequence of quantum operations.  $\Phi^{(N)}$  can be characterized in terms of its action on the elements of a basis of  $\mathcal{L}(\mathcal{H}_D^{\otimes N})$  as follows.

$$\Phi^{(N)} \left( \bigotimes_{i=1}^N |j_i^{Q_i}\rangle \langle k_i^{Q_i}| \right) = \sum_{\mathbf{l}, \mathbf{m}} S_{\mathbf{l}, \mathbf{j}, \mathbf{m}, \mathbf{k}}^{(N)} \bigotimes_{i=1}^N |l_i^{Q_i}\rangle \langle m_i^{Q_i}|. \quad (5.71)$$

The coefficients  $S_{\mathbf{l}, \mathbf{j}, \mathbf{m}, \mathbf{k}}^{(N)}$  specify  $\Phi^{(N)}$  uniquely. It follows from a construction due to Choi [89] that the  $S_{\mathbf{l}, \mathbf{j}, \mathbf{m}, \mathbf{k}}^{(N)}$  can be regarded as the matrix elements of a density operator on  $D^{2N}$ -dimensional Hilbert space  $\mathcal{H}_{D^2}^{\otimes N}$ . This can be seen as follows. Let

$$|\Psi\rangle = \frac{1}{\sqrt{D}} \sum_{k=1}^D |k^{R_i}\rangle |k^{Q_i}\rangle \in \mathcal{H}_D \otimes \mathcal{H}_D = \mathcal{H}_{D^2} \quad (5.72)$$

be a maximally entangled state in  $\mathcal{H}_{D^2}$ , where the  $|k^{R_i}\rangle$  ( $k = 1, \dots, D$ ) form orthonormal bases for the ancillary systems labelled  $R_i$  ( $i = 1, \dots, N$ ). The corresponding density operator is

$$|\Psi\rangle\langle\Psi| = \frac{1}{D} \sum_{j,k} |j^{R_i}\rangle\langle k^{R_i}| \otimes |j^{Q_i}\rangle\langle k^{Q_i}| \in \mathcal{L}(\mathcal{H}_{D^2}) . \quad (5.73)$$

Similarly, we define a map,  $J$ , from the set of quantum operations on  $\mathcal{H}_D^{\otimes N}$  to the set of density operators on  $\mathcal{H}_{D^2}^{\otimes N}$  by

$$\begin{aligned} J(\Phi^{(N)}) &\equiv \left( I^{(N)} \otimes \Phi^{(N)} \right) \left( (|\Psi\rangle\langle\Psi|)^{\otimes N} \right) \\ &= \frac{1}{D^N} \left( I^{(N)} \otimes \Phi^{(N)} \right) \left( \sum_{\mathbf{j}, \mathbf{k}} \bigotimes_{i=1}^N (|j_i^{R_i}\rangle\langle k_i^{R_i}| \otimes |j_i^{Q_i}\rangle\langle k_i^{Q_i}|) \right) \\ &= \frac{1}{D^N} \sum_{\mathbf{l}, \mathbf{j}, \mathbf{m}, \mathbf{k}} S_{\mathbf{l}, \mathbf{j}, \mathbf{m}, \mathbf{k}}^{(N)} \bigotimes_{i=1}^N (|j_i^{R_i}\rangle\langle k_i^{R_i}| \otimes |l_i^{Q_i}\rangle\langle m_i^{Q_i}|) . \end{aligned} \quad (5.74)$$

In this definition,  $I^{(N)}$  denotes the identity operation acting on the ancillary systems  $R_1, \dots, R_N$ . The map  $J$  is injective, i.e.  $J(\Phi_1^{(N)}) = J(\Phi_2^{(N)})$  if and only if  $\Phi_1^{(N)} = \Phi_2^{(N)}$ .

The first stage of the proof of the de Finetti theorem for operations is to show that the density operators  $J(\Phi^{(N)})$ ,  $N = 1, 2, \dots$ , form an exchangeable sequence when regarded as  $N$ -system states, with  $R_i$  and  $Q_i$  jointly forming the  $i$ th system. To do this, we first show that  $J(\Phi^{(N)})$  is symmetric, i.e., invariant under an arbitrary permutation  $\pi$  of the  $N$  systems.

Note that since the density operators  $\rho^{(N)}$  actually span the whole vector space  $\mathcal{L}(\mathcal{H}_D^{\otimes N})$ , enforcing Definition 1 above amounts to identifying the linear maps on the left- and right-hand sides of Eqs. (5.68) and (5.69). I.e.,

$$\Phi^{(k)} = \pi \circ \Phi^{(k)} \circ \pi^{-1} \quad (5.75)$$

and

$$\Phi^{(k)} \circ \text{tr}_{k+1} = \text{tr}_{k+1} \circ \Phi^{(k+1)} \quad (5.76)$$

Thus in much that we do it suffices to consider the action of these maps on an arbitrary basis state  $E^{(N)} = \bigotimes_{i=1}^N |j_i^{Q_i}\rangle\langle k_i^{Q_i}|$  for arbitrary  $\mathbf{j}$  and  $\mathbf{k}$ . In particular,

$$\begin{aligned} \pi \left( \Phi^{(N)}(\pi^{-1} E^{(N)}) \right) &= \pi \left( \Phi^{(N)} \left( \bigotimes_{i=1}^N |j_{\pi(i)}^{Q_i}\rangle\langle k_{\pi(i)}^{Q_i}| \right) \right) \\ &= \pi \sum_{\mathbf{l}, \mathbf{m}} S_{\mathbf{l}, \pi\mathbf{j}, \mathbf{m}, \pi\mathbf{k}}^{(N)} \bigotimes_{i=1}^N |l_i^{Q_i}\rangle\langle m_i^{Q_i}| \\ &= \sum_{\mathbf{l}, \mathbf{m}} S_{\pi\mathbf{l}, \pi\mathbf{j}, \pi\mathbf{m}, \pi\mathbf{k}}^{(N)} \bigotimes_{i=1}^N |l_i^{Q_i}\rangle\langle m_i^{Q_i}| . \end{aligned} \quad (5.77)$$

Assuming Eq. (5.68), i.e., symmetry of  $\Phi^{(N)}$ , for all  $\mathbf{j}$  and  $\mathbf{k}$ , it follows that

$$S_{\pi \mathbf{l}, \pi \mathbf{j}, \pi \mathbf{m}, \pi \mathbf{k}}^{(N)} = S_{\mathbf{l}, \mathbf{j}, \mathbf{m}, \mathbf{k}}^{(N)} \quad (5.78)$$

for all  $\mathbf{l}, \mathbf{j}, \mathbf{m}, \mathbf{k}$ , which, using Eq. (5.74), implies that

$$\pi(J(\Phi^{(N)})) = J(\Phi^{(N)}) , \quad (5.79)$$

i.e., symmetry of  $J(\Phi^{(N)})$ .

To prove extendibility of  $J(\Phi^{(N)})$ , we introduce the following notation for partial traces: we denote by  $\text{tr}_{N+1}^R$  the partial trace over the subsystem  $R_{N+1}$ , and by  $\text{tr}_{N+1}^Q$  the partial trace over the subsystem  $Q_{N+1}$ . In this notation, we need to show that  $\text{tr}_{N+1}^R \text{tr}_{N+1}^Q J(\Phi^{(N+1)}) = J(\Phi^{(N)})$ . Using Eqs. (5.69) and (5.74),

$$\begin{aligned} & \text{tr}_{N+1}^R \text{tr}_{N+1}^Q J(\Phi^{(N+1)}) \\ &= \text{tr}_{N+1}^R \text{tr}_{N+1}^Q \frac{1}{D^{N+1}} \left( I^{(N+1)} \otimes \Phi^{(N+1)} \right) \\ &\quad \times \left( \sum_{\mathbf{j}, j_{N+1}, \mathbf{k}, k_{N+1}} \bigotimes_{i=1}^{N+1} (|j_i^{R_i}\rangle \langle k_i^{R_i}| \otimes |j_i^{Q_i}\rangle \langle k_i^{Q_i}|) \right) \\ &= \text{tr}_{N+1}^Q \frac{1}{D^{N+1}} \left( I^{(N)} \otimes \Phi^{(N+1)} \right) \left( \sum_{\mathbf{j}, \mathbf{k}, k_{N+1}} \bigotimes_{i=1}^N (|j_i^{R_i}\rangle \langle k_i^{R_i}| \otimes |j_i^{Q_i}\rangle \langle k_i^{Q_i}|) \right. \\ &\quad \left. \otimes |k_{N+1}^{Q_{N+1}}\rangle \langle k_{N+1}^{Q_{N+1}}| \right) \\ &= \frac{1}{D^{N+1}} \sum_{\mathbf{j}, \mathbf{k}, k_{N+1}} \left( \bigotimes_{i=1}^N (|j_i^{R_i}\rangle \langle k_i^{R_i}|) \otimes \text{tr}_{N+1}^Q \Phi^{(N+1)} \left( \bigotimes_{l=1}^N |j_l^{Q_i}\rangle \langle k_l^{Q_i}| \right. \right. \\ &\quad \left. \left. \otimes |k_{N+1}^{Q_{N+1}}\rangle \langle k_{N+1}^{Q_{N+1}}| \right) \right) \\ &= \frac{1}{D^{N+1}} \sum_{\mathbf{j}, \mathbf{k}, k_{N+1}} \left( \bigotimes_{i=1}^N (|j_i^{R_i}\rangle \langle k_i^{R_i}|) \otimes \Phi^{(N)} \left( \bigotimes_{l=1}^N |j_l^{Q_i}\rangle \langle k_l^{Q_i}| \right) \right) \\ &= \frac{1}{D^{N+1}} \left( I^{(N)} \otimes \Phi^{(N)} \right) \left( \sum_{\mathbf{j}, \mathbf{k}, k_{N+1}} \bigotimes_{i=1}^N (|j_i^{R_i}\rangle \langle k_i^{R_i}| \otimes |j_i^{Q_i}\rangle \langle k_i^{Q_i}|) \right) \\ &= \frac{1}{D^N} \left( I^{(N)} \otimes \Phi^{(N)} \right) \left( \sum_{\mathbf{j}, \mathbf{k}} \bigotimes_{i=1}^N (|j_i^{R_i}\rangle \langle k_i^{R_i}| \otimes |j_i^{Q_i}\rangle \langle k_i^{Q_i}|) \right) \\ &= J(\Phi^{(N)}) . \end{aligned} \quad (5.80)$$

We have thus shown that  $J(\Phi^{(N)})$ ,  $N = 1, 2, \dots$ , form an exchangeable sequence. According to the quantum de Finetti theorem for density operators, we can write

$$J(\Phi^{(N)}) = \int p(\rho) \rho^{\otimes N} d\rho , \quad (5.81)$$

where  $p(\rho) \geq 0$  is unique, and  $\int d\rho p(\rho) = 1$ . With the parameterization

$$\rho = \frac{1}{D} \sum_{l,j,m,k} S_{l,j,m,k}^{(1)} |j^R\rangle\langle k^R| \otimes |l^Q\rangle\langle m^Q| , \quad (5.82)$$

Eq. (5.81) takes the form

$$\begin{aligned} J(\Phi^{(N)}) &= \frac{1}{D^N} \int_{\mathcal{D}} dS p(S) \left( \sum_{l,j,m,k} S_{l,j,m,k}^{(1)} |j^R\rangle\langle k^R| \otimes |l^Q\rangle\langle m^Q| \right)^{\otimes N} \\ &= \frac{1}{D^N} \int_{\mathcal{D}} dS p(S) \bigotimes_{i=1}^N \sum_{l_i,j_i,m_i,k_i} S_{l_i,j_i,m_i,k_i}^{(1)} |j_i^{R_i}\rangle\langle k_i^{R_i}| \otimes |l_i^{Q_i}\rangle\langle m_i^{Q_i}| \quad (5.83) \\ &= \frac{1}{D^N} \sum_{\mathbf{l},\mathbf{j},\mathbf{m},\mathbf{k}} \int_{\mathcal{D}} dS p(S) \bigotimes_{i=1}^N S_{l_i,j_i,m_i,k_i}^{(1)} |j_i^{R_i}\rangle\langle k_i^{R_i}| \otimes |l_i^{Q_i}\rangle\langle m_i^{Q_i}| , \end{aligned}$$

where the integration variable is a vector with  $D^4$  components,  $S = (S_{1,1,1,1}^{(1)}, \dots, S_{D,D,D,D}^{(1)})$ , and where the integration domain,  $\mathcal{D}$ , is the set of all  $S$  that represent matrix elements of a density operator. The function  $p(S)$  is unique,  $p(S) \geq 0$ , and  $\int_{\mathcal{D}} dS p(S) = 1$ . Notice the slight abuse of notation in the first line of Eq. (5.83), where the superscripts  $R$  and  $Q$  label the entire sequences of systems  $R_1, \dots, R_N$  and  $Q_1, \dots, Q_N$ , respectively.

Comparing Eq. (5.83) with Eq. (5.74), we can express the coefficients  $S_{\mathbf{l},\mathbf{j},\mathbf{m},\mathbf{k}}^{(N)}$  specifying the quantum operation  $\Phi^{(N)}$  [see Eq. (5.71)] in terms of the integral above:

$$S_{\mathbf{l},\mathbf{j},\mathbf{m},\mathbf{k}}^{(N)} = \int_{\mathcal{D}} dS p(S) \prod_{i=1}^N S_{l_i,j_i,m_i,k_i}^{(1)} . \quad (5.84)$$

Hence, for any  $\mathbf{j}$  and  $\mathbf{k}$ ,

$$\begin{aligned} \Phi^{(N)} \left( \bigotimes_{i=1}^N |j_i^{Q_i}\rangle\langle k_i^{Q_i}| \right) &= \sum_{\mathbf{l},\mathbf{m}} \int_{\mathcal{D}} dS p(S) \left( \prod_{i=1}^N S_{l_i,j_i,m_i,k_i}^{(1)} \right) \bigotimes_{i=1}^N |l_i^{Q_i}\rangle\langle m_i^{Q_i}| \\ &= \int_{\mathcal{D}} dS p(S) \bigotimes_{i=1}^N \sum_{l_i,m_i} S_{l_i,j_i,m_i,k_i}^{(1)} |l_i^{Q_i}\rangle\langle m_i^{Q_i}| . \quad (5.85) \end{aligned}$$

The  $D^4$  coefficients,  $S_{l,j,m,k}^{(1)}$ , of the vector  $S$  define a single-system map,  $\Phi_S$ , via

$$\Phi_S(|j^Q\rangle\langle k^Q|) \equiv \sum_{l,m} S_{l,j,m,k}^{(1)} |l^Q\rangle\langle m^Q| \quad (j, k = 1, \dots, D) . \quad (5.86)$$

Hence

$$\begin{aligned}\Phi^{(N)}\left(\bigotimes_{i=1}^N |j_i^{Q_i}\rangle\langle k_i^{Q_i}|\right) &= \int_{\mathcal{D}} dS p(S) \bigotimes_{i=1}^N \Phi_S\left(|j_i^{Q_i}\rangle\langle k_i^{Q_i}|\right) \\ &= \int_{\mathcal{D}} dS p(S) \Phi_S^{\otimes N}\left(\bigotimes_{i=1}^N |j_i^{Q_i}\rangle\langle k_i^{Q_i}|\right).\end{aligned}\quad (5.87)$$

Since this equality holds for arbitrary  $\mathbf{j}$  and  $\mathbf{k}$ , it implies the representation

$$\Phi^{(N)} = \int_{\mathcal{D}} dS p(S) \Phi_S^{\otimes N}.\quad (5.88)$$

For all  $S \in \mathcal{D}$ , the map  $\Phi_S$  is completely positive. This can be seen by considering

$$J(\Phi_S) = (I \otimes \Phi_S)(|\Psi\rangle\langle\Psi|) = \frac{1}{D} \sum_{l,j,m,k} S_{l,j,m,k}^{(1)} |j^R\rangle\langle k^R| \otimes |l^Q\rangle\langle m^Q|,$$

which, by definition of  $\mathcal{D}$ , is a density operator and therefore positive. It follows from a theorem by Choi [89] that  $\Phi_S$  is completely positive.

To complete the proof, we will now show that  $p(S) = 0$  almost everywhere unless  $\Phi_S$  is trace-preserving, i.e., a quantum operation. More precisely, we show that if  $U \in \mathcal{D}$  is such that  $\Phi_U$  is not trace-preserving, then there exists an open ball  $B$  containing  $U$  such that  $p(S) = 0$  in  $B \cap \mathcal{D}$ .

The essence of the argument can be most easily explained in the special case that the integral (5.64) takes the form of a sum,

$$\Phi^{(N)} = \sum_i p_i \Phi_i^{\otimes N},\quad (5.89)$$

where  $p_i > 0$ . It follows that

$$1 = \sum_i p_i (\text{tr}[\Phi_i(\rho)])^N\quad (5.90)$$

for all single-system density operators  $\rho$ . Now assume that the sum extends over some non-trace-preserving operation, which we take to be  $\Phi_1$  without loss of generality. This means that

$$\text{tr}[\Phi_1(\rho)] \neq 1\quad (5.91)$$

for some single-system density operator  $\rho$ . Now either  $\text{tr}[\Phi_1(\rho)] < 1$ , in which case normalization of  $\Phi^{(1)}(\rho)$  implies that  $\text{tr}[\Phi_k(\rho)] > 1$  for some  $k \neq 1$ , or  $\text{tr}[\Phi_1(\rho)] > 1$ , in which case we set  $k = 1$ . In both cases

$$p_k (\text{tr}[\Phi_k(\rho)])^N \rightarrow \infty,\quad (5.92)$$

which contradicts Eq. (5.90). We have thus shown that the sum (5.89) extends only over trace-preserving operations.

Now let us return to the general case where  $\Phi^{(N)}$  is represented by an integral. For  $\delta > 0$  and  $U \in \mathcal{D}$ , we define  $B_\delta(U)$  to be the set of all  $S$  such that  $|S - U| < \delta$ , i.e.,  $B_\delta(U)$  is the open ball of radius  $\delta$  centered at  $U$ . Furthermore, we define  $\bar{B}_\delta(U) = B_\delta(U) \cap \mathcal{D}$ .

Let  $U \in \mathcal{D}$  be such that  $\Phi_U$  is not trace-preserving, i.e., there exists a density operator  $\rho$  for which  $\text{tr}[\Phi_U(\rho)] \neq 1$ . We distinguish two cases.

**Case (i):**  $\text{tr}[\Phi_U(\rho)] = 1 + \epsilon$ , where  $\epsilon > 0$ . Since  $\text{tr}[\Phi_S(\rho)]$  is a linear and therefore continuous function of the vector  $S$ , there exists  $\delta > 0$  such that

$$\left| \text{tr}[\Phi_S(\rho)] - \text{tr}[\Phi_U(\rho)] \right| < \epsilon/2 \quad (5.93)$$

whenever  $S \in B_\delta(U)$ . For  $S \in \bar{B}_\delta(U)$ ,

$$\text{tr}[\Phi_S(\rho)] > 1 + \epsilon - \epsilon/2 = 1 + \epsilon/2. \quad (5.94)$$

Therefore

$$\begin{aligned} \text{tr}[\Phi^{(N)}(\rho^{\otimes N})] &= \text{tr} \left[ \int_{\mathcal{D}} dS p(S) \Phi_S^{\otimes N}(\rho^{\otimes N}) \right] \\ &= \int_{\mathcal{D}} dS p(S) (\text{tr}[\Phi_S(\rho)])^N \\ &= \int_{\mathcal{D} \setminus \bar{B}_\delta(U)} dS p(S) (\text{tr}[\Phi_S(\rho)])^N + \int_{\bar{B}_\delta(U)} dS p(S) (\text{tr}[\Phi_S(\rho)])^N \quad (5.95) \\ &\geq \int_{\bar{B}_\delta(U)} dS p(S) (\text{tr}[\Phi_S(\rho)])^N \\ &> (1 + \epsilon/2)^N \int_{\bar{B}_\delta(U)} dS p(S). \end{aligned}$$

Unless  $\int_{\bar{B}_\delta(U)} dS p(S) = 0$ , there exists  $N$  such that  $\text{tr}[\Phi^{(N)}(\rho^{\otimes N})] > 1$ , which contradicts the assumption that  $\Phi^{(N)}$  is trace-preserving. Hence  $p(S) = 0$  almost everywhere in  $\bar{B}_\delta(U)$ .

**Case (ii):**  $\text{tr}[\Phi_U(\rho)] = 1 - \epsilon$ , where  $0 < \epsilon \leq 1$ . Because of continuity, there exists  $\delta > 0$  such that

$$\left| \text{tr}[\Phi_S(\rho)] - \text{tr}[\Phi_U(\rho)] \right| < \epsilon/2 \quad (5.96)$$

whenever  $S \in B_\delta(U)$ . Hence, for  $S \in \bar{B}_\delta(U)$ ,

$$\text{tr}[\Phi_S(\rho)] < 1 - \epsilon + \epsilon/2 = 1 - \epsilon/2. \quad (5.97)$$

Now assume that  $\int_{\bar{B}_\delta(U)} dS p(S) = \eta > 0$ . Then, letting  $N = 1$ ,

$$\begin{aligned}
1 = \text{tr} [\Phi^{(1)}(\rho)] &= \text{tr} \left[ \int_{\mathcal{D}} dS p(S) \Phi_S(\rho) \right] \\
&= \int_{\mathcal{D} \setminus \bar{B}_\delta(U)} dS p(S) \text{tr}[\Phi_S(\rho)] + \int_{\bar{B}_\delta(U)} dS p(S) \text{tr}[\Phi_S(\rho)] \\
&< \int_{\mathcal{D} \setminus \bar{B}_\delta(U)} dS p(S) \text{tr}[\Phi_S(\rho)] + \eta(1 - \epsilon/2), \tag{5.98}
\end{aligned}$$

which implies that

$$\int_{\mathcal{D} \setminus \bar{B}_\delta(U)} dS p(S) \text{tr}[\Phi_S(\rho)] > 1 - \eta + \eta\epsilon/2 > 1 - \eta. \tag{5.99}$$

Since

$$\int_{\mathcal{D} \setminus \bar{B}_\delta(U)} dS p(S) = 1 - \eta, \tag{5.100}$$

it follows that there exist  $\zeta > 0$  and a point  $V \in \mathcal{D} \setminus \bar{B}_\delta(U)$  such that  $\text{tr}[\Phi_V(\rho)] > 1$  and

$$\int_{\bar{B}_\xi(V)} dS p(S) > 0 \quad \text{for all } \xi \leq \zeta. \tag{5.101}$$

We are thus back to case (i) above. Repeating the argument of case (i) one can show that this contradicts the assumption that  $\Phi^{(N)}$  is trace preserving for large  $N$ . It follows that  $\eta = 0$ , i.e.,  $p(S) = 0$  almost everywhere in  $\bar{B}_\delta(U)$ . This concludes the proof of the de Finetti theorem for quantum operations.

What we have proven here is a representation theorem. It shows us when an experimenter is warranted to think of his (prior) *known* quantum operation assignment as built out of a lack of knowledge of a “true” but *unknown* one. In that way, the theorem has the same kind of attraction as the previous de Finetti theorem for quantum states.

In particular for a Bayesian interpretation of quantum mechanics, it may be a necessary ingredient for its very consistency. In Refs. [4, 6], it has been argued strenuously that quantum operations should be considered of essentially the same physical meaning and status as quantum states themselves: They are Bayesian expressions of an experimenter’s judgment. This could be captured in the slogan “a quantum operation is really a quantum state in disguise.” In other words, the Choi representation theorem [89] is not just a mathematical nicety, but is instead of deep physical significance.<sup>12</sup>

Therefore, just as an unknown quantum state is an oxymoron in a Bayesian interpretation of quantum mechanics, so should be an unknown quantum operation. In the case of quantum states, the conundrum is solved by the existence of a de Finetti theorem for quantum tomography. In this section we have shown that the conundrum in quantum process tomography can be solved in the same way.

---

<sup>12</sup> There have been a few pieces of recent technical work that may be useful for shoring up this idea. On the contingency, see Ref. [90].

## 5.10 Concluding Remarks

Is there something in nature even when there are no observers or agents about? At the practical level, it would seem hard to deny this, and neither of the authors wish to be viewed as doing so. The world persists without the observer—there is no doubt in either of our minds about that. But then, does that require that two of the most celebrated elements (namely, quantum states and operations) in quantum theory—our best, most all-encompassing scientific theory to date—must be viewed as objective, agent-independent constructs? There is no reason to do so, we say. In fact, we think there is everything to be gained from carefully delineating which part of the structure of quantum theory is about the world and which part is about the agent's interface with the world.

From this perspective, much—but not all—of quantum mechanics is about disciplined uncertainty accounting, just as is Bayesian probability theory in general. Bernardo and Smith [11] write this of Bayesian theory,

What is the nature and scope of Bayesian Statistics ... ?

Bayesian Statistics offers a rationalist theory of personalistic beliefs in contexts of uncertainty, with the central aim of characterizing how an individual should act in order to avoid certain kinds of undesirable behavioural inconsistencies. The theory establishes that expected utility maximization provides the basis for rational decision making and that Bayes' theorem provides the key to the ways in which beliefs should fit together in the light of changing evidence. The goal, in effect, is to establish rules and procedures for individuals concerned with disciplined uncertainty accounting. The theory is not descriptive, in the sense of claiming to model actual behaviour. Rather, it is prescriptive, in the sense of saying “if you wish to avoid the possibility of these undesirable consequences you must act in the following way.

In fact, one might go further and say of quantum theory, that in those cases where it is not just Bayesian probability theory full stop, it is a theory of stimulation and response [4, 22]. The agent, through the process of quantum measurement stimulates the world external to himself. The world, in return, stimulates a response in the agent that is quantified by a change in his beliefs—i.e., by a change from a prior to a posterior quantum state. Somewhere in the structure of those belief changes lies quantum theory's most direct statement about what we believe of the world as it is without agents.

The present effort, showing how a Bayesian account of quantum states and operations is fully consistent with the laboratory practices of quantum-state and process tomography, is a necessary exercise along the way to pinpointing that direct statement.

## Acknowledgments

We thank Carl Caves and David Mermin for the extensive discussions that brought part of this chapter about. CAF thanks Science Foundation Ireland for the support of an E. T. S. Walton Visitor Award.

## References

1. C. M. Caves, C. A. Fuchs, and R. Schack: J. Math. Phys. **43**, 4537 (2002).
2. R. Schack, T. A. Brun, and C. M. Caves: Phys. Rev. A **64**, 014305 (2001).
3. C. A. Fuchs, R. Schack, and P. F. Scudo: A de Finetti Representation Theorem for Quantum Process Tomography, to appear in Phys. Rev. A. See also arXiv:quant-ph/0307198.
4. C. A. Fuchs: Quantum States: What the Hell Are They?, posted at <http://netlib.bell-labs.com/who/cafuchs>.
5. C. M. Caves, C. A. Fuchs and R. Schack: Phys. Rev. A **66**, 062111 (2002).
6. C. A. Fuchs: Quantum Mechanics as Quantum Information (and only a little more), arXiv:quant-ph/0205039.
7. E. T. Jaynes, *Probability Theory: The Logic of Science*, ed by G. L. Bretthorst (Cambridge University Press, Cambridge, 2003).
8. B. de Finetti: Erkenntnis **31**, 169–223 (1989).
9. B. de Finetti: *Theory of Probability* (Wiley, New York, 1990).
10. L. J. Savage: *The Foundations of Statistics* (Dover, New York, 1972).
11. J. M. Bernardo and A. F. M. Smith: *Bayesian Theory* (Wiley, Chichester, 1994).
12. R. Jeffrey: Subjective Probability (The Real Thing), posted at <http://www.princeton.edu/~bayesway/>.
13. C. A. Fuchs and A. Peres: Phys. Today **53**, 70 (2000); *ibid* **53**, 14 (2000).
14. C. M. Caves, C. A. Fuchs and R. Schack: Phys. Rev. A **65**, 022305 (2002).
15. C. M. Caves and C. A. Fuchs: Quantum Information: How Much Information in a State Vector? In: *The Dilemma of Einstein, Podolsky and Rosen - 60 Years Later (An International Symposium in Honour of Nathan Rosen - Haifa, March 1995)*, ed by A. Mann and M. Revzen, Ann. Israel Phys. Soc. **12**, 226 (1996).
16. T. A. Brun, C. M. Caves, and R. Schack: Phys. Rev. A **63**, 042309 (2001).
17. C. A. Fuchs: J. Mod. Opt. **50**, 987 (2003).
18. R. Schack: Found. Phys. **33**, 1461 (2003).
19. D. M. Appleby: Facts, Values and Quanta, arXiv:quant-ph/0402015.
20. S. J. van Enk and C. A. Fuchs: Phys. Rev. Lett. **88**, 027902 (2002); S. J. van Enk and C. A. Fuchs: Quant. Info. Comp. **2**, 151 (2002).
21. C. A. Fuchs: The Anti-Växjö Interpretation of Quantum Mechanics, arXiv:quant-ph/0204146.
22. C. A. Fuchs: *Notes on a Paulian Idea: Foundational, Historical, Anecdotal & Forward-Looking Thoughts on the Quantum* (Växjö University Press, Växjö, Sweden, 2003). See also arXiv:quant-ph/0105039.
23. A. Peres: Am. J. Phys. **52**, 644 (1984).

24. N. D. Mermin: Whose Knowledge? In *Quantum (Un)speakables: From Bell to Quantum Information*, ed by R. A. Bertlmann and A. Zeilinger (Springer-Verlag, Berlin, 2002).
25. T. A. Brun, J. Finkelstein, and N. D. Mermin: Phys. Rev. A **65**, 032315 (2002).
26. N. D. Mermin: J. Math. Phys. **43**, 4560 (2002).
27. R. W. Spekkens: In Defense of the Epistemic View of Quantum States: A Toy Theory, arXiv:quant-ph/0401052; L. Hardy: Quantum Theory From Five Reasonable Axioms, arXiv:quant-ph/0101012; O. Cohen: Classical Teleportation of Classical States, arXiv:quant-ph/0310017; R. Duvenhage: Found. Phys. **32**, 1399 (2002); P. G. L. Mana: Why Can States and Measurement Outcomes Be Represented as Vectors?, arXiv:quant-ph/0305117; A. Duwell, Quantum Information Does Not Exist, Stud. Hist. Phil. Mod. Phys. **34**, 479 (2003); I. Pitowski: Stud. Hist. Phil. Mod. Phys. **34**, 395 (2003); R. Clifton, J. Bub, and H. Halvorson: Found. Phys. **33**, 1561 (2003); J. Bub: Why the Quantum?, to appear in Stud. Hist. Phil. Mod. Phys., see also arXiv:quant-ph/0402149; D. M. Appleby: The Bell-Kochen-Specker Theorem, arXiv:quant-ph/0308114.
28. C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters: Phys. Rev. Lett. **70**, 1895 (1993).
29. D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu: Phys. Rev. Lett. **80**, 1121 (1998); D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger: Nature **390**, 575 (1997); A. Furusawa, J. L. Sørensen, S. L. Braunstein, C. A. Fuchs, H. J. Kimble, and E. S. Polzik: Science **282**, 706 (1998).
30. P. W. Shor: Phys. Rev. A **52**, R2493 (1995).
31. A. M. Steane: Phys. Rev. Lett. **77**, 793 (1996).
32. C. H. Bennett and G. Brassard: Quantum Cryptography: Public Key Distribution and Coin Tossing. In *Proc. IEEE International Conference on Computers, Systems and Signal Processing* (IEEE Press, New York, 1984) p. 175; C. H. Bennett: Phys. Rev. Lett. **68**, 3121 (1992).
33. A. Muller, H. Zbinden, and N. Gisin: Nature **378**, 449 (1995); W. T. Buttler, R. J. Hughes, P. G. Kwiat, S. K. Lamoreaux, G. G. Luther, G. L. Morgan, J. E. Nordholt, C. G. Peterson, and C. M. Simmons, Phys. Rev. Lett. **81**, 3283 (1998); R. J. Hughes, G. L. Morgan, and C. G. Peterson: J. Mod. Opt. **47**, 533 (2000).
34. K. Vogel and H. Risken: Phys. Rev. A **40**, 2847 (1989).
35. D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani: Phys. Rev. Lett. **70**, 1244 (1993).
36. U. Leonhardt: Phys. Rev. Lett. **74**, 4101 (1995).
37. Q. A. Turchette *et al.*: Phys. Rev. Lett. **75**, 4710 (1995).
38. I. L. Chuang and M. A. Nielsen: J. Mod. Opt. **44**, 2455 (1997).
39. J. F. Poyatos, J. I. Cirac, and P. Zoller: Phys. Rev. Lett. **78**, 390 (1997).
40. H. E. Kyburg, Jr. and H. E. Smokler, eds, *Studies in Subjective Probability*, 2nd edn (Robert E. Krieger Publishing, Huntington, NY, 1980).
41. For a collection of de Finetti's original papers and their translations into English, see P. Monari and D. Cocchi, eds, *Probabilità e Induzione—Induction and Probability* (Biblioteca di Statistica, CLUEB, Bologna, 1993).
42. R. L. Hudson and G. R. Moody: Z. Wahrschein. verw. Geb. **33**, 343 (1976).
43. R. L. Hudson: Found. Phys. **11**, 805 (1981).

44. D. Petz: Prob. Th. Rel. Fields **85**, 1 (1990); A. Bach: Europhys. Lett. **16**, 513 (1991); L. Accardi and Y. G. Lu: Ann. Prob. **21**, 1478 (1993); A. Bach: *Indistinguishable Classical Particles*, Lecture Notes in Mathematics, New Series, vol 44 (Springer, Berlin, 1997); R. L. Hudson: Int. J. Quant. Chem. **74**, 595 (1999).
45. C. A. Fuchs and K. Jacobs: Phys. Rev. A **63**, 062305 (2001).
46. D. Heath and W. Sudderth: Am. Stat. **30**, 188 (1976).
47. L. Daston: Hist. Math. **21**, 330 (1994).
48. T. A. Bass: *The Newtonian Casino* (Penguin Books, London, 1991), previously published as *The Eudaemonic Pie: Or why Would Anyone Play Roulette without a Computer in His Shoe* (Houghton Mifflin, New York, 1985).
49. R. N. Giere: Objective Single-Case Probabilities and the Foundations of Statistics. In *Logic, Methodology and Philosophy of Science IV*, ed by P. Suppes, L. Henkin, A. Jojo, and G. C. Moisil (North-Holland, Amsterdam, 1973) p. 467.
50. J. S. Bell: *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy* (Cambridge U. Press, Cambridge, 1987).
51. J. T. Cushing, A. Fine, and S. Goldstein, eds, *Bohmian Mechanics and Quantum Theory: An Appraisal* (Kluwer, Dordrecht, 1996).
52. E. T. Jaynes: Some Applications and Extensions of the de Finetti Representation Theorem. In *Bayesian Inference and Decision Techniques*, ed by P. Goel and A. Zellner (Elsevier, Amsterdam, 1986) p. 31.
53. A. Peres: Phys. Rev. A **61**, 022116 (2000).
54. V. Scarani, W. Tittel, H. Zbinden, and N. Gisin: Phys. Lett. A **276**, 1 (2000).
55. E. Schrödinger: Proc. Cam. Philo. Soc. **32**, 446 (1936).
56. E. T. Jaynes: Phys. Rev. **108**, 171 (1957).
57. L. P. Hughston, R. Jozsa, and W. K. Wootters: Phys. Lett. A **183**, 14 (1993).
58. E. Størmer: J. Func. Anal. **3**, 48 (1969).
59. C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters: Phys. Rev. A **54**, 3824 (1996).
60. N. D. Mermin: Phys. Today **43**, 9 (1990).
61. M. Koashi, V. Bužek, and N. Imoto: Phys. Rev. A **62**, 050302 (2000).
62. E. B. Davies and J. T. Lewis: Comm. Math. Phys. **17**, 239 (1970).
63. K. Kraus: *States, Effects, and Operations. Fundamental Notions of Quantum Theory*, Lecture Notes in Physics, vol 190 (Springer, Berlin, 1983).
64. A. Peres: *Quantum Theory: Concepts and Methods* (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993).
65. E. Merzbacher: *Quantum Mechanics*, 2nd edn (Wiley, New York, 1970); C. Cohen-Tannoudji: *Quantum Mechanics*, 2nd revised enlarged edition (Wiley, New York, 1977).
66. J. von Neumann: *Mathematical Foundations of Quantum Mechanics*, translated by T. A. Beyer (Princeton U. Press, Princeton, 1955); P. A. M. Dirac: *The Principles of Quantum Mechanics*, 4th edn (Oxford University Press, Oxford, 1958).
67. A. S. Holevo: Prob. Info. Trans. **9**, 110 (1973).
68. E. Prugovečki: Int. J. Theo. Phys. **16**, 321 (1977).
69. G. M. D'Ariano, P. Perinotti, M. F. Sacchi: Informationally complete measurements and groups representation, arXiv:quant-ph/0310013.

70. J. M. Renes, R. Blume-Kohout, A. J. Scott, C. M. Caves: Symmetric Informationally Complete Quantum Measurements to appear in *J. Math. Phys.* (2004); see also arXiv:quant-ph/0310075.
71. For a small sampling of more recent considerations, see: G. M. D'Ariano, L. Maccone, and M. G. A. Paris: *J. Phys. A* **34**, 93 (2001); S. Weigert: *Phys. Rev. Lett.* **84**, 802 (2000); V. Bužek, G. Drobný, R. Derka, G. Adam, and H. Wiedemann: *Chaos Sol. Fract.* **10**, 981 (1999).
72. This question appears to have been considered much earlier than the current interest: W. Band and J. L. Park: *Found. Phys.* **1**, 133 (1970); J. L. Park and W. Band: *Found. Phys.* **1**, 211 (1971); W. Band and J. L. Park: *Found. Phys.* **1**, 339 (1971).
73. A. Fujiwara and H. Nagaoka: *IEEE Trans. Inf. Theory* **44**, 1071 (1998).
74. One can get a feeling for this from the large review article, D. J. Aldous: Exchangeability and Related Topics. In *École d'Été de Probabilités de Saint-Flour XIII – 1983*, ed by P. L. Hennequin, Lecture Notes in Mathematics vol 1117 (Springer-Verlag, Berlin, 1985) pp. 1–198.
75. P. Diaconis: *Synthese* **36**, 271 (1977); P. Diaconis and D. Freedman: *Ann. Prob.* **8**, 745 (1980).
76. G. G. Emch: Is There a Quantum de Finetti Programme? In *Proceedings of the Conference: Foundations of Probability and Physics – 2*, ed by A. Khrennikov (Växjö University Press, Växjö, Sweden, 2002) pp. 159–178.
77. H.-K. Lo, H. F. Chau, and M. Ardehali: Efficient Quantum Key Distribution Scheme And Proof of Its Unconditional Security, arXiv:quant-ph/0011056.
78. K. Tamaki, M. Koashi, and N. Imoto: *Phys. Rev. Lett.* **90**, 167904 (2003).
79. M. A. Nielsen and I. L. Chuang: *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
80. M. A. Nielsen, E. Knill, and R. Laflamme: *Nature* **396**, 52 (1998).
81. A. M. Childs, I. L. Chuang, and D. W. Leung: *Phys. Rev. A* **64**, 012314 (2001).
82. Y. Nambu *et al.*: Experimental Investigation of Pulsed Entangled Photons and Photonic Quantum Channels. In *Quantum Optics in Computing and Communications, Proceedings of SPIE*, vol 4917, ed by S. Liu, G. Guo, H.-K. Lo, and N. Imoto (2002), pp. 13–24; e-print arXiv:quant-ph/0210147.
83. J. B. Altepeter *et al.*: *Phys. Rev. Lett.* **90**, 193601 (2003).
84. F. De Martini, A. Mazzei, M. Ricci, and G. M. D'Ariano: *Phys. Rev. A* **67**, 062307 (2003).
85. D. W. Leung: Towards Robust Quantum Computation. PhD thesis, Stanford University (2000); e-print cs.CC/0012017.
86. D. W. Leung: *J. Math. Phys.* **44**, 528 (2003).
87. W. Dür and J. I. Cirac: *Phys. Rev. A* **64**, 012317 (2001).
88. G. M. D'Ariano and P. Lo Presti: *Phys. Rev. Lett.* **86**, 4195 (2001).
89. M.-D. Choi: *Lin. Alg. App.* **10**, 285 (1975).
90. G. M. D'Ariano and P. Lo Presti: Imprinting a Complete Information about a Quantum Channel on Its Output State, arXiv:quant-ph/0211133; K. Życzkowski and I. Bengtsson: On duality between Quantum Maps and Quantum States, arXiv:quant-ph/0401119; P. Arrighi and C. Patricot: On Quantum Operations as Quantum States, arXiv:quant-ph/0307024; J. Oppenheim and B. Reznik: A Probabilistic and Information Theoretic Interpretation of Quantum Evolutions, arXiv:quant-ph/0312149; A. Fujiwara and P. Algoet, *Phys. Rev. A* **59**, 3290–3294 (1999).

# 6 Quantum Tomography from Incomplete Data via *MaxEnt* Principle

Vladimír Bužek<sup>1,2</sup>

<sup>1</sup> Research Center for Quantum Information, Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 845 11 Bratislava, Slovakia

<sup>2</sup> Faculty of Informatics, Masaryk University, Botanická 68a, 602 00 Brno, Czech Republic  
`buzek@savba.sk`

*...existing quantum theory must be supplemented with some principle that tells us how to translate, or encode, the results of measurements into a definite state description  $\hat{\rho}$ . Note that the problem is not to find  $\hat{\rho}$  which correctly describes true physical situation. That is unknown, and always remains so, because of incomplete information. In order to have a usable theory we must ask the much more modest question: What  $\hat{\rho}$  best describes our state of knowledge about the physical situation?*

E. T. Jaynes [1]

## 6.1 Modest question

The concept of a quantum state represents one of the most fundamental pillars of the paradigm of quantum theory [2–4]. Contrary to its mathematical elegance and convenience in calculations, the physical interpretation of a quantum state is not so transparent. The problem is that the quantum state (described either by a state vector, or density operator or a phase-space probability density distribution) does not have a well-defined objective status, i.e. a state vector is not an *objective* property of a particle. According to Peres (see [2], pp. 373-374): “...there is no physical evidence whatsoever that every physical system has at every instant a well-defined state... In a strict interpretation of quantum theory these mathematical symbols [i.e., state vectors] represent *statistical information* enabling us to compute the *probabilities* of occurrence of specific events.” Once this point of view is adopted then it becomes clear that any reconstruction of a density operator (or its mathematical equivalent) can be understood exclusively as an expression of our knowledge about the quantum mechanical state based on a certain set of measured data. To be more specific, any quantum-mechanical reconstruction scheme is nothing more than an *a posteriori* estimation of the density operator of a quantum-mechanical (microscopic) system based on data obtained with the help of a macroscopic measurement apparatus [4]. The quality of

the reconstruction depends on the “quality” of the measured data and the efficiency of the reconstruction procedure with the help of which the data analysis is performed. In general, we can specify three different situations:

- Firstly, when all system observables are precisely measured. In this case the *complete reconstruction* of an initially unknown state can be performed. We can call this the reconstruction on the *complete* observation level. A typical example is a tomographic reconstruction of quantum states of light as discussed in Section 6.2 of this Chapter.
- Secondly, when just part of the system observables is precisely measured then one cannot perform a complete reconstruction of the measured state. Nevertheless, the reconstructed density operator still uniquely determines mean values of the measured observables. We can denote this scheme as reconstruction on *incomplete* observation levels. In this Chapter we will investigate in detail this specific situation and we will present a very efficient method of reconstruction of quantum states based on the principle of *Maximum entropy* (MaxEnt).
- Finally, when measurement does not provide us with sufficient information to specify the exact mean values (or probability distributions) but only frequencies of appearances of eigenstates of the measured observables, then one can perform an estimation (reconstruction) using methods such as maximum likelihood estimation (see Chapter by Hradil *et al.* of this book) or quantum Bayesian inference (see, for instance, the recent recent review [5]).

## 6.2 Complete observation level

Providing all system observables (i.e., the quorum [2, 4]) have been precisely measured, then the density operator of a quantum-mechanical system can be completely reconstructed (i.e., the density operator can be uniquely determined based on the available data). In principle, we can consider two different schemes for reconstruction of the density operator of the given quantum-mechanical system. The difference between these two schemes is based on the way in which information about the quantum-mechanical system is obtained. The first type of measurement is such that on each element of the ensemble of the measured states only a *single* observable is measured. In the second type of measurement a *simultaneous* measurement of conjugate observables is assumed. We note that in both cases we will assume ideal, i.e., unit-efficiency, measurements. In what follows as an illustration we will consider a specific complete observation level that is realized via a quantum homodyne tomography of a single-mode electromagnetic field.

### 6.2.1 Quantum states of light

Utilizing a close analogy between the operator for the electric component  $\hat{E}(r, t)$  of a monochromatic light field and the quantum-mechanical harmonic oscillator we will consider a dynamical system that is described by a pair of canonically conjugated Hermitean observables  $\hat{q}$  and  $\hat{p}$ ,

$$[\hat{q}, \hat{p}] = i\hbar. \quad (6.1)$$

Eigenvalues of these operators range continuously from  $-\infty$  to  $+\infty$ . The annihilation and creation operators  $\hat{a}$  and  $\hat{a}^\dagger$  can be expressed as a complex linear combination of  $\hat{q}$  and  $\hat{p}$ :

$$\hat{a} = \frac{1}{\sqrt{2\hbar}} (\lambda\hat{q} + i\lambda^{-1}\hat{p}); \quad \hat{a}^\dagger = \frac{1}{\sqrt{2\hbar}} (\lambda\hat{q} - i\lambda^{-1}\hat{p}), \quad (6.2)$$

where  $\lambda$  is a real parameter. The operators  $\hat{a}$  and  $\hat{a}^\dagger$  obey the Weyl-Heisenberg commutation relation

$$[\hat{a}, \hat{a}^\dagger] = 1, \quad (6.3)$$

and therefore possess the same algebraic properties as the operator associated with the complex amplitude of a harmonic oscillator (in this case  $\lambda = \sqrt{m\omega}$ , where  $m$  and  $\omega$  are the mass and the frequency of the quantum-mechanical oscillator, respectively) or the photon annihilation and creation operators of a single mode of the quantum electromagnetic field. In this case  $\lambda = \sqrt{\epsilon_0\omega}$  ( $\epsilon_0$  is the dielectric constant and  $\omega$  is the frequency of the field mode) and the operator for the electric field reads (we do not take into account polarization of the field)

$$\hat{E}(r, t) = \sqrt{2}\mathcal{E}_0 (\hat{a}e^{-i\omega t} + \hat{a}^\dagger e^{i\omega t}) u(r), \quad (6.4)$$

where  $u(r)$  describes the spatial field distribution and is the same in both classical and quantum theories. The constant  $\mathcal{E}_0 = (\hbar\omega/2\epsilon_0 V)^{1/2}$  is equal to the “electric field per photon” in the cavity of volume  $V$ .

### 6.2.2 Wigner functions

In general, states of a quantum mechanical system are described by positive Hermitean density operators  $\hat{\rho} = \hat{\rho}^\dagger$  that act on a Hilbert space  $H$ . The density operators  $\hat{\rho}$  form a convex space  $\mathcal{S}$ . The extreme points of this state space correspond to a manifold of all one-dimensional projectors (pure states) acting on the given Hilbert space  $H$ . In the case of a harmonic oscillator one can introduce a quasi-probability density distributions in a phase space that can be associated with density operators (states) of the oscillator under consideration. In particular, the Wigner function [6, 7] can be defined as a

particular Fourier transform of the density operator  $\hat{\rho}$  of a harmonic oscillator expressed in the basis of the eigenvectors  $|q\rangle$  of the position operator  $\hat{q}$ :

$$W_{\hat{\rho}}(q, p) \equiv \int_{-\infty}^{\infty} d\zeta \langle q - \zeta/2 | \hat{\rho} | q + \zeta/2 \rangle e^{ip\zeta/\hbar}. \quad (6.5)$$

Alternatively, the Wigner function (WF) can be rewritten in the form

$$W_{\hat{\rho}}(q, p) = \frac{1}{2\pi\hbar} \int dp' dq' C_{\hat{\rho}}^{(W)}(q', p') \exp\left[-\frac{i}{\hbar}(qp' - pq')\right], \quad (6.6)$$

where the characteristic function  $C_{\hat{\rho}}^{(W)}(q, p)$  is given by the relation

$$C_{\hat{\rho}}^{(W)}(q, p) = \text{Tr}\left[\hat{\rho}\hat{D}(q, p)\right]. \quad (6.7)$$

The displacement operator  $\hat{D}(q, p)$  in terms of the position and the momentum operators reads

$$\hat{D}(q, p) = \exp\left[\frac{i}{\hbar}(\hat{q}p - \hat{p}q)\right]. \quad (6.8)$$

The Wigner function can be interpreted as the quasi-probability density distribution through which a probability can be expressed to find a quantum-mechanical system (harmonic oscillator) around the “point”  $(q, p)$  of the phase space. With the help of the Wigner function  $W_{\hat{\rho}}(q, p)$  the position and momentum probability distributions  $w_{\hat{\rho}}(q)$  and  $w_{\hat{\rho}}(p)$  can be expressed from  $W_{\hat{\rho}}(q, p)$  via marginal integration over the conjugated variable (in what follows we assume  $\lambda = 1$ )

$$w_{\hat{\rho}}(q) \equiv \frac{1}{\sqrt{2\pi\hbar}} \int dp W_{\hat{\rho}}(q, p) = \sqrt{2\pi\hbar} \langle q | \hat{\rho} | q \rangle, \quad (6.9)$$

where  $|q\rangle$  is the eigenstate of the position operator  $\hat{q}$ . The marginal probability distribution  $W_{\hat{\rho}}(q)$  is normalized to unity, i.e.,

$$\frac{1}{\sqrt{2\pi\hbar}} \int dq w_{\hat{\rho}}(q) = 1. \quad (6.10)$$

As an illustration, let us consider a Wigner function of a specific superposition of two coherent states:

$$|\alpha_e\rangle = N_e^{1/2} (|\alpha\rangle + |-\alpha\rangle); \quad N_e^{-1} = 2 [1 + \exp(-2|\alpha|^2)], \quad (6.11)$$

which is called the even coherent state [8]. The coherent state  $|\alpha\rangle$  is defined as usually,  $|\alpha\rangle = \hat{D}(\bar{q}, \bar{p})|0\rangle$  where  $|0\rangle$  is the vacuum state of the harmonic oscillator. The parameter  $\alpha = \alpha_x + i\alpha_y$  is defined via the relations  $\bar{q} =$

$\sqrt{2\hbar}\alpha_x/\lambda$  and  $\bar{p} = \sqrt{2\hbar}\alpha_y\lambda$ . The Wigner function of the coherent state  $|\alpha\rangle\langle\alpha|$  has a Gaussian form

$$W_{|\alpha\rangle}(q, p) = \frac{1}{\sigma_q\sigma_p} \exp \left[ -\frac{1}{2\hbar} \frac{(q - \bar{q})^2}{\sigma_q^2} - \frac{1}{2\hbar} \frac{(p - \bar{p})^2}{\sigma_p^2} \right], \quad (6.12)$$

where

$$\sigma_q^2 = \frac{1}{2\lambda^2} \quad \text{and} \quad \sigma_p^2 = \frac{\lambda^2}{2}. \quad (6.13)$$

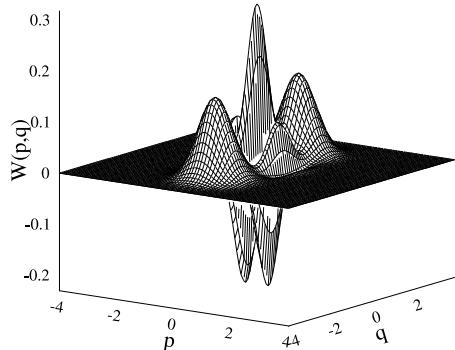
If we assume  $\alpha$  to be real, then the Wigner function of the even coherent state reads

$$W_{|\alpha_e\rangle}(q, p) = N_e [W_{|\alpha\rangle}(q, p) + W_{|-\alpha\rangle}(q, p) + W_{int}(q, p)]; \quad (6.14)$$

where  $W_{|\pm\alpha\rangle}(q, p)$  is the WF of coherent states  $|\pm\alpha\rangle$ . The interference part of the Wigner function (6.14) is given by the relation

$$W_{int}(q, p) = \frac{2}{\sigma_q\sigma_p} \exp \left[ -\frac{q^2}{2\hbar\sigma_q^2} - \frac{p^2}{2\hbar\sigma_p^2} \right] \cos \left( \frac{\bar{q}p}{\hbar\sigma_q\sigma_p} \right), \quad (6.15)$$

We plot the Wigner function of the even coherent state in Fig. 6.1. From the figure it is clearly seen that the interference term (6.15) results in oscillations of the Wigner function around the origin of the phase space.



**Fig. 6.1.** The Wigner function of the even coherent state with  $\alpha = 2$ .

### 6.2.3 Quantum homodyne tomography

The relation (6.9) for the probability distribution  $w_{\hat{p}}(q)$  of the position operator  $\hat{q}$  can be generalized to the case of the distribution of the rotated quadrature operator  $\hat{x}_\theta$ . This operator is defined as

$$\hat{x}_\theta = \sqrt{\frac{\hbar}{2}} [\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta}], \quad (6.16)$$

and the corresponding conjugated operator  $\hat{x}_{\theta+\pi/2}$ , such that  $[\hat{x}_\theta, \hat{x}_{\theta+\pi/2}] = i\hbar$ , reads

$$\hat{x}_{\theta+\pi/2} = \frac{\sqrt{\hbar}}{i\sqrt{2}} [\hat{a}e^{-i\theta} - \hat{a}^\dagger e^{i\theta}]. \quad (6.17)$$

The position and the momentum operators are related to the operator  $\hat{x}_\theta$  as,  $\hat{q} = \hat{x}_0$  and  $\hat{x}_{\pi/2} = \hat{p}$ . The rotation (i.e., the linear homogeneous canonical transformation) given by Eqs.(6.16) and (6.17) can be performed by the unitary operator  $\hat{U}(\theta)$ :

$$\hat{U}(\theta) = \exp [-i\theta \hat{a}^\dagger \hat{a}], \quad (6.18)$$

so that

$$\hat{x}_\theta = \hat{U}^\dagger(\theta) \hat{x}_0 \hat{U}(\theta); \quad \hat{x}_{\theta+\pi/2} = \hat{U}^\dagger(\theta) \hat{x}_{\pi/2} \hat{U}(\theta). \quad (6.19)$$

Alternatively, in the vector formalism we can rewrite the transformation (6.19) as

$$\begin{pmatrix} \hat{x}_\theta \\ \hat{x}_{\theta+\pi/2} \end{pmatrix} = \mathbf{F} \begin{pmatrix} \hat{q} \\ \hat{p} \end{pmatrix}; \quad \mathbf{F} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (6.20)$$

Eigenvalues  $x_\theta$  and  $x_{\theta+\pi/2}$  of the operators  $\hat{x}_\theta$  and  $\hat{x}_{\theta+\pi/2}$  can be expressed in terms of the eigenvalues  $q$  and  $p$  of the position and momentum operators as:

$$\begin{pmatrix} x_\theta \\ x_{\theta+\pi/2} \end{pmatrix} = \mathbf{F} \begin{pmatrix} q \\ p \end{pmatrix}; \quad \begin{pmatrix} q \\ p \end{pmatrix} = \mathbf{F}^{-1} \begin{pmatrix} x_\theta \\ x_{\theta+\pi/2} \end{pmatrix};$$

$$\mathbf{F}^{-1} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad (6.21)$$

where the matrix  $\mathbf{F}$  is given by Eq.(6.20) and  $\mathbf{F}^{-1}$  is the corresponding inverse matrix. It has been shown by Ekert and Knight [9] that Wigner functions are transformed under the action of the linear canonical transformation (6.20) as:

$$\begin{aligned} W_{\hat{\rho}}(q, p) &\rightarrow W_{\hat{\rho}}(\mathbf{F}^{-1}(x_\theta, x_{\theta+\pi/2})) \\ &= W_{\hat{\rho}}(x_\theta \cos \theta - x_{\theta+\pi/2} \sin \theta; x_\theta \sin \theta + x_{\theta+\pi/2} \cos \theta), \end{aligned} \quad (6.22)$$

which means that the probability distribution  $w_{\hat{\rho}}(x_\theta, \theta) = \sqrt{2\pi\hbar} \langle x_\theta | \hat{\rho} | x_\theta \rangle$  can be evaluated as

$$\begin{aligned} w_{\hat{\rho}}(x_\theta, \theta) &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dx_{\theta+\pi/2} \\ &\times W_{\hat{\rho}}(x_\theta \cos \theta - x_{\theta+\pi/2} \sin \theta; x_\theta \sin \theta + x_{\theta+\pi/2} \cos \theta). \end{aligned} \quad (6.23)$$

As shown by Vogel and Risken [10] (see also [11–14]) the knowledge of  $w_{\hat{\rho}}(x_\theta, \theta)$  for all values of  $\theta$  (such that  $[0 < \theta \leq \pi]$ ) is equivalent to the knowledge of the Wigner function itself. This Wigner function can be obtained from the set of distributions  $w_{\hat{\rho}}(x_\theta, \theta)$  via the inverse Radon transformation:

$$\begin{aligned} W_{\hat{\rho}}(q, p) = & \frac{1}{(2\pi\hbar)^{3/2}} \int_{-\infty}^{\infty} dx_\theta \int_{-\infty}^{\infty} d\xi |\xi| \\ & \times \int_0^\pi d\theta w_{\hat{\rho}}(x_\theta, \theta) \exp \left[ \frac{i}{\hbar} \xi(x_\theta - q \cos \theta - p \sin \theta) \right]. \end{aligned} \quad (6.24)$$

We stress that the transformation (6.24) is a “deterministic” inverse transformation with the help of which the complete knowledge about the state encoded in the marginal distributions  $w_{\hat{\rho}}(x_\theta, \theta)$  is rewritten in a form of the Wigner function.

This reconstruction scheme has been used by Raymer and his coworkers [15, 16]; see also chapter by Raymer and Beck of this volume for more details. In their experiments the Wigner functions of a coherent state and a squeezed vacuum state have been reconstructed from tomographic data.

Quantum-state tomography can be applied not only to optical fields but also for reconstruction of other physical systems. In particular, Janicke and Wilkens [17] have suggested that Wigner functions of atomic waves can be tomographically reconstructed. Kurtsiefer et al. [18] have performed experiments in which Wigner functions of matter wave packets have been reconstructed. Yet another example of the tomographic reconstruction is a reconstruction of Wigner functions of vibrational states of trapped atomic ions theoretically described by a number of groups [19] and experimentally measured by Leibfried et al. [20]. Vibrational motional states of molecules have also been reconstructed by this kind of quantum tomography by Dunn et al. [21].

The problem with the inverse Radon transformation is that it does not take into account the possibility of a finite number of measured distributions. As we will show later, in the case of incomplete tomographic data the transformation (6.24) can lead to unphysical reconstructions (e.g. non-positive density operators). In what follows we briefly review a quantum tomography scheme that is based on the sampling via the pattern functions which is equivalent to the inverse Radon transformation.

#### 6.2.4 Quantum tomography via pattern functions

In a sequence of papers D’Ariano et al. [13], Leonhardt et al. [22] and Richter [23] have shown that Wigner functions can be very efficiently reconstructed from tomographic data with the help of the so-called pattern functions. This reconstruction procedure is more efficient than the usual Radon transformation [24]. To be specific, D’Ariano et al. [13] and Kühn et al. [14] have shown that the density matrix  $\rho_{mn}$  in the Fock basis can be reconstructed directly

from the tomographic data, i.e. from the quadrature-amplitude “histograms” (probabilities),  $w(x_\theta, \theta)$  via the so-called *direct sampling method* when

$$\rho_{mn} = \int_0^\pi \int_{-\infty}^\infty w(x_\theta, \theta) F_{mn}(x_\theta, \theta) dx_\theta d\theta, \quad (6.25)$$

where  $F_{mn}(x_\theta, \theta)$  is a set of specific *sampling* functions (see below). Once the density matrix elements are reconstructed with the help of Eq.(6.25) then the Wigner function of the corresponding state can be directly obtained using the relation

$$W_{\hat{\rho}}(q, p) = \sum_{m,n} \rho_{mn} W_{|m\rangle\langle n|}(q, p), \quad (6.26)$$

where  $W_{|m\rangle\langle n|}(q, p)$  is the Wigner function of the operator  $|m\rangle\langle n|$ .

A serious problem with the direct sampling method as proposed by D’Ariano et al. [13] is that the sampling functions  $F_{mn}(x_\theta, \theta)$  are difficult to compute. Later D’Ariano, Leonhardt and Paul [22, 25] have simplified the expression for the sampling function and have found that it can be expressed as

$$F_{mn}(x_\theta, \theta) = f_{mn}(x_\theta) \exp [i(m-n)\theta], \quad (6.27)$$

where the so-called *pattern* function “picks up” the pattern in the quadrature histograms (probability distributions)  $w_{mn}(x_\theta, \theta)$  that just match the corresponding density-matrix elements. Leonhardt et al. [24] have shown that the pattern function  $f_{mn}(x_\theta)$  can be expressed as derivatives

$$f_{mn}(x) = \frac{\partial}{\partial x} g_{mn}(x), \quad (6.28)$$

of functions  $g_{mn}(x)$  that are given by the Hilbert transformation

$$g_{mn}(x) = \frac{\mathcal{P}}{\pi} \int_{-\infty}^{\infty} \frac{\psi_m(\zeta) \psi_n(\zeta)}{x - \zeta} d\zeta, \quad (6.29)$$

where  $\mathcal{P}$  stands for the principal value of the integral and  $\psi_n(x)$  are the real energy eigenfunctions of the harmonic oscillator, i.e. the normalizable solutions of the Schrödinger equation

$$\left( -\frac{\hbar^2}{2} \frac{d^2}{dx^2} + \frac{x^2}{2} \right) \psi_n(x) = \hbar(n + 1/2) \psi_n(x), \quad (6.30)$$

(we assume  $m = \omega = 1$ ). Further details of possible applications and discussion devoted to numerical procedures of the reconstruction of density operators via the direct sampling method can be found in Ref. [24].

### 6.3 Maxent principle and observation levels

The state of a quantum system can always be described by a statistical density operator  $\hat{\rho}$ . Depending on the system preparation, the density operator represents either a pure quantum state (complete system preparation) or a statistical mixture of pure states (incomplete preparation). The degree of deviation of a statistical mixture from the pure state can be best described by the *uncertainty measure*  $\eta[\hat{\rho}]$  (see [26, 27])

$$\eta[\hat{\rho}] = -\text{Tr}(\hat{\rho} \ln \hat{\rho}). \quad (6.31)$$

The uncertainty measure  $\eta[\hat{\rho}]$  possesses the following properties:

**1.** In the eigenrepresentation of the density operator  $\hat{\rho}$

$$\hat{\rho} |r_m\rangle = r_m |r_m\rangle, \quad (6.32)$$

we can write

$$\eta[\hat{\rho}] = -\sum_m r_m \ln r_m \geq 0, \quad (6.33)$$

where  $r_m$  are eigenvalues and  $|r_m\rangle$  the eigenstates of  $\hat{\rho}$ .

**2.** For uncertainty measure  $\eta[\hat{\rho}]$  the following inequality holds:

$$0 \leq \eta[\hat{\rho}] \leq \ln N, \quad (6.34)$$

where  $N$  denotes the dimension of the Hilbert space of the system and  $\eta[\hat{\rho}]$  takes its maximum value when

$$\hat{\rho} = \frac{\hat{1}}{\text{Tr}\hat{1}} = \frac{\hat{1}}{N}. \quad (6.35)$$

In this case all pure states in the mixture appear with the same probability equal to  $1/N$ . If the system is prepared in a pure state then it holds that  $\eta[\hat{\rho}] = 0$ .

**3.** It can be shown with the help of the Liouville equation

$$\frac{\partial}{\partial t} \hat{\rho}(t) = -\frac{i}{\hbar} [\hat{H}, \hat{\rho}(t)], \quad (6.36)$$

that in the case of an isolated system the uncertainty measure is a constant of motion, i.e.,

$$\frac{d\eta(t)}{dt} = 0. \quad (6.37)$$

### 6.3.1 MaxEnt principle

When instead of the density operator  $\hat{\rho}$ , expectation values  $G_\nu$  of a set  $\mathcal{O}$  of operators  $\hat{G}_\nu$  ( $\nu = 1, \dots, n$ ) are given, then the uncertainty measure can be determined as well. The set of linearly independent operators is referred to as the *observation level*  $\mathcal{O}$  [28, 29]. The operators  $\hat{G}_\nu$  that belong to a given observation level do not commute necessarily. A large number of density operators that fulfill the conditions

$$\begin{aligned} \text{Tr } \hat{\rho}_{\{\hat{G}\}} &= 1, \\ \text{Tr } (\hat{\rho}_{\{\hat{G}\}} \hat{G}_\nu) &= G_\nu, \quad \nu = 1, 2, \dots, n; \end{aligned} \quad (6.38)$$

can be found for a given set of expectation values  $G_\nu = \langle \hat{G}_\nu \rangle$ . That is, the conditions (6.38) specify a set  $\mathcal{C}$  of density operators, which has to be considered. Each of these density operators  $\hat{\rho}_{\{\hat{G}\}}$  can possess a different value of the uncertainty measure  $\eta[\hat{\rho}_{\{\hat{G}\}}]$ . If we wish to use only the expectation values  $G_\nu$  of the chosen observation level for determining the density operator, we must select a particular density operator  $\hat{\rho}_{\{\hat{G}\}} = \hat{\sigma}_{\{\hat{G}\}}$  in an unbiased manner. According to the Jaynes principle of the Maximum Entropy [28–33] this density operator  $\hat{\sigma}_{\{\hat{G}\}}$  must be the one that has the largest uncertainty measure

$$\eta_{\max} \equiv \max \left\{ \eta[\hat{\sigma}_{\{\hat{G}\}}] \right\} \quad (6.39)$$

and simultaneously fulfills constraints (6.38). As a consequence of Eq.(6.39) the following fundamental inequality holds

$$\eta[\hat{\sigma}_{\{\hat{G}\}}] = -\text{Tr}(\hat{\sigma}_{\{\hat{G}\}} \ln \hat{\sigma}_{\{\hat{G}\}}) \geq \eta[\hat{\rho}_{\{\hat{G}\}}] = -\text{Tr}(\hat{\rho}_{\{\hat{G}\}} \ln \hat{\rho}_{\{\hat{G}\}}) \quad (6.40)$$

for all possible  $\hat{\rho}_{\{\hat{G}\}}$  that fulfill Eqs.(6.38). The variation determining the maximum of  $\eta[\hat{\sigma}_{\{\hat{G}\}}]$  under the conditions (6.38) leads to a generalized canonical density operator [28–30, 32]

$$\hat{\sigma}_{\{\hat{G}\}} = \frac{1}{Z_{\{\hat{G}\}}} \exp \left( -\sum_\nu \lambda_\nu \hat{G}_\nu \right); \quad (6.41)$$

$$Z_{\{\hat{G}\}}(\lambda_1, \dots, \lambda_n) = \text{Tr} \left[ \exp \left( -\sum_\nu \lambda_\nu \hat{G}_\nu \right) \right], \quad (6.42)$$

where  $\lambda_n$  are the Lagrange multipliers and  $Z_{\{\hat{G}\}}(\lambda_1, \dots, \lambda_n)$  is the generalized partition function. By using the derivatives of the partition function we obtain the expectation values  $G_\nu$  as

$$G_\nu = \text{Tr}(\hat{\sigma}_{\{\hat{G}\}} \hat{G}_\nu) = -\frac{\partial}{\partial \lambda_\nu} \ln Z_{\{\hat{G}\}}(\lambda_1, \dots, \lambda_n), \quad (6.43)$$

where in the case of noncommuting operators the following relation has to be used

$$\frac{\partial}{\partial a} \exp[-\hat{X}(a)] = \exp[-\hat{X}(a)] \int_0^1 \exp[\mu \hat{X}(a)] \frac{\partial \hat{X}(a)}{\partial a} \exp[-\mu \hat{X}(a)] d\mu \quad (6.44)$$

By using Eq.(6.43) the Lagrange multipliers can, in principle, be expressed as functions of the expectation values

$$\lambda_\nu = \lambda_\nu(G_1, \dots, G_n). \quad (6.45)$$

We note that Eqs.(6.43) for Lagrange multipliers do not always have solutions which lead to physical results, which means that in these cases states of quantum systems cannot be reconstructed on a given observation level.

The maximum uncertainty measure regarding an observation level  $\mathcal{O}_{\{\hat{G}\}}$  will be referred to as the entropy  $S_{\{\hat{G}\}}$

$$S_{\{\hat{G}\}} \equiv \eta_{\max} = -\text{Tr}(\hat{\sigma}_{\{\hat{G}\}} \ln \hat{\sigma}_{\{\hat{G}\}}). \quad (6.46)$$

This means that to different observation levels different entropies are related. By inserting  $\sigma_{\{\hat{G}\}}$  [cf. Eq.(6.41)] into Eq.(6.46), we obtain the following expression for the entropy

$$S_{\{\hat{G}\}} = \ln Z_{\{\hat{G}\}} + \sum_\nu \lambda_\nu G_\nu. \quad (6.47)$$

By making use of Eq.(6.45), the parameters  $\lambda_\nu$  in the above equation can be expressed as functions of the expectation values  $G_\nu$ , and this leads to a new expression for the entropy

$$S_{\{\hat{G}\}} = S(G_1, \dots, G_n). \quad (6.48)$$

We note that using the expression

$$dS_{\{\hat{G}\}} = \sum_\nu \lambda_\nu dG_\nu, \quad (6.49)$$

which follows from Eqs.(6.43) and (6.47) the following relation can be obtained

$$\lambda_\nu = \frac{\partial}{\partial G_\nu} S(G_1, \dots, G_n). \quad (6.50)$$

### 6.3.2 Linear transformations within an observation level

An observation level can be defined either by a set of linearly independent operators  $\{\hat{G}_\nu\}$ , or by a set of independent linear combinations of the same operators

$$\hat{G}'_\mu = \sum_\nu c_{\mu\nu} \hat{G}'_\nu. \quad (6.51)$$

Therefore,  $\hat{\sigma}$  and  $S$  are invariant under a linear transformation:

$$\hat{\sigma}'_{\{\hat{G}'\}} = \frac{\exp(-\sum_\mu \lambda'_\mu \hat{G}'_\mu)}{\text{Tr} \exp(-\sum_\mu \lambda'_\mu \hat{G}'_\mu)} = \hat{\sigma}_{\{\hat{G}\}}. \quad (6.52)$$

As a result, the Lagrange multipliers transform contravariantly to Eq.(6.51), i.e.,

$$\lambda'_\mu = \sum_\nu c'_{\mu\nu} \lambda_\nu, \quad (6.53)$$

$$\sum_\mu c'_{\nu\mu} c_{\mu\rho} = \delta_{\nu\rho}. \quad (6.54)$$

### 6.3.3 Extension and reduction of the observation level

If an observation level  $\mathcal{O}_{\{\hat{G}\}} \equiv \hat{G}_1, \dots, \hat{G}_n$  is extended by including further operators  $\hat{M}_1, \dots, \hat{M}_l$ , then additional expectation values  $M_1 = \langle \hat{M}_1 \rangle, \dots, M_l = \langle \hat{M}_l \rangle$  can only increase the amount of available information about the state of the system. This procedure is called the *extension* of the observation level (from  $\mathcal{O}_{\{\hat{G}\}}$  to  $\mathcal{O}_{\{\hat{G}, \hat{M}\}}$ ) and is associated with a decrease of the entropy. More precisely, the entropy  $S_{\{\hat{G}, \hat{M}\}}$  of the extended observation level  $\mathcal{O}_{\{\hat{G}, \hat{M}\}}$  can be only smaller or equal to the entropy  $S_{\{\hat{G}\}}$  of the original observation level  $\mathcal{O}_{\{\hat{G}\}}$ ,

$$S_{\{\hat{G}, \hat{M}\}} \leq S_{\{\hat{G}\}}. \quad (6.55)$$

The generalized canonical density operator of the observation level  $\mathcal{O}_{\{\hat{G}, \hat{M}\}}$

$$\hat{\sigma}_{\{\hat{G}, \hat{M}\}} = \frac{1}{Z_{\{\hat{G}, \hat{M}\}}} \exp \left( - \sum_{\nu=1}^n \lambda_\nu \hat{G}_\nu - \sum_{\mu=1}^l \kappa_\mu \hat{M}_\mu \right), \quad (6.56)$$

with

$$Z_{\{\hat{G}, \hat{M}\}} = \text{Tr} \left[ \exp \left( - \sum_{\nu=1}^n \lambda_\nu \hat{G}_\nu - \sum_{\mu=1}^l \kappa_\mu \hat{M}_\mu \right) \right], \quad (6.57)$$

belongs to the set of density operators  $\hat{\rho}_{\{\hat{G}\}}$  fulfilling Eq.(6.38). Therefore, Eq.(6.56) is a special case of Eq.(6.41). Analogously to Eqs.(6.43) and (6.45), the Lagrange multipliers can be expressed by functions of the expectation values

$$\begin{aligned}\lambda_\nu &= \lambda_\nu(G_1, \dots, G_n, M_1, \dots, M_l), \\ \kappa_\mu &= \kappa_\mu(G_1, \dots, G_n, M_1, \dots, M_l).\end{aligned}\quad (6.58)$$

The sign of equality in Eq.(6.55) holds only for  $\kappa_\mu = 0$ . In this special case the expectation values  $M_\mu$  are functions of the expectation values  $G_\nu$ . The measurement of observables  $\hat{M}_\mu$  does not increase information about the system. Consequently,  $\hat{\rho}_{\{\hat{G}, \hat{M}\}} = \hat{\rho}_{\{\hat{G}\}}$  and  $S_{\{\hat{G}, \hat{M}\}} = S_{\{\hat{G}\}}$ .

We can also consider a *reduction of the observation level* if we decrease the number of independent observables that are measured, e.g.,  $\mathcal{O}_{\{\hat{G}, \hat{M}\}} \rightarrow \mathcal{O}_{\{\hat{G}\}}$ . This reduction is accompanied by an increase of the entropy due to the decrease of the information available about the system.

### 6.3.4 Wigner functions on different observation levels

With the help of a generalized canonical density operator  $\hat{\sigma}_{\{\hat{G}\}}$  we define the Wigner function in the  $\xi$  phase space at the corresponding observation level as

$$W_{\{\hat{G}\}}(\xi) = \frac{1}{\pi} \int d^2\eta \text{Tr} [\hat{D}(\eta) \hat{\sigma}_{\{\hat{G}\}}] \exp(\xi\eta^* - \xi^*\eta). \quad (6.59)$$

Analogous expression can be found for the Wigner function in the  $(q, p)$  phase space.

### 6.3.5 MaxEnt principle and laws of physics

It has been pointed out by Jaynes in his Brandeis lectures [1] that there is a strong formal resemblance between the MaxEnt formalism and the rules of calculations in statistical mechanics and thermodynamics. Simultaneously he has emphasized that the MaxEnt principle “has nothing to do with the laws of physics”<sup>3</sup>. To be more specific it is worth citing a paragraph from the Jaynes’ Brandeis lectures (see p. 183 of these lectures [1]): “Conventional quantum theory has provided an answer to the problem of setting up initial state descriptions only in the limiting case where measurements of a “complete set of commuting observables” have been made, the density matrix  $\hat{\rho}(0)$  then reducing to the projection operator onto a pure state  $\psi(0)$  which is the appropriate simultaneous eigenstate of all measured quantities. But there is almost no experimental situation in which we really have all this information, and before we have a theory able to treat actual experimental situations, existing quantum theory must be supplemented with some principle that tells us how to translate, or encode, the results of measurements into a definite state description  $\hat{\rho}(0)$ . Note that the problem is not to find  $\hat{\rho}(0)$  which correctly

---

<sup>3</sup> In fact, this is the reason why the MaxEnt principle is applicable in so many fields of human activities, for instance we can mention economy or sociology (for more details see the book by Kapur and Kevasan [31]).

describes “true physical situation”. That is unknown, and always remains so, because of incomplete information. In order to have a usable theory we must ask the much more modest question: *What  $\hat{\rho}(0)$  best describes our state of knowledge about the physical situation?*”.

In other words, the MaxEnt principle is *the most conservative assignment in the sense that it does not permit one to draw any conclusions not warranted by the data*. From this point of view the MaxEnt principle has a very close relation (or can be understood as the generalization) of the Laplace’s principle of *indifference* [34] which states that where nothing is known one should choose a constant valued function to reflect this ignorance. Then it is just a question of how to quantify a degree of this ignorance. If we choose an entropy to quantify the ignorance, then the relation between the Laplace’s indifference principle and the Jaynes principle of the Maximum Entropy is transparent, i.e. for a constant-valued probability distribution the entropy takes its maximum value.

We can conclude that a measurement itself is a physical process and is governed by the laws of physics. On the other hand formal procedures by means of which we reconstruct information about the system from the measured data are based on certain principles that cannot be directly expressed in terms of the physical laws.

### 6.3.6 Quantum tomography via MaxEnt principle

The probability density distribution  $w_{\hat{\rho}}(x_\theta)$  [see Eq.(6.23)] for rotated quadratures  $\hat{x}_\theta$  can be represented as a result of the measurement of the continuous set of projectors  $|x_\theta\rangle\langle x_\theta|$ . Based on the measurement of the distributions  $w_{\hat{\rho}}(x_\theta)$  for all values of  $\theta \in [0, \pi]$  we can formally “reconstruct” the density operator according to the formula

$$\hat{\rho}_{ME} = \frac{1}{Z_0} \exp \left[ - \int_0^\pi d\theta \int_{-\infty}^\infty dx_\theta |x_\theta\rangle\langle x_\theta| \lambda(x_\theta) \right], \quad (6.60)$$

where the Lagrange multipliers  $\lambda(x_\theta)$  are given by an infinite set of equations

$$w_{\hat{\rho}}(x_\theta) = \sqrt{2\pi\hbar} \langle x_\theta | \hat{\rho}_{ME} | x_\theta \rangle; \quad \forall x_\theta \in (-\infty, \infty). \quad (6.61)$$

If the distributions  $w_{\hat{\rho}}(x_\theta)$  are measured for all values of  $x_\theta$  and all angles  $\theta$  then the density operator  $\hat{\rho}_{ME}$  is reconstructed precisely and is equal to the density operator obtained with the help of the inverse Radon transformation (or with the help of the pattern functions).

In practical experimental situations (see the experiments by Raymer et al. [15] and by Mlynek et al. [18]) it is impossible to measure the distributions  $w_{\hat{\rho}}(x_\theta)$  for all values of  $x_\theta$  and all angles  $\theta$ . What is measured are distributions (histograms) for finite number  $N_\theta$  quadrature angles  $\theta$  and the finite number  $N_x$  of bins for quadrature operators. This means that practical experiments are associated with an observation level specified by a finite number of observables

$$\hat{Q}_{nm} = |x_{\theta_m}^{(n)}\rangle\langle x_{\theta_m}^{(n)}| \quad (6.62)$$

with number of quadrature angles equal to  $N_\theta$  and the number of bins for each quadrature equal to  $N_x$ . These observables in the Fock basis can be represented as

$$\left(\hat{Q}_{nm}\right)_{k_1, k_2} = \psi_{k_1}^*(x_n)\psi_{k_2}(x_n)\exp[i\theta_m(k_1 - k_2)], \quad (6.63)$$

where  $\theta_m$  is the quadrature phase,  $x_n$  is eigenvalue of the quadrature operator and  $\psi_k(x)$  is the wave function of the  $k$ th energy eigenstate (Fock state) of the harmonic oscillator. We can therefore assume that from the measurement of the observables  $\hat{Q}_{nm}$  the mean values  $\bar{Q}_{nm}$  are determined (these mean values correspond to “discretized” quadrature distributions). In addition, it is usually the case that the mean photon number of the state is known (measured) as well.

The operators  $\hat{Q}_{nm}$  together with  $\hat{n}$  form a specific observation level corresponding to the incomplete tomographic measurement. In this case we can express the generalized canonical density operator in the form

$$\hat{\rho}_{ME} = \frac{1}{Z} \exp \left( -\lambda_0 \hat{n} - \sum_{n=1}^{N_x} \sum_{m=1}^{N_\theta} \lambda_{n,m} |x_{\theta_m}^{(n)}\rangle\langle x_{\theta_m}^{(n)}| \right) \quad (6.64)$$

The knowledge of the mean photon number is essential for the *MaxEnt* reconstruction because it formally regularizes the *MaxEnt* reconstruction scheme (the generalized partition function is finite in this case).

## 6.4 Numerical implementation

Let us summarize what is supposed to be known as a result of the measurement - these are the measured mean values  $\bar{Q}_{nm}$  and  $\bar{n}$  of the observables  $\hat{Q}_{nm}$  and  $\hat{n}$ , respectively. Further, the experimental setup gives us the numbers  $N_\theta$  and  $N_x$  as well as the size  $\Delta x$  of quadrature bins. These last two numbers specify the range of measured quadratures  $-N_x\Delta x/2 \leq x \leq N_x\Delta x/2$ .

In addition to these “experimental” parameters we have to specify also the dimensionality  $N_{max}$  of the Hilbert space in which we reconstruct the density operator. In the case of the *MaxEnt* reconstruction  $N_{max}$  has to be chosen so that the “truncation” of the Hilbert space does not affect the reconstruction of the state of original light field (i.e.  $N_{max} \gg \bar{n}$  so that the reconstructed state “fits” into the truncated Hilbert space).

To perform the reconstruction we have to determine the Lagrange multipliers  $\lambda_{n,m}$  in the expression for the generalized canonical density operator (6.64). These multipliers are given by the constraints (6.38) and numerically can be determined via the minimization of a deviation function  $\Delta Q$  with

respect to the measured mean photon number  $\bar{n}$  and the set of histograms  $\bar{Q}_{nm} = \text{Tr}\{\hat{\rho}\hat{Q}_{nm}\}$ :

$$\Delta Q = (\bar{n} - \text{Tr}\{\hat{\rho}_{ME}\hat{n}\})^2 + \sum_{n,m=1}^{N_\theta, N_x} \left( \bar{Q}_{nm} - \text{Tr}\{\hat{\rho}_{ME}\hat{Q}_{nm}\} \right)^2. \quad (6.65)$$

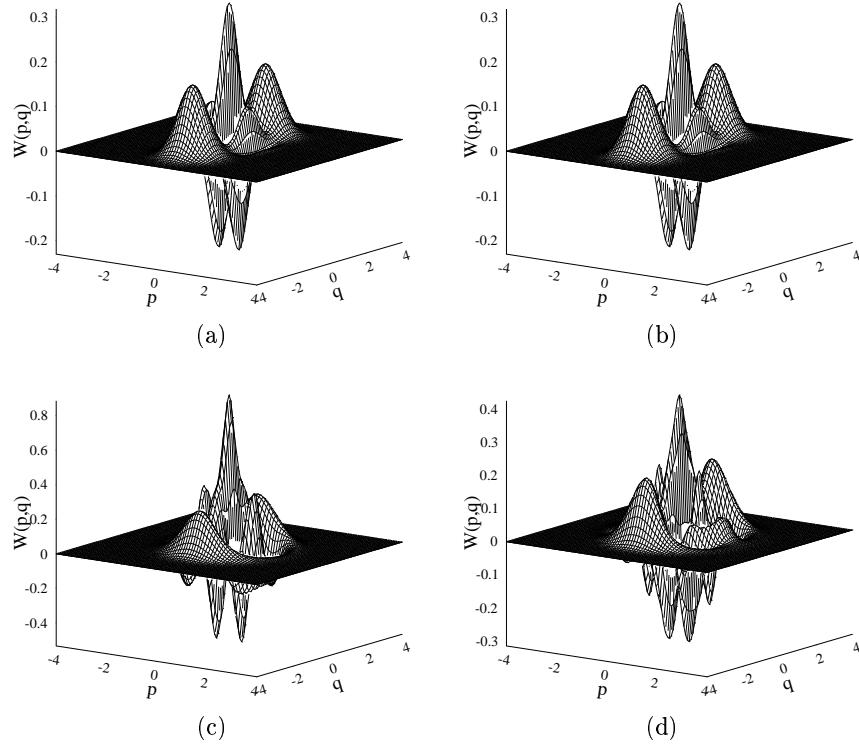
The trace is performed within the truncated Hilbert space specified by the parameter  $N_{max}$ . When  $\Delta Q = 0$  the Lagrange multipliers are determined precisely and the reconstructed density operator  $\hat{\rho}_{ME}$  *ideally* satisfies the mean values of measured observables.

In order to find the minimal value of the function  $\Delta Q$  and to determine the Lagrange multipliers we utilize the Levenberg-Marquardt algorithm with a finite difference Jacobian (see a standard routine from the IMSL library, Visual Numerics, Inc., <http://www.vni.com>).

Once the Lagrange multipliers are specified, then, using the expression for the generalized canonical density operator (6.64), we can plot the corresponding Wigner function. The fidelity of the reconstruction is given by three parameters. Firstly, it is the minimal value of the function  $\Delta Q$  that determines the deviation between the measured mean values of the observables and the corresponding mean values evaluated from the reconstructed density operator. Secondly, if it is a priori known that the measured system is prepared in a pure state then the von Neumann entropy  $S$  of the density operator  $\hat{\rho}_{ME}$  is a measure of the fidelity of the reconstruction. Specifically, if the entropy is equal to zero then the pure state is perfectly reconstructed. Thirdly, if we want to test the reconstruction scheme we can compare the reconstructed density operator with the known original  $\hat{\rho}$ . In this case we can use the measure

$$\Delta\rho = \sum_{m,n}^{N_x, N_\theta} |(\hat{\rho})_{mn} - (\hat{\rho}_{ME})_{mn}|^2. \quad (6.66)$$

Let us test the *MaxEnt* reconstruction scheme and assume the mean values of the observables  $\hat{Q}_{mn}$  to be given by the even coherent state (6.11) with the real amplitude  $\alpha = 2$  (we plot the Wigner function of this state in Fig. 6.2a). Let us further assume just two quadrature angles  $N_\theta = 2$  corresponding to the measurement of the position and the momentum of the harmonic oscillator. The total number of bins for each quadrature is taken to be  $N_x = 40$  with the size of the bin equal to  $\Delta x = 0.2$  (we assume  $\hbar = 1$ ), which corresponds to the measurement of the quadrature distributions on the interval  $\langle -4, 4 \rangle$ . For the given mean photon number ( $\bar{n} \simeq 4$ ) it is enough to consider the Hilbert space of the dimension  $N_{max} = 20$ . The even coherent state with  $\alpha = 2$  can be very well approximated as a superposition of even number states up to  $n = 8$  (the higher number states are occupied with 1% probability), so  $N_{max} = 20$  is a very good truncation.



**Fig. 6.2.** (a) The Wigner function of the even coherent state with  $\alpha = 2$ . (b) Reconstruction of the Wigner function via the *MaxEnt* principle from two marginal distributions ( $N_\theta = 2$ ) for the position and momentum. The measured marginal distributions are divided into  $N_x = 40$  bins of the width  $\Delta x = 0.2$  covering the interval  $\langle -4, 4 \rangle$ . The fidelity of the reconstruction is  $\Delta Q = 5.4 \times 10^{-12}$  and  $\Delta \rho = 1.0 \times 10^{-8}$ , the corresponding entropy is  $S \approx 10^{-6}$ . (c) Optical tomography via the direct sampling using pattern functions with  $N_\theta = 2$  and with an artificial truncation at  $N_{max} = 4$ . This value of  $N_{max}$  is chosen such that the deviation  $\Delta Q$  is minimized. In this case  $\Delta Q \simeq 9.17$  and  $\Delta \rho \simeq 3.7$ . The reconstructed Wigner function is unphysical because the corresponding density operator has negative probabilities for odd Fock states ( $P_1 \approx -0.35$ ,  $P_3 \approx -0.54$ ). (d) The result of the tomography can be improved when the number of quadratures is larger. For  $N_\theta = 4$  and the truncation at  $N_{max} = 6$  we can improve the fidelity of the reconstruction ( $\Delta Q \simeq 1.33$  and  $\Delta \rho = 0.51$ ) but it is still unphysical because  $P_1 \approx -0.1$ .

With these values of the parameters we have performed the reconstruction of the state via the *MaxEnt* principle. Using the minimization procedure we have achieved the deviation with respect to the “experimental” data  $\Delta Q = 5.4 \times 10^{-12}$ . The difference between the reconstructed density operator and the original measured in terms of (6.66) in this particular case is  $\Delta\rho = 1.0 \times 10^{-10}$ . We see that the reconstruction is indeed very precise even for a very small number of experimental data. A high quality of the reconstruction is indicated by the corresponding value of the von Neumann entropy  $S \approx 10^{-6}$ . We plot the reconstructed Wigner function in Fig. 6.2b. From this figure we see that the reconstruction and the original are essentially identical. We also note that the quality of the reconstruction practically does not depend on the choice of  $N_{max}$  when this is larger than some minimal value related to  $\bar{n}$  (in our case for  $N_{max} > 12$  the reconstruction is almost perfect but even for  $N_{max} = 8$  the fidelity is very high).

In order to illustrate the *MaxEnt* reconstruction scheme for statistical mixtures we will consider briefly a statistical mixture described by the density operator

$$\hat{\rho} = \frac{1}{2}|\alpha\rangle\langle\alpha| + \frac{1}{2}|-\alpha\rangle\langle-\alpha| \quad (6.67)$$

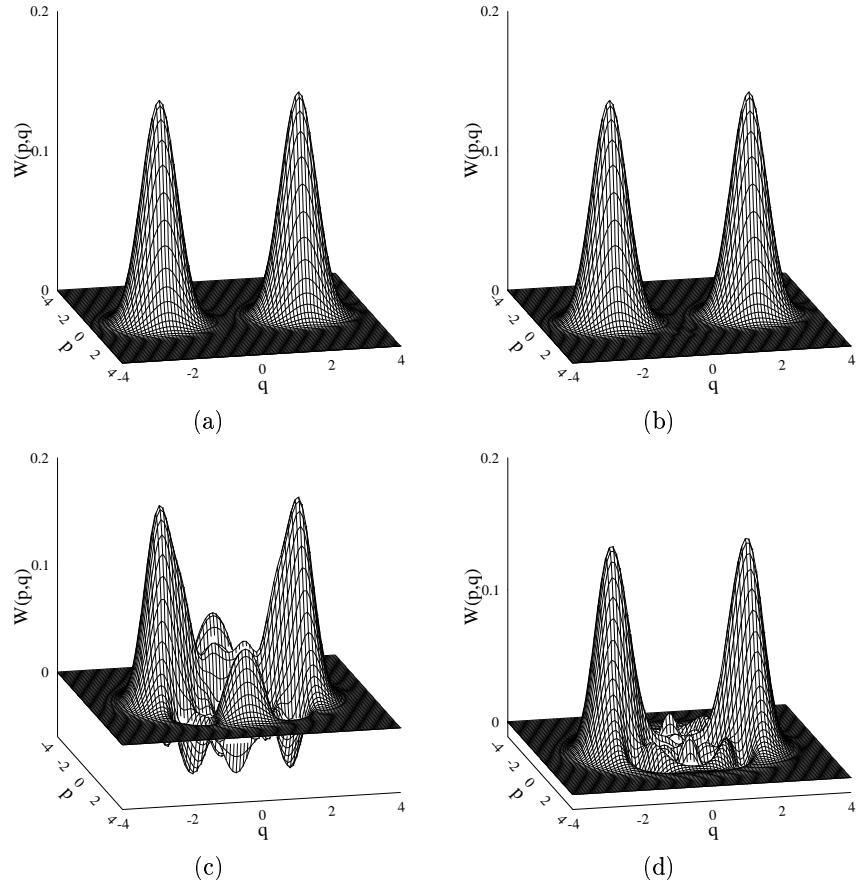
with the real amplitude  $\alpha = 2$ . The Wigner function of this state is plotted in Fig. 6.3a. In the figure Fig. 6.3b we plot the reconstructed Wigner function under the same conditions as the Fig. 6.2b, i.e.  $N_\theta = 2$ ,  $N_x = 40$ ,  $\Delta x = 0.2$  and  $N_{max} = 20$ . The two measured quadratures are the position and the momentum. We see from the figure that the reconstruction is almost perfect. For this reconstruction we have  $\Delta Q = 1.4 \times 10^{-8}$  and  $\Delta\rho = 1.1 \times 10^{-8}$ . The corresponding entropy  $S = 0.694$  is close to  $\ln 2$ .

Here we note that the size of the bin  $\Delta x$  does not play a significant role in the reconstruction via the *MaxEnt* principle. We will discuss this issue in more detail below, but now we concentrate our attention on the role of the number  $N_\theta$  of quadrature angles.

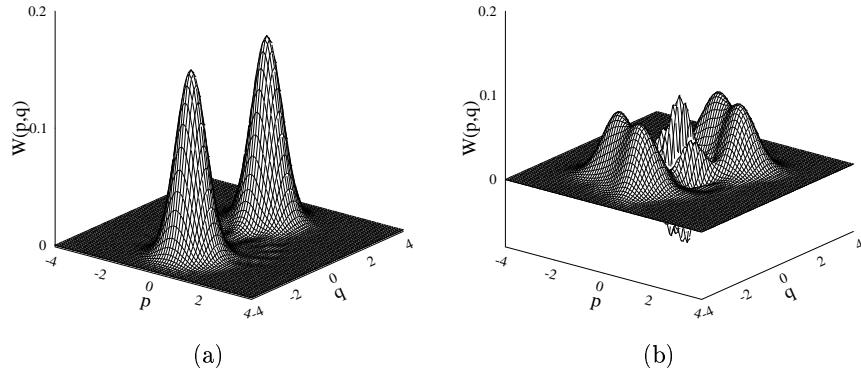
#### 6.4.1 Minimal number of measured quadratures

The two quadratures  $\hat{q} = \hat{x}_{\theta=0}$  and  $\hat{p} = \hat{x}_{\theta=\pi/2}$  are sufficient for a reconstruction of the even coherent state when  $\alpha$  is real (this corresponds to specific a priori information about the state). If we consider the most general case, when  $\alpha$  is complex, then the reconstruction based on the measurement of just two quadratures is not very good (in what follows, instead of choosing the complex  $\alpha$  we will use correspondingly rotated quadratures).

Firstly, let us consider a reconstruction of the even coherent state with the real amplitude  $\alpha$  based on the measurement of two rotated quadratures which are not mutually orthogonal (see Fig. 6.4). In particular, let us assume  $\hat{x}_{\theta=0}$  and  $\hat{x}_{\theta=\pi/8}$ . Other settings are the same as in figure Fig 6.2b. We plot the reconstructed Wigner function in Fig. 6.4a.



**Fig. 6.3.** (a) The Wigner function of the mixture of two coherent states with  $\alpha = \pm 2$ . (b) The reconstruction of the Wigner function via the *MaxEnt* principle based on the measurement of marginal distributions for the position and the momentum ( $N_\theta = 2$ ). The measured marginal distributions are divided into  $N_x = 40$  bins of width  $\Delta x = 0.2$  covering the interval  $\langle -4, 4 \rangle$ . The fidelity of the reconstruction is  $\Delta Q = 1.4 \times 10^{-8}$  and  $\Delta \rho = 1.1 \times 10^{-8}$ , with entropy  $S = 0.694$ . (c) The tomography via pattern functions with  $N_\theta = 4$  and  $N_{max} = 8$  for which the minimum deviation  $\Delta Q = 0.71$  is obtained. The fidelity of the reconstruction is  $\Delta \rho = 0.17$ . The reconstructed density operator is unphysical. (d) For the larger number of quadratures  $N_\theta = 8$  with  $N_{max} = 8$  we obtain  $\Delta Q = 0.13$  and  $\Delta \rho = 0.023$ . The reconstructed Wigner function still exhibits some fictitious interference structure.



**Fig. 6.4.** The Wigner function of the even coherent state  $\alpha = 2$  reconstructed via *MaxEnt* principle from two particular marginal distributions ( $N_\theta = 2$ ). In Fig. (a) we consider two rotated quadratures for  $\theta = 0$  and  $\theta = \pi/8$ . The reconstruction leads to a statistical mixture of two coherent states. The fidelity of the reconstruction is  $\Delta Q \approx \times 10^{-6}$ ,  $\Delta\rho = 0.49$ , and the entropy is  $S = 0.7$ . (b) The choice of “measurement angles”  $\theta = \pi/8, \pi/2 + \pi/8$  leads to a mixture with  $\Delta Q \approx \times 10^{-7}$ ,  $\Delta\rho = 0.60$ , and the entropy is  $S = 1.31$ . Other settings (the number of bins and their width) are the same as in Fig. 6.2.

It is very similar to the Wigner function of the statistical mixture (which is also indicated by the value of the corresponding von Neumann entropy close to  $\ln 2$ ). The reason is that the type of the measurement considered in the example does not provide us with enough information about the interference pattern in the phase space (the two measured quadratures are “too” close).

Our next example is the case when the two measured quadratures are mutually orthogonal, but are rotated with respect to the position and the momentum. Specifically,  $\hat{x}_{\theta=\pi/8}$  and  $\hat{x}_{\theta=\pi/2+\pi/8}$  and other settings are the same as in figure Fig 6.2b. We plot the reconstructed Wigner function in Fig. 6.4b. In this Wigner function we see some interference pattern but the information from the measurement does not allow us to perform a reliable reconstruction ( $\Delta Q \approx 10^{-7}$ ,  $\Delta\rho = 0.60$ , and  $S = 1.31$ ). We can observe some “fictitious” peaks in the reconstructed Wigner function.

To improve the fidelity of the reconstruction we have to consider larger numbers of the rotated quadratures. In fact, it is our empirical experience that three rotated quadratures ( $N_\theta = 3$ ) are always sufficient to perform a very reliable *MaxEnt* reconstruction of an arbitrary unknown state. We have not found yet a rigorous proof for this empirical observation.

#### 6.4.2 Comparison with the reconstruction via direct sampling

It has been shown by Leonhardt et al. [24] that for a reliable reconstruction via direct sampling with the help of the pattern functions two conditions have

to be satisfied:

$$\begin{aligned} N_\theta &= N_{max} \\ \Delta x &< \pi/2\sqrt{2N_{max} + 1}. \end{aligned} \quad (6.68)$$

This means that the truncation of the Hilbert space in which the reconstructed density operator is defined specifies how many quadrature angles have to be considered as well as it puts some restriction on the size of the bin.

### The role of $N_\theta$

We start with the analysis of the first condition. In our case of the even coherent state with  $\alpha = 2$  we have to consider at least  $N_{max} = 8$ . Consequently, following Leonhardt we would have to consider a measurement of  $N_\theta = 8$  quadratures. In this case the precision of reconstruction is  $\Delta Q = 0.13$  and  $\Delta\rho = 0.03$  which is reasonable, but much smaller than in the case of the *MaxEnt* reconstruction. It is important to remember that any deviation of  $N_\theta$  from  $N_{max}$  causes a dramatic deterioration of the reconstruction scheme (for more details see Refs. [24, 35]). In particular, for  $N_{max} > N_\theta$  higher “ghost” Fock states appear in the reconstructed density matrix. This effect of aliasing (see Ref. [35]) is caused by the fact that in the sampling method the matrix elements  $\rho_{mn}$  with  $(m - n) \bmod N_\theta$  cannot be distinguished.

To see the effect of an insufficient number of phases for the sampling via pattern functions we plot in Fig. 6.2 the results for  $N_\theta = 2$  [see (c)] and  $N_\theta = 4$  [see (d)] marginal distributions. In both cases we chose  $N_{max}$  such that the parameter  $\Delta Q$  (deviation from the measured data) is minimized. In particular, for  $N_\theta = 2$  using the numerical search we have found that  $\Delta Q$  is minimized for  $N_{max} = 4$  when  $\Delta Q = 9.17$  and  $\Delta\rho = 3.7$ . However, the reconstructed density operator is unphysical – we obtain *negative* probabilities of odd Fock states:  $P_1 = -0.35$ ,  $P_3 = -0.54$ . The corresponding Wigner function is plotted in Fig. 6.2c. Analogously, for  $N_\theta = 4$  we have found the optimal truncation to be  $N_{max} = 6$ . In this case  $\Delta Q = 1.33$  and  $\Delta\rho = 0.51$ . The fidelity of the reconstruction is now better, but it still gives us an unphysical result with  $P_1 = -0.1$  (the corresponding Wigner function is plotted in Fig. 6.2d). We have checked that higher values of  $N_{max}$  significantly deteriorate the quality of reconstructions. Comparing the sampling method with the result of the *MaxEnt* approach we see the great advantage of the latter for a small number of quadrature phases.

Analogous results are obtained also for statistical mixtures (see Fig. 6.3). Specific values for the fidelities of the reconstruction are given in the figure caption.

### The role of $N_x$

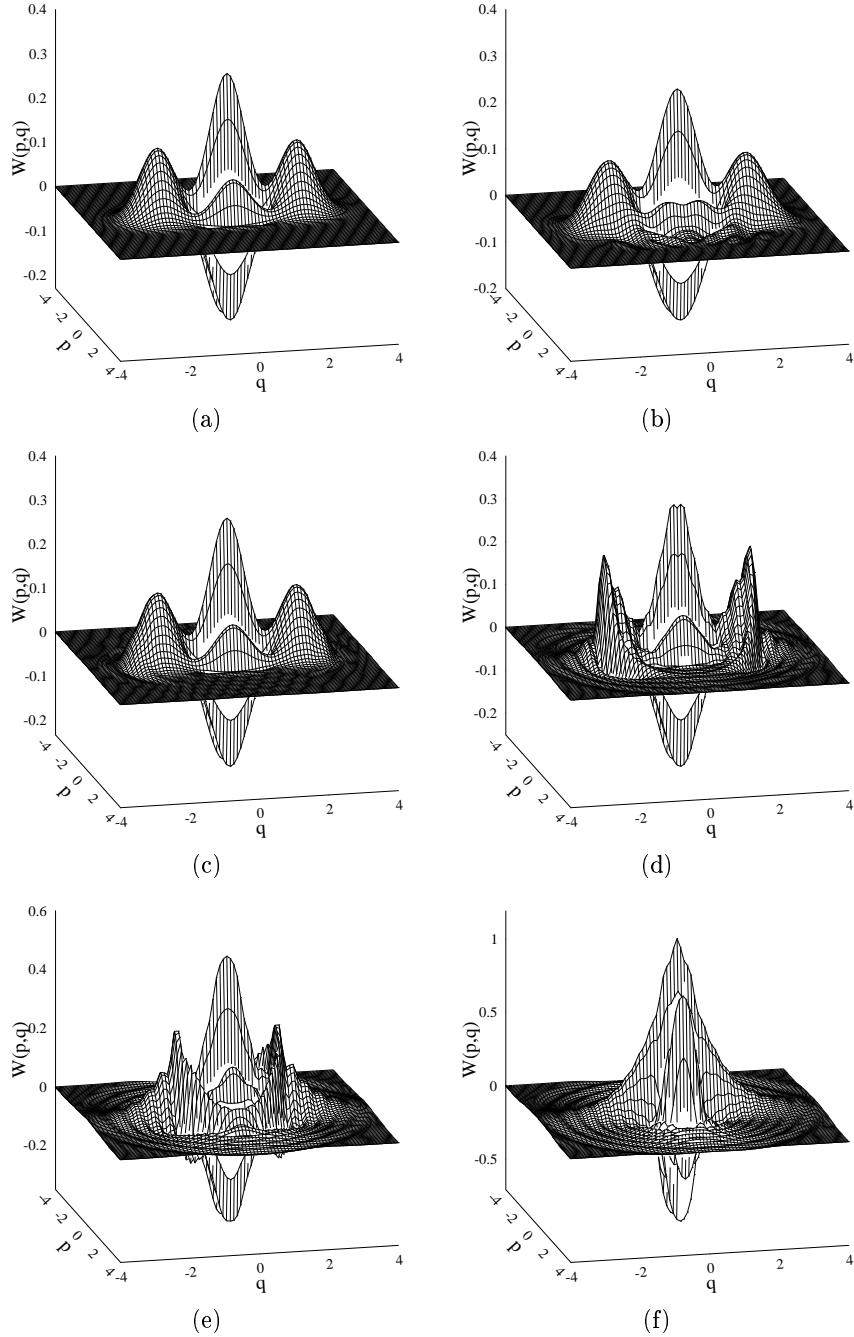
In addition to the required resolution of bins, i.e.  $\Delta x < \pi/2\sqrt{2N_{max} + 1}$ , the sampling via pattern functions is also very sensitive with respect to the size of the interval on which the marginals are measured. To apply the sampling via pattern functions marginal distributions have been measured on the whole interval where they take non-zero values. The importance of the distribution “tails” of the marginals can be illustrated on the example of the even coherent state. Let us consider the size of the bin (i.e. the resolution) to be  $\Delta x = 0.2$  and let us change the values of  $N_x$ .

In Fig. 6.5 we plot the reconstructed Wigner functions which are obtained from the incomplete marginal distributions via the *MaxEnt* [(a)-(b)] and via the sampling [(c)-(f)] reconstruction schemes.

In Fig. 6.5a we plot the reconstruction via the *MaxEnt* principle with two marginal distributions ( $N_\theta = 2$ ). The “measured” marginal distributions are divided into  $N_x = 20$  bins of width  $\Delta x = 0.2$  covering the interval  $\langle -2, 2 \rangle$ . The quality of the reconstruction is  $\Delta Q = 2 \times 10^{-15}$ ,  $\Delta\rho = 2 \times 10^{-10}$ , and  $S \approx 10^{-8}$  which is comparable with Fig. 6.2b when  $N_x = 40$  and  $x \in \langle -4, 4 \rangle$ . We see that even though the interval on which the marginal distributions are measured is shorter by the factor two the fidelity of reconstruction is not affected. In Fig. 6.5b we plot the reconstructed Wigner function under the same conditions except  $N_x = 10$ , i.e.  $x \in \langle -1, 1 \rangle$ . The quality of the reconstruction decreases to  $\Delta Q = 3 \times 10^{-15}$ ,  $\Delta\rho = 0.029$ , and  $S = 0.20$ , but still is rather reliable.

In Figs. 6.5c-f we present results of the reconstruction via direct sampling. We assume  $N_\theta = 20$  and  $N_{max} = 20$ . That is, we consider significantly more data than in the previous two cases. Nevertheless, results of the reconstruction are much worse. Specifically, let us assume the marginal distributions to be taken within the intervals (c)  $\langle -4, 4 \rangle$  which results in the reconstruction with the fidelity  $\Delta Q = 0.05$  and  $\Delta\rho = 0.002$ . Analogously, (d)  $\langle -3, 3 \rangle$  with  $\Delta Q = 0.88$ ,  $\Delta\rho = 0.154$ ; (e)  $\langle -2, 2 \rangle$  with  $\Delta Q = 8.6$ ,  $\Delta\rho = 1.01$ ; (f)  $\langle -1, 1 \rangle$  with  $\Delta Q = 47.7$ ,  $\Delta\rho = 5.82$ . We conclude that the shorter the interval on which the quadrature distribution is measured the less reliable the reconstruction is. From Figs. 6.5 it is seen that the sampling via pattern functions can reconstruct only a structure within a measured region of the phase space.

From above we can conclude, that for a reliable use of the reconstruction via the direct sampling we have to measure a sufficient number of quadrature distributions ( $N_\theta = N_{max}$ ) on the whole interval of  $x$ . On the contrary, with the *MaxEnt* approach we need just a small number of quadrature distributions and the interval on which the distributions are measured can be rather small. We have also analyzed the situation when the distributions are measured on the interval which is not symmetric with respect to the origin of the phase space. In this case the *MaxEnt* scheme works very reliably while the direct sampling fails completely.



**Fig. 6.5.** The reconstruction of the Wigner function of the even coherent state with  $\alpha = 2$ . (a) The reconstruction via the *MaxEnt* principle with two marginal distributions ( $N_\theta = 2$ ) for the position and the momentum. The “measured” marginal distributions are divided into  $N_x = 20$  bins of width  $\Delta x = 0.2$  covering the interval  $\langle -2, 2 \rangle$ . The fidelity of the reconstruction is  $\Delta Q = 2 \times 10^{-15}$  and  $\Delta \rho = 2 \times 10^{-10}$  which is comparable with Fig. 6.2b (where  $N_x = 40$  corresponding to  $x \in \langle -4, 4 \rangle$ ). (b) This is the same example as (a) except  $N_x = 10$ , i.e.  $x \in \langle -1, 1 \rangle$ . The quality of the reconstruction decreases to  $\Delta Q = 3 \times 10^{-15}$  and  $\Delta \rho = 0.029$ . (c)-(f) Reconstruction via pattern functions with  $N_\theta = 20$  and  $N_{max} = 20$ . The marginals are taken within the intervals (c)  $\langle -4, 4 \rangle$  [ $\Delta Q = 0.05$ ,  $\Delta \rho = 0.002$ ]; (d)  $\langle -3, 3 \rangle$  [ $\Delta Q = 0.88$ ,  $\Delta \rho = 0.154$ ]; (e)  $\langle -2, 2 \rangle$  [ $\Delta Q = 8.6$ ,  $\Delta \rho = 1.01$ ]; (f)  $\langle -1, 1 \rangle$  [ $\Delta Q = 47.7$ ,  $\Delta \rho = 5.82$ ]. We see that the shorter the interval on which the quadrature distribution is measured the less reliable the reconstruction is.

The *MaxEnt* reconstruction scheme can be applied for various quantum systems. In particular, in what follows we will analyze how vibrational states of trapped neutral atoms can be reconstructed.

## 6.5 Reconstruction of motional states of neutral atoms

In atomic optics a highly coherent control of motional degrees of freedom has been achieved for neutral atoms [36,37]. In order to verify the degree (fidelity) of coherent control over motional degrees of freedom of neutral atoms one can consider a reconstruction of their motional states from measured data. In what follows we will perform this type of analysis.

Recently, experimental manipulations of motional quantum states of neutral atoms have been reported by the group of C. Salomon in Paris [36,37]. Cold Cs atoms can be cooled into specific quantum states of a far detuned 1D optical lattice. The optical lattice is induced by the interference of two laser beams. Along the vertical  $z$  axis a periodic potential of “harmonic” microwells is produced with a period of 665 nm and with an amplitude of about 0.2  $\mu\text{K}$  [37]. The vertical oscillation frequency in a microwell at the center of the trap is  $\omega_z/2\pi = 85$  kHz. The corresponding ground state has the rms size  $\Delta z_0 = \sqrt{\hbar/2m\omega_z} \approx 21$  nm and  $\Delta p_0/m = \sqrt{\hbar\omega_z/2m} \approx 11$  mm/s is its rms velocity width. The trapped cloud of neutral Cs atoms has a nearly Gaussian shape with a vertical rms size  $\Delta\xi_0 = 53$   $\mu\text{m}$ . With the help of deterministic manipulations the neutral atoms can be prepared in non-classical 1D motional states along the vertical axis such as squeezed states, number states, or specific superpositions of number states [37]. The measurement of the prepared quantum state  $\hat{\rho}$  is performed as follows: The system is evolved within the harmonic potential during the time  $\tau$ . Then the lasers are turned off and the system undergoes the ballistic expansion (BE). After the time of flight  $T = 8.7$  ms a 2D absorption image of the cloud is taken in 50  $\mu\text{s}$  with a horizontal beam [37]. Integration of 2D absorption images in the horizontal direction gives us the spatial distribution along the vertical  $z$  axis. Therefore we will consider only a 1D quantum-mechanical system along the vertical axis.

To confirm that a desired quantum state has been obtained (engineered) one can compare the spatial distributions along the vertical axis with the predicted ones. The coincidence of these spatial distributions is a necessary but not the sufficient requirement. A complete verification of the fidelity of the preparation of desired quantum states requires a quantum state reconstruction procedure. In order to perform this task we adopt the *MaxEnt* principle [29]. To do so we utilize a close analogy between quantum homodyne tomography [11] and the BE absorption imaging for the case of the point-like cloud (with the rms size equal to zero).

### 6.5.1 Motional states of atoms via MaxEnt principle: Formalism

In the quantum homodyne tomography the probability distributions are measured for the rotated quadrature operators  $\hat{x}_\theta$ . The annihilation and creation operators of motional quanta,  $\hat{a}$  and  $\hat{a}^\dagger$ , are related to the position and momentum operators,  $\hat{z}$  and  $\hat{p}$ , via expressions  $\hat{z} = \frac{1}{\sqrt{2}}(\hat{a} + \hat{a}^\dagger)$  and  $\hat{p} = \frac{1}{\sqrt{2}}i(\hat{a} - \hat{a}^\dagger)$ , respectively. The angle  $\theta$  of the quadrature operator corresponds to  $\omega_z\tau$  and vertical “cuts” of the absorption images (taken after the BE) can be associated with quadrature probability distributions. However, for a real physical situation with a nonzero rms size of the cloud the vertical “cuts” of absorption images correspond to a coarse-grained quadrature probability distributions. In particular, the vertical cuts of measured absorption images (taken in 2D) give us (after integration along the horizontal direction) the spatial distribution along the vertical axis. The spatial distribution along the vertical  $z$  axis can be expressed as (in order to reflect a different physical origin of the system under consideration in this Section we will denote observables as  $\hat{F}$  instead of  $\hat{Q}$  that has been used in Section 6.4)

$$\bar{F}_\tau(z) = T^{-1} \int F_0(\xi_0) P_\tau((z - \xi_0)/T) d\xi_0, \quad (6.69)$$

where  $F_0(\xi_0)$  is the initial spatial distribution of the cloud in the  $z$ -direction (i.e., a Gaussian distribution with the rms size  $\Delta\xi_0$ ). The function  $P_\tau(v)$  denotes the velocity probability distribution of the measured quantum state which has been evolved for time  $\tau$  in the harmonic potential before the BE, i.e.

$$P_\tau(v) = |\langle v | \psi(\tau) \rangle|^2, \quad |\psi(\tau)\rangle = \hat{U}(\tau)|\psi(0)\rangle. \quad (6.70)$$

Here  $\hat{U}(\tau) = \exp(-i\hat{H}\tau/\hbar)$  represents the time-evolution operator for the harmonic oscillator with the Hamiltonian  $\hat{H} = \hat{p}^2/2m + m\omega_z^2\hat{z}^2/2$ . Now we can treat the measured “cuts” as the mean values of specific observables:  $\bar{F}_\tau(z) = \text{Tr}[\hat{\rho}\hat{F}_\tau(z)]$ . In practice just a few discrete times  $\tau_j$  ( $j = 1, \dots, N_\tau$ ) are considered and the  $z$  coordinate is discretized into the bins  $z_k$  ( $k = -N_z, \dots, N_z$ ) of a given resolution  $\Delta z$ . The set of operators that enters the equation for the MaxEnt reconstruction then takes the form

$$\hat{F}_{\tau_j}(z_k) = T^{-1} \int F_0(\xi_0) \hat{U}^\dagger(\tau_j) \left| \frac{z_k - \xi_0}{T} \right\rangle \left\langle \frac{z_k - \xi_0}{T} \right| \hat{U}(\tau_j) d\xi_0, \quad (6.71)$$

where ( $j = 1, \dots, N_\tau; k = -N_z, \dots, N_z$ ). We have already commented in Section 6.3 that the operator of mean phonon number  $\hat{n}$  is added to the set of observables  $\{\hat{F}_{\tau_j}(z_k)\}$ . Knowledge of the mean excitation number  $\bar{n}$  is essential in the case of an incomplete set of observables [29]. Knowledge of the mean excitation number leads to a natural “truncation” of the Hilbert space. The inclusion of the mean phonon number into the MaxEnt reconstruction

scheme does not represent its limitation as the mean energy represents one of basic characteristics of any system that should be inferred from the measurement.

The experimental “cuts” of the BE absorption images  $\{\hat{F}_{\tau_j}(z)\}$  can be taken at few selected times, for example  $\omega_z \tau_j = 0, \pi/4, \pi/2, 3\pi/4$  ( $N_\tau = 4$ ). To perform the reconstruction we have to determine the Lagrange multipliers  $\{\lambda_{j,k}\}$  and  $\lambda_{\bar{n}}$  associated with  $\{\hat{F}_{\tau_j}(z_k)\}$  and  $\hat{n}$ , respectively, in the expression for the generalized canonical density operator. The Lagrange multipliers can be determined via the minimization of a deviation function  $\Delta F$  with respect to the measured data, i.e.

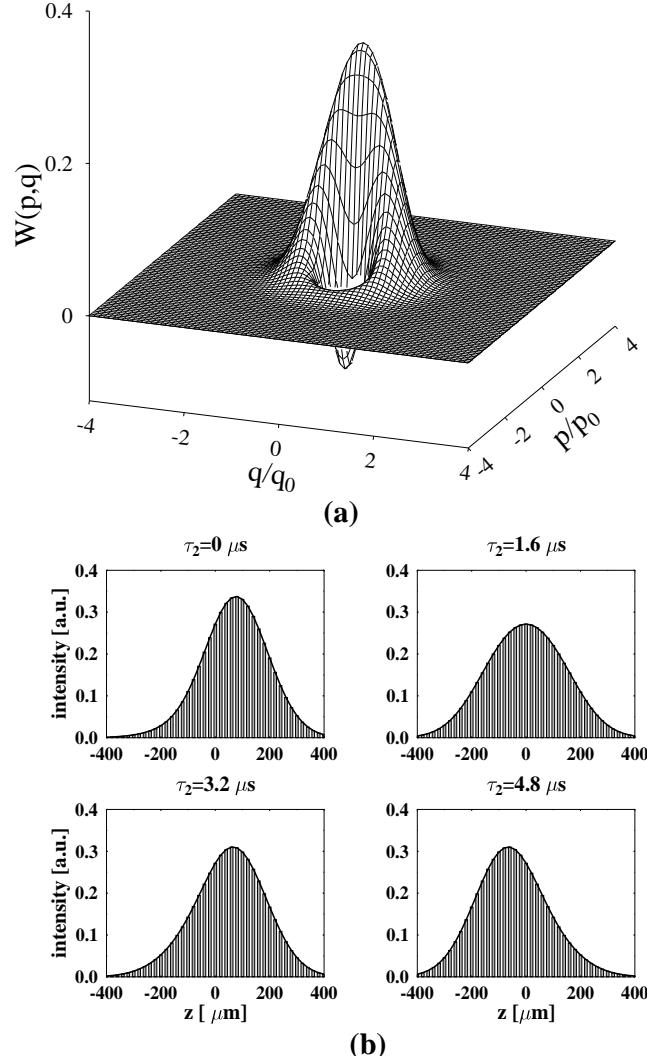
$$\begin{aligned} \Delta F = & \sum_{j,k} w_{j,k} \left\{ \bar{F}_{\tau_j}(z_k) - \text{Tr} \left( \hat{\rho}_r \hat{F}_{\tau_j}(z_k) \right) \right\}^2 \\ & + w_{\bar{n}} \{ \bar{n} - \text{Tr} (\hat{\rho}_r \hat{n}) \}^2. \end{aligned} \quad (6.72)$$

Here  $\{w_{j,k}\}$  and  $w_{\bar{n}}$  represent positive weight factors for particular observables. Without any prior knowledge about the state, for simplicity we can take  $w_{i,j} = 1$ . The weight factor  $w_{\bar{n}}$  associated with the mean phonon number can be chosen according to our preference either to fit better the “cuts” of the BE images or the mean phonon number. In the case of the perfect measurement and the complete reconstruction the result has to be independent of the choice of the weight factors (in this case we can take  $w_{\bar{n}} = 1$ ). The weight factors could also be associated with the *prior* information about the dispersion of the measured observables. In particular, the weight factors can be taken as  $w_\nu \sim \sigma_\nu^{-2}$  to reflect the knowledge of variances  $\sigma_\nu$  for the measured observables  $\hat{G}_\nu$ . When the mean values of the observables for the MaxEnt estimate  $\hat{\rho}_r$  fit within desired interval  $\bar{G}_\nu \pm \sigma_\nu$  then contributions of the observables to the deviation function  $\Delta F$  are of the same order ( $\sim 1$ ). However, in our case we do not assume the knowledge of variances for the measured discretized probability distributions (taking  $w_\nu = 1$ ).

Once the Lagrange multipliers are numerically fitted, the result of the reconstruction – the generalized canonical density operator  $\hat{\rho}_r$  – can be visualized, for example, via the corresponding Wigner function.

### 6.5.2 Numerical simulation of MaxEnt tomography

To test our reconstruction procedure let us consider the reconstruction of the Wigner function of the motional quantum state  $|\psi(0)\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$  of Cs atoms trapped in the optical lattice. This kind of state has been demonstrated in recent experiments [37]. We assume the following setup parameters:  $\omega_z/2\pi = 80$  kHz, the rms size of the ground state  $\Delta z_0 = 22$  nm, the rms velocity width  $\Delta p_0/m = 11$  mm/s and the rms width of the cloud of the atoms about 60  $\mu\text{m}$ . Before BE (with BE time  $T = 8.7$  ms) the atoms evolve within the harmonic trapping potential for  $\tau = 0, 1.6, 3.2, 4.8$   $\mu\text{s}$ . As the input for



**Fig. 6.6.** (a) Numerical simulation of the reconstruction of the Wigner function of the motional quantum state  $(|0\rangle + |1\rangle)/\sqrt{2}$  of Cs atoms trapped in the optical lattice (assuming  $\omega_z/2\pi = 80$  kHz, the rms size of the ground state  $\Delta z_0 = 22$  nm and the rms velocity width  $\Delta p_0/m = 11$  mm/s). For the reconstruction via the MaxEnt principle four vertical cuts of the absorption images (with BE time  $T = 8.7$  ms) have been taken (b). The histograms correspond to the measured data while the solid lines are obtained from the reconstructed Wigner function (i.e. they correspond to reconstructed marginal distributions). Before BE the atoms evolve within the trapping potential for the times  $\tau = 0, 1.6, 3.2, 4.8 \mu s$ . In addition, the mean number of motional quanta  $\bar{n} = 0.5$  and the rms width of the cloud of the atoms about 60  $\mu m$  have been assumed.

the reconstruction via the MaxEnt principle four vertical “ideal” cuts of the BE absorption images are taken as shown in Fig. 6.6(b).

In addition, for the phonon number operator  $\hat{n}$ , which is included in the set of measured observables (see discussion above), we assume the mean value  $\bar{n} = 0.5$ . The result of the ideal reconstruction is shown in Fig. 6.6. The fidelity of the measured and the reconstructed quantum states is close to unity, which means a perfect reconstruction with  $\Delta F = 10^{-10}$ , entropy  $S = 10^{-7}$ ,  $\Delta\rho = 10^{-8}$  has been achieved. Here  $\Delta\rho = \sum_{m,n} |(\hat{\rho} - \hat{\rho}_r)_{mn}|^2$  denotes a deviation of the original and reconstructed density operators.

Obviously, in a real measurement the measured values are always fluctuating around the exact ones due to an experimental noise. Therefore we simulate a non-ideal measurement introducing random fluctuations to the measured values of observables. It means that instead of the ideal values  $\bar{F}_{\tau_j}(z_k)$  we use for the MaxEnt reconstruction procedure the fluctuating (“noisy”) values

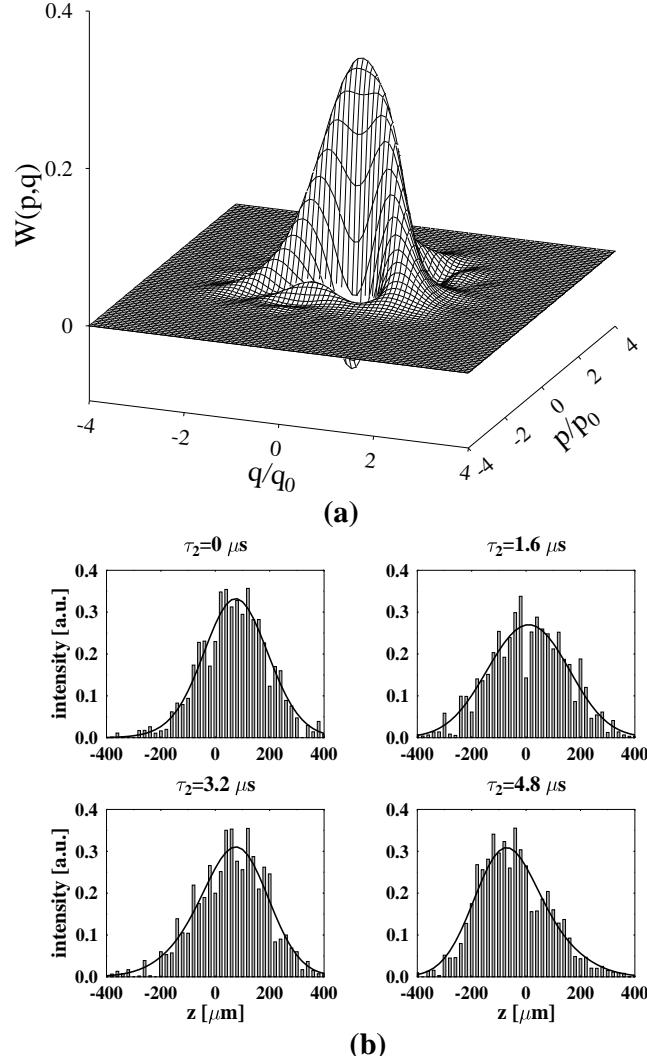
$$\bar{F}'_{\tau_j}(z_k) = \bar{F}_{\tau_j}(z_k) + \eta \xi_{j,k} (\bar{F}_{\tau_j}(z_k))^{1/2}. \quad (6.73)$$

Here  $\eta$  is a relative-error parameter that characterizes the quality of the measurement and  $\{\xi_{j,k}\}$  represents a Gaussian noise for observables. The result of the reconstruction is shown in Fig. 6.7 for  $\eta = 0.1$ . Noisy mean values of the observables are shown in Fig. 6.7(b). Inspite of a significant relative error, the reconstruction is almost perfect, with the fidelity of the measured and the reconstructed states still close to one ( $\Delta F = 0.16$ , entropy  $S = 0.01$ ,  $\Delta\rho = 0.05$ ). The minimum value of the deviation function  $\Delta F = 0.16$  can serve also as a measure of the imperfection of the given measurement (due to a technical noise) [38].

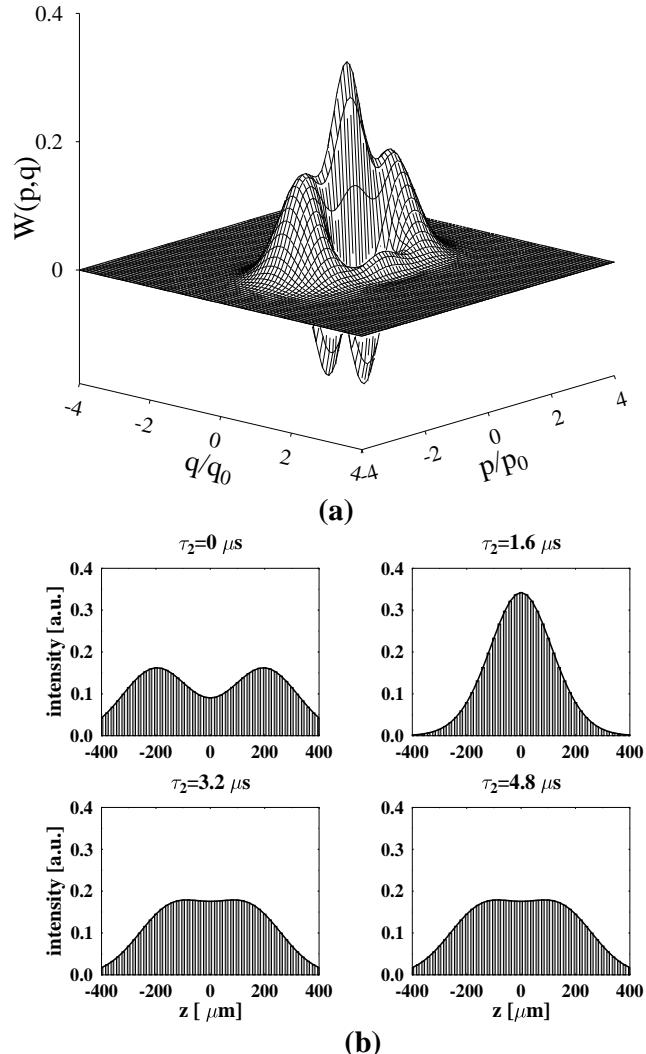
A typical non-classical state that we can utilize for a further test is the even coherent state  $\mathcal{N}_e(|\alpha\rangle + |-|\alpha\rangle)$  [8]. We have presented a numerical simulation of the reconstruction of this state in Fig. 6.8. For the amplitude  $\alpha = \sqrt{2}$  we obtained  $\Delta F = 10^{-8}$ , the entropy  $S = 0.026$  and  $\Delta\rho = 10^{-4}$  (under assumption that the exact mean phonon number  $\bar{n} = 1.928$  is known). In the case of the imperfect measurement with  $\eta = 0.1$  the reconstruction leads to  $\Delta F = 0.14$ , entropy  $S = 0.13$ , and  $\Delta\rho = 0.06$  for  $\bar{n} = 2.09$ . The fidelity of the reconstructed and the measured states is, in this case, also close to one.

In order to model a technical noise in the measurement we have considered Gaussian fluctuations proportional to the square root of the mean values. It means that tails of the “cuts” of BE images do not introduce a significant error [compare Fig. 6.8(b) and Fig. 6.9(b)]. However, in the current measurements the situation seems to be different and the fluctuations do not decrease with the amplitude of the expected values.

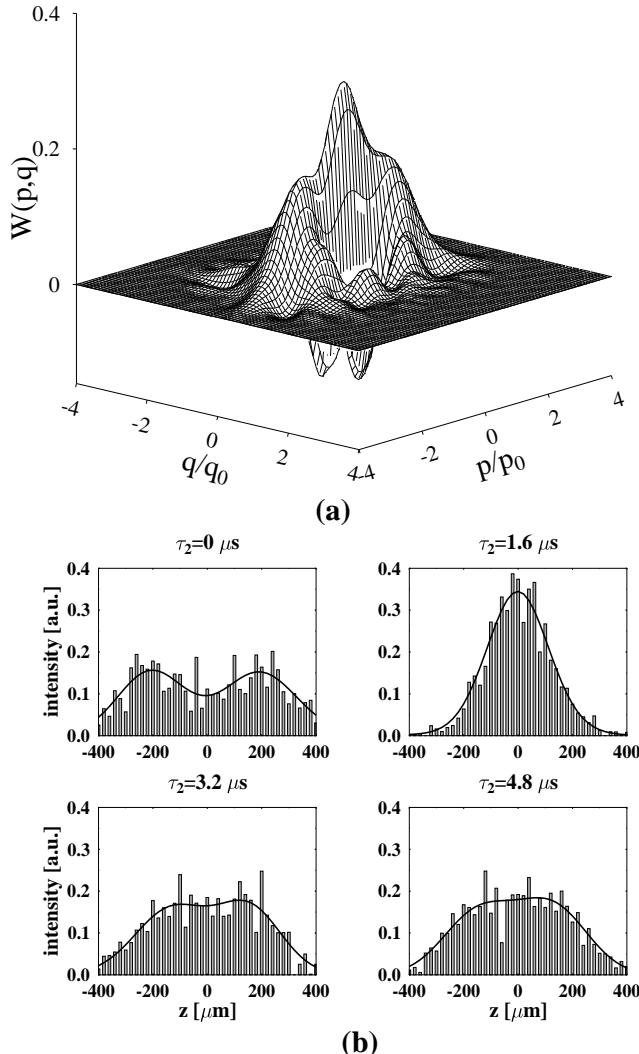
The fundamental question in the context of the *MaxEnt* reconstruction of states from incomplete tomographic data is whether the quality of the reconstruction can be improved using additional data from subsequent time moments  $\tau$  and how many such time moments  $\tau$  are required for the complete



**Fig. 6.7.** (a) Numerical simulation of the reconstruction of the Wigner function of the atomic motional quantum state  $(|0\rangle + |1\rangle)/\sqrt{2}$  for the same settings as in Fig. 6.6. (b) Four vertical cuts of the absorption images taken for reconstruction are fluctuating randomly around their ideal values shown in Fig. 6.6(b) with the relative error  $\eta = 0.1$ . The histograms correspond to the measured data while the solid lines are obtained from the reconstructed Wigner function. In addition, the mean phonon number  $\bar{n}' = 0.6$  has been considered.



**Fig. 6.8.** Numerical simulation of the reconstruction of the Wigner function of the motional quantum state  $\mathcal{N}_e(|\alpha\rangle + |-\alpha\rangle)$  with  $\alpha = \sqrt{2}$  in the case of the ideal measurement. The mean number of motional quanta  $\bar{n} = 1.928$ . Other settings are the same as in Fig. 6.6.



**Fig. 6.9.** Numerical simulation of the reconstruction of the Wigner function of the motional quantum state  $\mathcal{N}_e(|\alpha\rangle + |-\alpha\rangle)$  with  $\alpha = \sqrt{2}$  in the case of noisy measurement with  $\eta = 0.1$ . The “measured” mean number of the motional quanta  $\bar{n}' = 2.09$ . Other settings are the same as in Fig. 6.7.

reconstruction of the unknown state  $\hat{\rho}$ . As we have shown in Section 6.4 for the efficient quantum tomography just *three* quadrature distributions are sufficient for a complete reconstruction using the MaxEnt principle (in the case of the perfect measurement). This corresponds to the ideal case without the spatial dispersion of the cloud of atoms, i.e. the choice with  $\omega_z \tau_j = 0, \pi/4, \pi/2$  ( $N_\tau = 3$ ) is sufficient for  $\Delta\xi_0 \rightarrow 0$ . Obviously, in experiments with neutral atoms the spatial size of the atomic cloud is nonzero. However, in the case of the ideal measurement three BE absorption images associated with three “rotations”  $\omega_z \tau_j$  are still sufficient for a complete reconstruction of tested examples of quantum states. On the other hand, it seems that for higher mean phonon numbers the spatial distributions along the vertical axis that are directly determined from absorption images should be known with improving precision (and on a wider interval of values as well). In the above examples we have considered for convenience BE images for four “rotations” ( $N_\tau = 4$ ) which results in a very good reconstruction.

### 6.5.3 Reconstruction from experimental data

In what follows we will apply the MaxEnt reconstruction scheme to the data obtained at the ENS in Paris [39]. Firstly we note that the unknown quantum state should belong to a Hilbert subspace that can be determined easily. Thus we can limit ourselves to the subspace spanned by Fock (number) states  $|0\rangle, |1\rangle, \dots, |N-1\rangle$ . The upper bound on the accessible phonon number  $N$  is given by experimental limitations such as, a feasible depth of microwells of the optical lattice and the validity of the harmonic potential approximation. For recent experiments  $N$  has been typically of the order of 10. This value is large enough to demonstrate the preparation of many non-classical states but on the other hand excludes highly squeezed states from coherent processing.

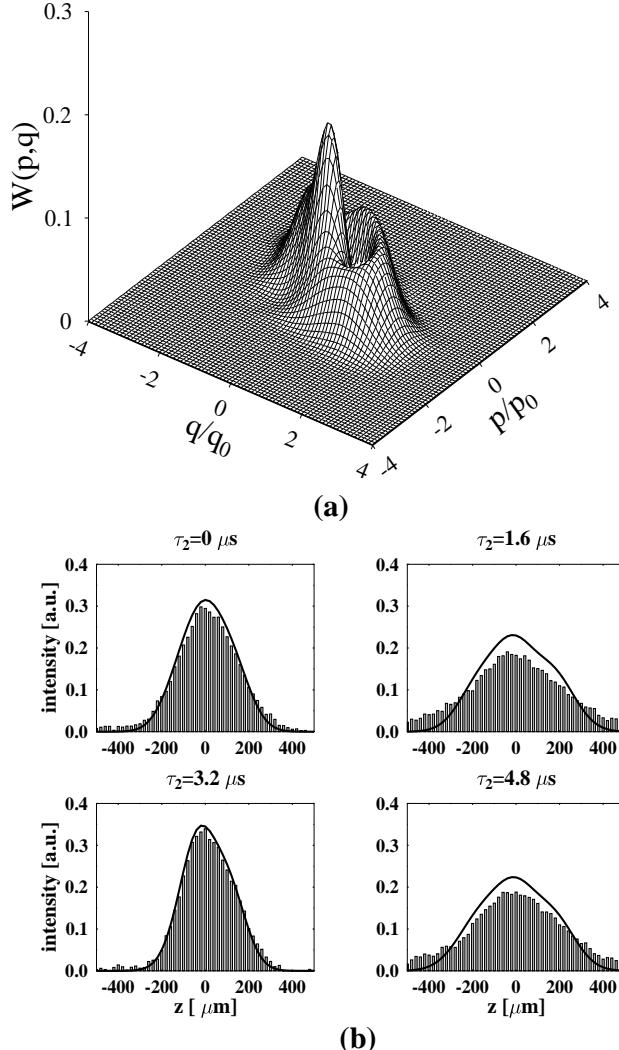
Let us consider the experimental arrangement used in Paris [39] with the parameters:  $\omega_z/2\pi = 80$  kHz, the rms size of the ground state  $\Delta z_0 = 22$  nm, the rms velocity width  $\Delta p_0/m = 11$  mm/s, the rms width of the cloud of the atoms about 60  $\mu\text{m}$  and BE time  $T = 8.7$  ms. Initially the atoms are prepared in a well-defined motional state  $|\psi_0\rangle$  (e.g. in the vacuum state  $|0\rangle$ ). Then the optical lattice is switched off for the time period  $t_1$  during which the atoms evolve freely towards the state  $|\psi_1\rangle = \exp(-it_1\hat{p}^2/2m)|\psi_0\rangle$ . Next, the optical lattice is again switched on for the time  $\tau$  during which the atoms evolve within the harmonic trapping potential. The measurement is performed after the BE time. The first two stages can be considered as the preparation of the state  $|\psi_1\rangle$ . After its “rotation” by  $\omega_z \tau$  (within the phase space of the harmonic oscillator) and the subsequent BE the absorption images are taken.

The considered data are for the initial vacuum state which means that under ideal conditions a squeezed state  $|\psi_1\rangle = \exp(-it_1\hat{p}^2/2m)|0\rangle$  should be prepared. The vertical spatial distributions obtained from the measured 2D-absorption images are discretized into pixels (bins) with the pixel width 5.45  $\mu\text{m}$ . The optical density of each pixel is averaged in the horizontal direction

in which the absorption images are divided into 50 rows, each  $3.9 \mu\text{m}$  wide (these rows cover the size of the cloud in the horizontal direction). For the reconstruction via the MaxEnt principle four vertical spatial distributions for “rotation” times  $\tau = 0, 1.6, 3.2$  and  $4.8 \mu\text{s}$  are taken. The selected times roughly correspond to rotations within the phase space by  $\omega_z \tau = 0, \pi/4, \pi/2$  and  $3\pi/4$ , respectively. Unfortunately, the mean excitation number  $\hat{n}$  for measured state  $|\psi_1\rangle$  was not measured explicitly in the experiment, therefore we have to estimate it as follows: During the free expansion period the rms size of the cloud increases by  $\Delta x = p_0 \tau_1 / m$ . The corresponding increase of the potential energy  $\frac{1}{2}m\omega_z^2(\Delta x)^2$  in units  $\hbar\omega_z$  gives us the increase of the number of excitation quanta with respect to the initial state  $|\psi_0\rangle$ . For  $\tau_1 = 4 \mu\text{s}$  and the initial vacuum it means  $\bar{n} \approx 1$ . Experiments can be realized also for higher  $\tau_1$ . For example,  $\tau_1 = 8 \mu\text{s}$  leads to  $\bar{n} \approx 4$ . However, as mentioned above, such “squeezed” states with a significant contribution of higher phonon number states violate the underlying harmonic approximation for the potential. To keep a coherent control an anharmonic part of the potential has to be taken into account.

The result of the reconstruction via the MaxEnt principle is shown in Fig. 6.10. The deviation of the fitted and measured values is  $\Delta F = 0.09$  and the entropy of the reconstructed state is  $S = 1.0$ . It means that the reconstructed state is a statistical mixture. We see a two peak structure that suggests that there is a mixture of two squeezed states coherently displaced from each other. It is caused by the fact that the vertical center of the cloud was not fixed in the experiment and it has to be determined by our fit for each measured BE absorption image separately. Assuming a priori knowledge that the Wigner function has a symmetric shape with respect to the origin of the phase space (i.e. there is no coherent amplitude) a Gaussian fit can be used to determine the center of the cloud for each vertical distribution. For states with a non-zero coherent amplitude the center of the cloud should be fixed already in the experiment.

It turns out that the reconstruction results do not describe the squeezed vacuum state as was originally expected [39]. The main reason is that the mean phonon number was not measured directly in the experiment. It can be inferred only indirectly from the ideal case without any incoherence during preparation or measurement. As we discussed above, it is essential to include the information about the mean number of vibrational quanta into the MaxEnt reconstruction scheme. In optical tomography the analogous information about mean photon number can be obtained from distributions of two “orthogonal” quadratures. In our case it could correspond to two absorption images such that  $\omega_z(\tau_j - \tau_k) = \pi/2$ . However, it would require a precise timing of the evolution within the harmonic trapping potential. Therefore the mean number of vibrational quanta should be determined in an independent measurement.



**Fig. 6.10.** The Wigner function reconstructed from the experimental data obtained at the ENS, Paris. The experimental setting is the same as for Fig. 6.6. From the experimental data we have inferred the mean number of motional quanta  $\bar{n} \approx 1.0$ , while the reconstructed value is  $\bar{n}' \approx 1.1$ . Deviation of the measured and predicted values of observables is  $\Delta F = 0.09$  and entropy of the reconstructed mixture state is  $S = 1.0$ . Subtraction of a background from the measured marginals gives almost the same Wigner function and reduces significantly a difference between measured and reconstructed marginals.

Another problem arises from a slow convergence of anti-squeezed spatial distributions that are derived directly from the measured absorption images. In particular, the convergence of tails is too slow for those “rotations” that correspond to anti-squeezed phases, i.e.  $\tau = 1.6, 4.8 \mu\text{s}$  [see Fig. 6.10(b)]. The slow convergence is reflected by the presence of non-negligible backgrounds for Gaussian fits to these spatial distributions. If we eliminate (subtract) these backgrounds from the measured distributions the MaxEnt reconstruction gives almost the same Wigner function as in Fig. 6.10(a) but with a highly reduced deviation function  $\Delta F = 0.02$  (comparing to  $\Delta F = 0.09$  in Fig. 6.10). Such background in these absorption images can be caused by an incoherence associated with a violation of the harmonic approximation. In fact, in our analysis we have neglected the change of the oscillation frequency along the  $z$ -axis. In recent experiments, the oscillation frequency decreases 10% from  $\omega_z$  for microwells at the edge of the initial cloud.

In this section we have applied the *MaxEnt* scheme for a reconstruction of motional quantum states of neutral atoms. As an example we have analyzed the experimental data obtained by the group of C. Salomon at the ENS in Paris and we reconstruct the Wigner function of motional quantum states of Cs atoms trapped in the optical lattice. In our analysis we have neglected the change of the oscillation frequency along  $z$  axis in recent experiments. The dispersion of the oscillation frequency is of the order of a few percent. This source of errors can significantly affect the quality of a quantum state preparation and its reconstruction. In addition, only up to the first 10 bound states of microwells of the optical lattice can be approximated by a harmonic potential. It implies limits on coherent manipulations of quantum states. It means that states with a significant contribution of higher number (Fock) states cannot be prepared and manipulated in a controlled way.

## 6.6 Direct measurement of Wigner functions of single mode fields in cavities

We have shown in previous sections that Wigner functions of bosonic fields can be reconstructed from the tomographic data either via the inverse Radon transformation or using the method of pattern functions. In the case where the experimental data are not complete both these procedures can be substituted by a very efficient method of reconstruction based on the MaxEnt principle. Recently several groups have discussed the possibility of performing a reconstruction of Wigner functions of bosonic fields based on other tomographic measurements. Specifically, Englert *et al.* [40] have shown that the Wigner function can be measured *directly* via a measurement of the parity operator of the bosonic field. This method has been further developed by Lütterbach and Davidovich [41]. The essence of the method is based on a simple relation between the value of the Wigner function  $W(\alpha)$  at a phase-space point  $\alpha$  and the mean value of the parity operator  $\hat{P}$ :

$$W(\alpha) = 2\text{Tr} \left[ \hat{D}(-\alpha)\hat{\rho}\hat{D}(\alpha)\hat{P} \right], \quad (6.74)$$

where the  $\hat{D}(\alpha) = \exp(\alpha\hat{a}^\dagger - \alpha^*\hat{a})$  is displacement operator [see also Eq. (6.8)] and the field parity operator  $\hat{P}$  is defined as

$$\hat{P} = \exp [i\pi \hat{a}^\dagger \hat{a}] . \quad (6.75)$$

This operator acts on a Fock state  $|n\rangle$  as  $\hat{P}|n\rangle = (-1)^n|n\rangle$ . With this normalization,  $-2 \leq W \leq 2$  and  $\int W(\alpha) d\alpha = \pi$  (see also Section 6.2.2).

Lutterbach and Davidovich have shown that if the single-mode field in the state  $\hat{\rho}$  is stored in a cavity  $C$  and if it can be shifted (displaced) by injecting in  $C$  a coherent field with an amplitude  $-\alpha$ , then one can perform a measurement of the parity operator  $\hat{P}$  on the resulting field  $\hat{\rho}(\alpha) = \hat{D}(-\alpha)\hat{\rho}\hat{D}(\alpha)$ . Repeating this measurement many times for each value of  $\alpha$ , one obtains a value of the Wigner function at the given phase-space point  $\alpha$

$$\frac{W(\alpha)}{2} = \text{Tr} \left[ \hat{\rho}(\alpha) \hat{P} \right] = \langle \hat{P} \rangle = \sum_{n=0}^{\infty} (-1)^n \rho_{nn}(\alpha) . \quad (6.76)$$

In order to measure the mean value of the parity operator  $\langle \hat{P} \rangle$  in the cavity it is enough to send across the cavity a sequence of two-level atoms with a known velocity. The transition between two levels (denoted as  $|e\rangle$  and  $|g\rangle$ , respectively) of each atom is slightly de-tuned from the resonance with the cavity mode (the frequency mismatch is denoted as  $\Delta$ ). When the cavity mode is in a Fock state  $|n\rangle$  the atomic transition frequency is light shifted at the center of the cavity by  $\Omega^2 n / 2\Delta$  where  $\Omega$  is the resonant vacuum Rabi frequency (see for instance Ref. [42]). The frequency mismatch  $\Delta$  is chosen so that a single-photon process produces a  $\pi$ -phase shift on an  $|e\rangle \rightarrow |g\rangle$  transition during the atom-cavity interaction time (that is dictated by the velocity of the atom).

As shown by Lutterbach and Davidovich this phase shift can be determined (measured) by the Ramsey interferometry [43] by subjecting each atom to two resonant  $\pi/2$  pulses that mix  $|e\rangle$  and  $|g\rangle$  levels of the atom before and after the interaction with the cavity field. The probability  $p_e$  ( $p_g$ ) to detect the atom at the exit from the cavity in the state  $|e\rangle$  ( $|g\rangle$ ) exhibits modulations with respect to the phase  $\phi$  of the interferometer. For a cavity that has been prepared in the vacuum state  $|0\rangle$  and a proper choice of the phase reference the probability  $p_e$  oscillates as  $p_e(\phi) = (1 + \cos \phi)/2$ . The phase shift induced by the  $n$ -photon field is  $n\phi$ . Correspondingly, the probability to find an atom in the excited state  $|e\rangle$  when the cavity mode is displaced by  $\alpha$  reads

$$\begin{aligned} p_e(\phi, \alpha) &= \left[ 1 + \sum_{n=0}^{\infty} (-1)^n \rho_{n,n}(\alpha) \cos \phi \right] / 2 \\ &= \left[ 1 + \langle \hat{P} \rangle \cos \phi \right] / 2 . \end{aligned} \quad (6.77)$$

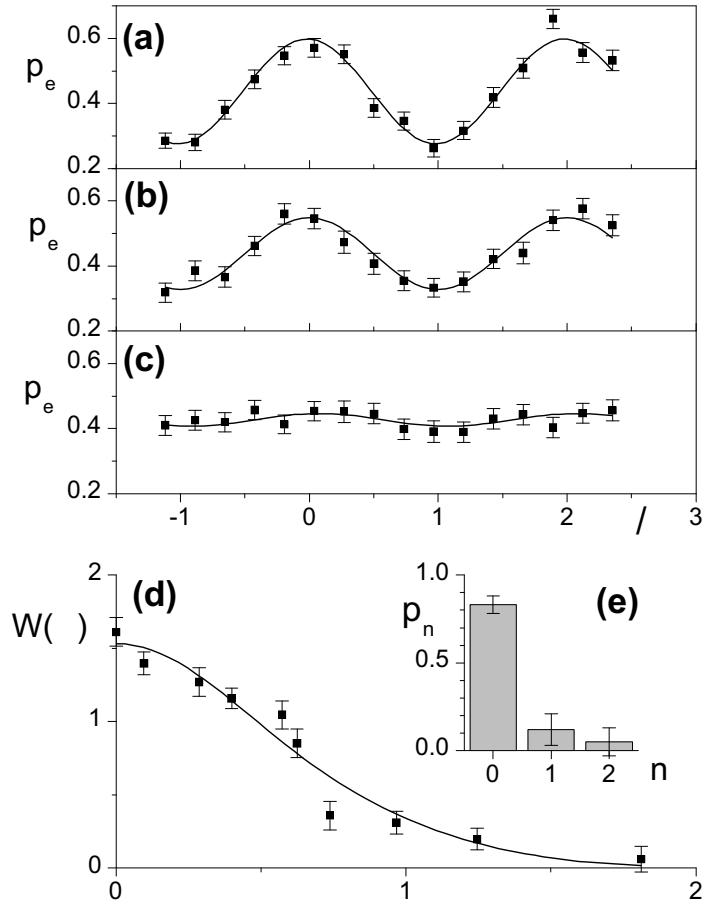
From above we see that the Wigner function at the phase-space point  $\alpha$  is directly related to the fringes contrast  $c(\alpha)$ :

$$\begin{aligned} W(\alpha) &= 2\langle \hat{P} \rangle = 2c(\alpha) \\ &= 2[p_e(0, \alpha) - p_e(\pi, \alpha)] . \end{aligned} \quad (6.78)$$

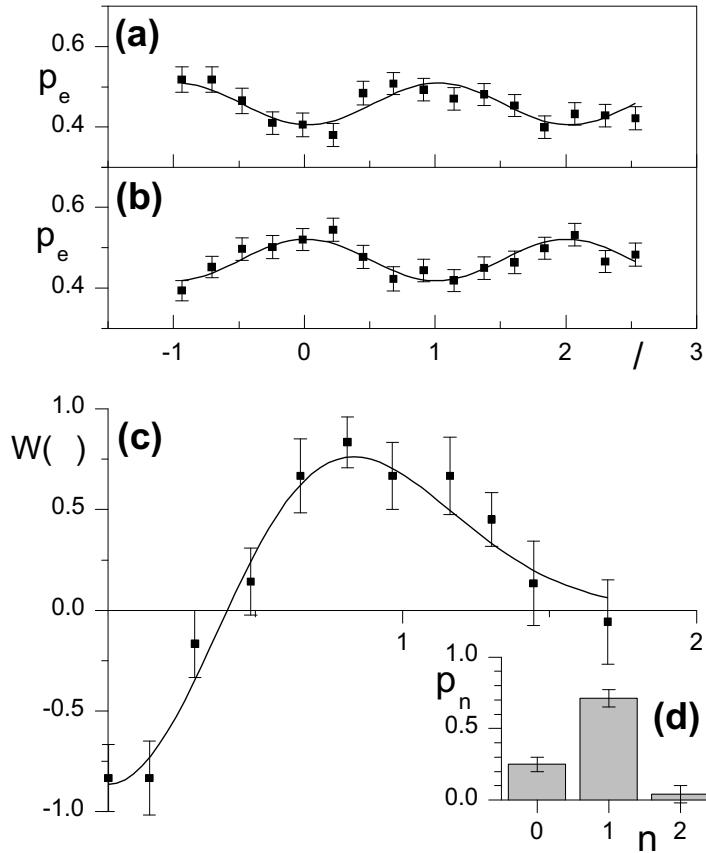
In other words the Wigner function of the cavity field is determined from the expectation values of a measurement performed on the atom after its interaction with the displaced cavity mode. The attractive feature of this procedure is that there is no need to perform any further inversion procedure to determine the value of the Wigner function at the given phase-space point. On the other hand, to determine the complete Wigner function one has to “scan” the whole phase space, i.e. one has to displace the original cavity mode with a continuum set of values of  $\alpha$ .

The group of Serge Haroche from the ENS in Paris has recently reported [44] experimental realization of the Lutherbach-Davidovich method. The experiment itself has been very challenging since a large dispersive phase shift per photon had to be achieved. In addition the measurement itself has to be performed in a time that is shorter than the field damping time. Technical details of the experiment and the description of the cavity QED setup can be found in Refs. [44] and [45], respectively. In what follows we will utilize the data obtained in the experiment. In particular, Haroche *et al.* have performed the direct measurement of the Wigner function for the vacuum state and the one-photon Fock state. In their experiment only a finite set of the values  $W(\alpha)$  of the Wigner function for *real non-negative*  $\alpha$  have been measured. In Fig. 6.11 data from the measurement of the Wigner function of a vacuum cavity field are presented. Analogously, in Fig. 6.12 data from the measurement of the Wigner function of a single-photon Fock state of the cavity field are shown.

Given the experimental data and *a priori* knowledge that the cavity mode is initially prepared in a phase-insensitive state, one might conclude that the experimental data presented in Fig. 6.11 and Fig. 6.12 correspond to the vacuum state and the one-photon Fock state of the cavity field. Nevertheless, a conservative approach is more appropriate, that is the reconstruction of the Wigner function has to be warranted only by experimental data (i.e., the mean values of the parity operator for given values of  $\alpha$ ) and no further assumptions should be made. Under this circumstance, we can utilize the reconstruction scheme based on the principle of Maximum Entropy as described in previous sections of this Chapter. Here the observables are the displaced parity operators  $\hat{D}(\alpha)\hat{P}\hat{D}^\dagger(\alpha)$  for various discrete values of real non-negative parameter  $\alpha$ . In order to use the MaxEnt reconstruction scheme as discussed in Section 6.3 we also have to know the mean-value of the photon number operator of the cavity field. Unfortunately, this operator has not been measured in the experiment, so this number is only estimated from the experimen-



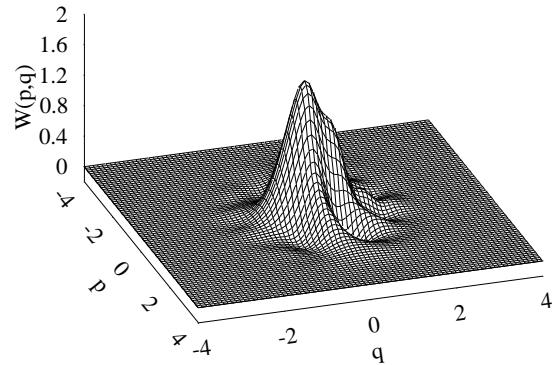
**Fig. 6.11.** Results of the measurement of a set of values  $W(\alpha)$  of the Wigner function of a single-mode cavity field. (a) Ramsey fringes for an injected amplitude  $\alpha = 0$ . Probability  $p_e$  for detecting the atom in the state  $|e\rangle$  as a function of the Ramsey interferometer phase  $\phi/\pi$ . Dots are experimental with error bars reflecting the variance of the binomial detection statistics. The solid curve is a sine fit. (b) and (c) Ramsey fringes for  $\alpha = 0.57$  and  $\alpha = 1.25$ , respectively. (d) Dots represent experimentally measured values of the Wigner function versus the phase-space parameter  $\alpha$  with error bars reflecting the uncertainty in the Ramsey fringes fit. The solid line represents a theoretical fit performed by Haroche *et al.*. (e) Corresponding photon number distribution. The figure is shown with the kind permission of J.M. Raimond.



**Fig. 6.12.** Results of the measurement of a set of values  $W(\alpha)$  of the Wigner function of a single-mode cavity field. Here a single-photon field has been initially prepared in the cavity. (a) Ramsey fringes for an injected amplitude  $\alpha = 0$ . Probability  $p_e$  for detecting the atom in the state  $|e\rangle$  as a function of the Ramsey interferometer phase  $\phi/\pi$ . Dots are experimental with error bars reflecting the variance of the binomial detection statistics. The solid curve is a sine fit. (b) Ramsey fringes for  $\alpha = 0.81$ . (c) Dots do represent experimentally measured set of values of the Wigner function versus the phase-space parameter  $\alpha$  with error bars reflecting the uncertainty on the Ramsey fringes fit. The solid line represents a theoretical fit performed by Haroche *et al.*. (d) Inferred photon number distribution. The figure is shown with the kind permission of J.M. Raimond.

tal data. With these inputs we can perform the reconstruction of Wigner function.

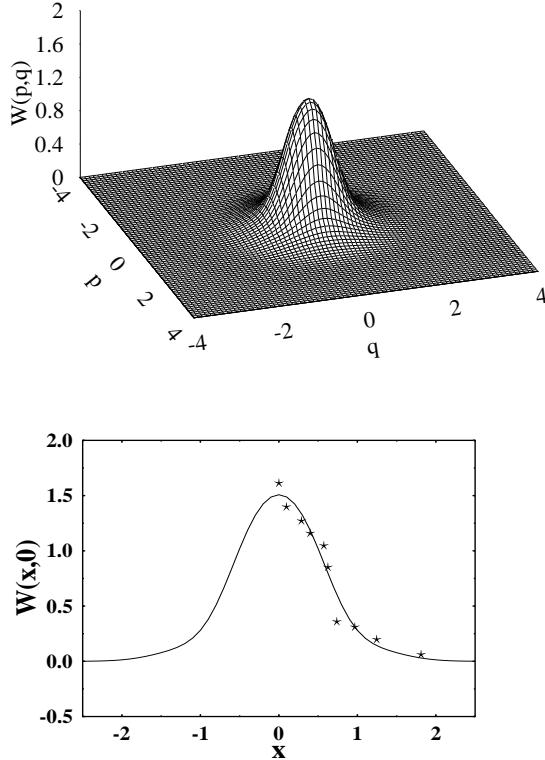
In Fig. 6.13 we present a reconstruction of the Wigner function of the cavity mode based on the experimental data presented in Fig. 6.11.



**Fig. 6.13.** The Wigner function for the vacuum-like state (with the mean photon number  $\bar{n} = 0.22$  determined from the measured data) reconstructed via the MaxEnt principle. The measured observable - the shifted parity operator - corresponds to the direct measurement of the Wigner function  $W(x_j, 0)$  at  $\{x_j, j = 1, \dots, 10\}$ . Here I denote  $x_j = \alpha$ .

The reconstructed Wigner function is assymetric since the data obtained from the measurement do not warrant any phase symmetry. The mean-photon number is not a direct observable, it is not a constraint in the MaxEnt reconstruction scheme. From the reconstructed Wigner function one obtains the mean-photon number  $\bar{n} \simeq 0.3$  and the photon number distribution  $P_n \simeq \{0.88, 0.07, 0.00, \dots\}$ .

In principle one can assume that the measured data are rotationally invariant. That is, if in the direction of other rotated quadrature the same mean values of the displaced parity operator are measured, then utilizing this additional (though not warranted by the measured data) information we can use the MaxEnt principle again and perform the reconstruction with these additional mean-values of the “measured” parity operators. In Fig. 6.14 we present the reconstruction of the Wigner function under the assumption that the same set of values of the Wigner functions has been measured for 16 values of angles of rotation. It is interesting to note, that from the reconstructed Wigner function one obtains the mean-photon number  $\bar{n} \simeq 0.285$  and the photon number distribution is equal to  $P_n \simeq \{0.835, 0.084, 0.042, 0.039, \dots\}$ . That is, from the MaxEnt reconstruction scheme under the given assumptions we obtain a non-zero probability of occupation of the Fock state  $|3\rangle$  that has not been considered by the authors of the experiment (see Ref. [44]). In the

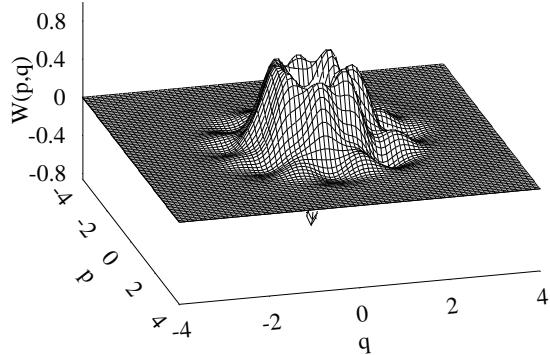


**Fig. 6.14.** The Wigner function for the vacuum-like state (with the “measured” mean value of the photon number  $\bar{n} = 0.22$ ) reconstructed via the MaxEnt principle. The measured observable - the shifted parity operator - corresponds to the direct measurement of the Wigner function  $W(x_j, 0)$  at  $\{x_j, j = 1, \dots, 10\}$ . We consider the same setting as in Fig. 6.13 but we perform a reconstruction under the assumption that the distribution  $W(x_j, \phi_n)$  has been measured for 16 values of the angles, i.e.  $\phi_n = n\pi/8$ . In the figure we also plot the distribution  $W(x, 0)$  that is obtained from the MaxEnt reconstruction scheme and we compare it with the experimental data (displayed as  $\star$ ).

figure we also plot the distribution  $W(x, 0)$  that is obtained from the MaxEnt reconstruction scheme and we compare it with the experimental data (displayed as  $\star$ ). We see that the MaxEnt scheme very nicely reconstructs the Wigner function of the cavity mode. This state is not exactly the vacuum, but rather a thermal mixture at a very low temperature.

Let us consider the reconstruction of the Wigner function based on the experimental data presented in Fig. 6.12.

This reconstructed Wigner function is assymmetric since the data obtained from the measurement do not warrant any phase symmetry. The mean-photon

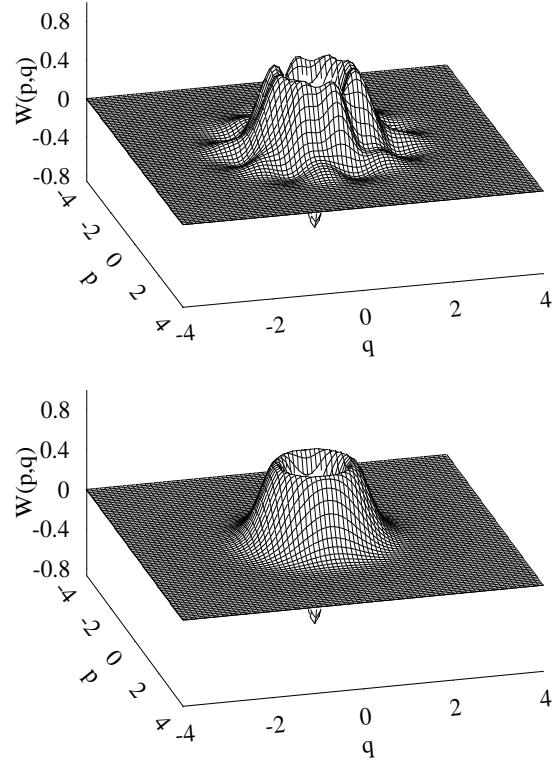


**Fig. 6.15.** The Wigner function for the single-photon-like state (with the “measured” meanvalue of the photon number  $\bar{n} = 0.79$ ) reconstructed via MaxEnt principle. The measured observable - the shifted parity operator - corresponds to the direct measurement of the Wigner function  $W(x_j, 0)$  at  $\{x_j, j = 1, \dots, 10\}$ .

number is not a direct observable, it is not a constraint in the MaxEnt reconstruction scheme. From the reconstructed Wigner function one obtains the mean-photon number  $\bar{n} \simeq 0.82$  and the photon number distribution  $P_n \simeq \{0.28, 0.70, 0.00, \dots\}$ .

We can again make an assumption about the fact that the displaced parity operator is measured for different values of the phase-space rotations (i.e. down different rotated quadratures). In this case we can extend the observation level and perform a new reconstruction via the MaxEnt principle. In Fig. 6.16 we assume that the displaced parity operator has been measured for 20 values of phases. That is, we consider 20 rotations of the original distribution  $W(x_j, 0)$ . For pedagogical purposes we will consider two situations: first we consider these 20 angles to be chosen randomly. In the second case the angles are regular, with  $\phi_n = n\pi/10$ .

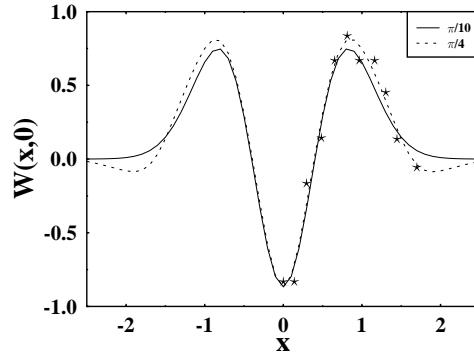
Intuitively it is clear that once we extend the observation level, i.e. we increase the number of angles for which the displaced parity operator is measured, we should obtain a better reconstruction. This is nicely illustrated in the lower picture in Fig. 6.16 that really reminds the Wigner function of the single-photon Fock state. The difference between the upper and lower figures is only in the distribution of phases. In the upper case phases are random in the lower case phases are regular. So the question is why there is such a difference between the two figures that has been obtained from the same experimental data and using the same reconstruction scheme. The reason is purely numerical. In the case of random phases the dimension of the Hilbert space that is used for the MaxEnt reconstruction has to be larger. The oscillations of the reconstructed Wigner function in the top figure have their origin in a wrong assumption about the dimension of the Hilbert space that has to be considered in the reconstruction (see discussion in Section 6.4).



**Fig. 6.16.** The Wigner function for the one-photon Fock state (with the “measured” mean value of the photon number  $\bar{n} = 0.79$ ) reconstructed via the MaxEnt principle. The measured observable - the shifted parity operator - corresponds to the direct measurement of the Wigner function  $W(x_j, 0)$  at  $\{x_j, j = 1, \dots, 10\}$ . The reconstruction of the Wigner function is performed under the assumption that the distribution  $W(x_j, \phi_n)$  has been measured for 20 values of the phase, while  $W(x_j, \phi_m) = W(x_j, \phi_n)$ . The upper figure corresponds to the situation when the angles have been chosen randomly, the bottom figure corresponds to the regular case when  $\phi_n = n\pi/10$ .

This example illustrates that one has to be very careful with handling the data.

Finally, in Fig. 6.17 we present a comparison between the reconstructed distribution  $W(x, 0)$  and measured values of  $W(x_j, 0)$  obtained in the measurement.



**Fig. 6.17.** The distribution  $W(x, 0)$  that is obtained from the MaxEnt reconstruction scheme of the Fock state  $|1\rangle$  under two assumptions. The number of angles for which the distribution  $W(x_j, \phi_n)$  is 8 (dotted line) and 20 (solid line). We compare these distributions with the experimental data (displayed as  $\star$ ).

## 6.7 Conclusions

We have shown that the reconstruction of Wigner functions from incomplete tomographic data can be very reliably performed with the help of the Jaynes principle of the Maximum Entropy. He have presented a generalized canonical density operator that is suitable for the incomplete tomographic data. We have implemented a numerical procedure with the help of which the *MaxEnt* reconstruction can be performed. We have compared the *MaxEnt* scheme with the sampling via pattern functions. The comparison is very clear - the *MaxEnt* approach is much more efficient. It requires less data, it gives the reconstruction with much higher fidelity and is more stable with respect to the choice of parameters such as the size of the quadrature bin, or the interval on which the quadrature distributions are measured. Our empirical experience shows that three rotated quadratures are always sufficient to perform a very reliable *MaxEnt* reconstruction of an arbitrary unknown state.

In this Chapter the method has been demonstrated on two sets of experimental data. Firstly, we have reconstructed vibrational states of neutral atoms. Secondly, we have performed a reconstruction of a single-mode electromagnetic field inside a high- $Q$  cavity. In both these examples the efficiency of the proposed method is clearly seen. Recently, our method based on the Jayne's principe of maximum entropy has been applied (see Ref. [46]) for the tomographic reconstruction of a complete internuclear quantum state, represented by the Wigner function, of a dissociating I-2 molecule based on femtosecond time resolved position and momentum distributions of the atomic fragments. These examples illustrates versatility of the MaxEnt reconstruc-

tion scheme and its efficiency in situations when mean values of a restricted set of observables have been measured.

#### Acknowledgements

I would like to thank Gabriel Drobny for a longstanding collaboration and numerous discussions devoted to technical as well as conceptual problems related to reconstruction of states of quantum systems. I also thank Christophe Salomon and Isabelle Boucouse (experiment with neutral atoms) and Michel Brune and Jean-Michel Raimond (cavity QED experiment) for providing me with their experimental data and for helpful discussions and correspondence. This work was supported by the European Union projects QGATES and (CON)QUEST.

## References

1. E.T. Jaynes: Information theory and statistical mechanics. In: *1962 Brandeis Lectures*, vol 3, ed by K.W. Ford (Benjamin, Elmsford, New York 1963) p 181
2. A. Peres: *Quantum Theory: Concepts and Methods* (Kluwer Academic Publishers, Dordrecht, 1995)
3. R. Omnès: *The Interpretation of Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1994)
4. L.E. Ballentine: *Quantum Mechanics* (Prentice Hall, Englewood Cliffs, New Jersey, 1990)
5. V. Bužek and R. Derka: *Quantum observations*. In *Coherence and Statistics of Photons and Atoms*, ed by J. Peřina (John Wiley & Sons, New York, 2001) pp 198 - 261
6. E.P. Wigner: Phys. Rev. **40**, 749 (1932)
7. M. Hillery, R.F. O'Connell, M.O. Scully, and E.P. Wigner: Phys. Rep. **106**, 121 (1984)
8. V. Bužek and P.L. Knight: Quantum interference, superposition states of light and nonclassical effects. In *Progress in Optics*, vol 34, ed by E. Wolf (North Holland, Amsterdam, 1995) p.1
9. A.K. Ekert and P.L. Knight: Phys. Rev. A **43**, 3934 (1991)
10. K. Vogel and H. Risken: Phys. Rev. A **40**, 2847 (1989)
11. U. Leonhardt: *Measuring the quantum state of light*. (Cambridge University Press, Cambridge, 1997)
12. D.-G. Welsch, W. Vogel, and T. Opatrný: Homodyne detection and quantum state reconstruction. In *Progress in Optics*, vol 39, ed by E. Wolf (North Holland, Amsterdam, 1999) p.63
13. G.M. D'Ariano, C. Machiavelo, and M.G.A. Paris: Phys. Rev. A **50**, 4298 (1994)
14. H. Kühn, D.-G. Welsch, and W. Vogel: J. Mod. Opt. **41**, 1607 (1994)
15. D.T. Smithey, M. Beck, M.G. Raymer, and A. Faridani: Phys. Rev. Lett. **70**, 1244 (1993);  
D.T. Smithey, M. Beck, J. Cooper, and M.G. Raymer: Phys. Scr. T **48**, 35 (1993);  
M. Beck, M.G. Raymer, I.A. Walmsley, and V. Wong: Opt. Lett. **18**, 2041 (1993).

- M. Beck, D.T. Smithey, and M.G. Raymer: Phys. Rev. A **48**, R890 (1993);  
 M.G. Raymer, M. Beck, and D.F. Mc Alister: Phys. Rev. Lett. **72**, 1137 (1994);  
 M.G. Raymer, D.T. Smithey, M. Beck, and J. Cooper: Acta Phys. Pol. **86**, 71 (1994)
- 16. M. Munroe, D. Boggavarapu, M.E. Anderson, and M.G. Raymer: Phys. Rev. A **52**, R924 (1995)
  - 17. U. Janicke and M. Wilkens: J. Mod. Opt. **42**, 2183 (1995)
  - 18. S. Schiller, G. Breitenbach, S.F. Pereira, T. Müller, and J. Mlynek: Phys. Rev. Lett. **77**, 2933 (1996);  
 Ch. Kurtsiefer, T. Pfau, and J. Mlynek: Nature **386**, 150 (1997).
  - 19. S. Wallentowitz and W. Vogel: Phys. Rev. Lett. **75**, 2932 (1995);  
 J.F. Poyatos, R. Walser, J.I. Cirac, P. Zoller, and R. Blatt: Phys. Rev. A **53**, R1966 (1996);  
 C.D'Helon and G.J. Milburn: Phys. Rev. A **53**, R25 (1996)
  - 20. D. Leibfried, D.M. Meekhof, B.E. King, C. Monroe, W.M. Itano, and D.J. Wineland: Phys. Rev. Lett. **77**, 4281 (1996);  
 D. Leibfried, D.M. Meekhof, C. Monroe, B.E. King, W.M. Itano, and D.J. Wineland: J. Mod. Opt. **44**, 2485 (1997);  
 D. Wineland, C. Monroe, D.M. Meekhof, B.E. King, D. Liebfried, W.M. Itano, J.C. Bergquist, D. Berkeland, J.J. Bollinger, and J. Miller: Proc. Roy. Soc. A **454**, 411 (1998)
  - 21. T.J. Dunn, I.A. Walmsley, and S. Mukamel: Phys. Rev. Lett. **74**, 884 (1995)
  - 22. U. Leonhardt, H. Paul, and G.M. D'Ariano: Phys. Rev. A **52**, 4899 (1995)
  - 23. Th. Richter: Phys. Lett. A **211**, 327 (1996);  
 Th. Richter: Phys. Rev. A **53**, 1197 (1996)
  - 24. U. Leonhardt, M. Munroe, T. Kiss, T. Richter, and M.G. Raymer: Opt. Commun. **127**, 144 (1996)
  - 25. G.M. D'Ariano, U. Leonhardt, and H. Paul: Phys. Rev. A **52**, R1801 (1995)
  - 26. J. von Neumann: *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955)
  - 27. A. Wehr: Rev. Mod. Phys. **50**, 221 (1978)
  - 28. E. Fick and G. Sauermann: *The Quantum Statistics of Dynamic Processes*. (Springer Verlag, Berlin, 1990)
  - 29. V. Bužek, G. Adam, and G. Drobný: Ann. Phys. (N.Y.) **245**, 37 (1996)
  - 30. E.T. Jaynes: Phys. Rev. **108**, 171 (1957); *ibid.* **108**, 620 (1957) 620; Am. J. Phys. **31**, 66 (1963)
  - 31. J.N. Kapur and H.K. Kesavan: *Entropic Optimization Principles with Applications* (Academic Press, New York, 1992)
  - 32. A. Katz: *Principles of Statistical Mechanics* (W.H. Freeman and Company, San Francisco, 1967)
  - 33. A. Hobson: *Concepts in Statistical Mechanics* (Gordon Breach Science Publishers, New York, 1971)
  - 34. H. Jeffreys: *Theory of Probability*, 3rd ed (Oxford Univ. Press, Oxford, 1960);  
 R. von Mises: *Mathematical Theory of Probability and Statistics* (Academy Press, New York, 1964)
  - 35. U. Leonhardt and M. Munroe, Phys. Rev. A **54**, 3682 (1996)
  - 36. I. Bouchoule, H. Perrin, A. Kuhn, M. Morinaga, and C. Salomon: Phys. Rev. A **59**, R8 (1999)
  - 37. M. Morinaga, I. Bouchoule, J.-C. Karam, and C. Salomon: Phys. Rev. Lett. **83**, 4037 (1999)

38. We should stress that due to random fluctuations of “observed” mean values (6.73) physically incompatible data might be obtained. It means that there does not exist a physical density operator which could fit perfectly the results of the noisy measurement (i.e. leading to  $\Delta F \rightarrow 0$ ). Obviously increasing the relative error of the measurement we increase also a number of incompatible results and the reconstruction becomes meaningless.
39. C. Salomon and I. Bouchoule: private communication
40. B.-G. Englert, N. Sterpi, and H. Walther: Opt. Commun. **100**, 526 (1993)
41. L.G. Lutterbach and L. Davidovich: Phys. Rev. Lett. **78**, 2547 (1997)
42. C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg: *Atom-Photon Interactions* (John Wiley & Sons, New York, 1992)
43. N.F. Ramsey: *Molecular Beams* (Oxford University Press, New York, 1985)
44. P. Bertet, A. Auffeves, P. Maioli, S. Osnaghi, T. Meunier, M. Brune, J.M. Raimond, and S. Haroche: Phys. Rev. Lett. **89**, 200402 (2002)
45. J.M. Raimond, M. Brune, and S. Haroche: Rev. Mod. Phys. **73**, 565 (2001)
46. E. Skovsen, H. Stapelfeldt, S. Juhl, and K. Molmer: Phys. Rev. Lett. **91**, 090406 (2003)



# 7 Experimental Quantum State Measurement of Optical Fields and Ultrafast Statistical Sampling

Michael G. Raymer<sup>1</sup> and Mark Beck<sup>2</sup>

<sup>1</sup> Department of Physics, University of Oregon, Eugene, OR 97403  
[raymer@uoregon.edu](mailto:raymer@uoregon.edu)

<sup>2</sup> Department of Physics, Whitman College, Walla Walla, WA 99362  
[beckmk@whitman.edu](mailto:beckmk@whitman.edu)

## 7.1 Introduction

How can the quantum state of a physical system be completely determined by measurements? Before answering, it is useful to define what is ordinarily meant by a quantum state. The standard interpretation, well stated by Leonhardt, is that “knowing the state means knowing the maximally available statistical information about all physical quantities of a physical object” [1]. Typically by “maximally available statistical information” we mean probability distributions. Hence, knowing the state of a system means knowing the probability distributions corresponding to measurements of any possible observable pertaining to that system. For multiparticle (or multimode) systems this means knowing joint probability distributions corresponding to joint measurement of multiple particles (or modes). Quantum mechanics is a theory of information, and the state of a system is a convenient means of describing the statistical information about that system. Interestingly, it has been shown that any extension of quantum theory in which the state (density matrix) does not contain all statistical information requires that an isolated system could receive information via EPR (Einstein-Podolski-Rosen) correlations, an effect which is widely viewed as nonphysical [2].

Since knowing the state means knowing all the statistical information about a system, is the inverse true? If one knows all of the statistical information about a system, does one then know (or can one infer) the quantum state of that system? Clearly a real experiment cannot measure all possible statistical information. A more practical question is then: can one infer reasonably well the quantum state of a system by measuring statistical information corresponding to a finite number of observables?

The answer to this question is a resounding yes, but there are some caveats. By now it is well established that the state of an individual system cannot, even in principle, be measured [3, 4]. This is easily seen by the fact that a single measurement of some observable yields a single value, corresponding to a projection of the original state onto an eigenstate that has nonzero probability. Clearly this does not reveal much information about the

original state. This same measurement simultaneously disturbs the individual system being measured, so that it is no longer in the same state after the measurement. This means that subsequent measurements of this same system are no longer helpful in determining the original state. Since state measurement requires statistical information, multiple measurements are needed, each of which disturbs the system being measured. And attempting to use “weaker” measurements also does not help in this regard [5–7].

However, the state of an ensemble of identically prepared systems can be measured [3, 4]. Here each member of an ensemble of systems is prepared by the same state-preparation procedure. Each member is measured only once, and then discarded (or possibly recycled). Thus, multiple measurements can be performed on systems all in the same state, without worrying about the measurement apparatus disturbing the system. A mathematical transformation, of which there are several (see Refs. [1, 8–10] and the rest of this volume,) is then applied to the data in order to reconstruct, or infer, the state. The relevant interpretation of the measured state in this case is that it is the state of the ensemble. (Although we believe that any single system subsequently prepared by this same procedure will be described by this state as well.)

How many different observables need to be statistically characterized for an accurate reconstruction of the state? This is an interesting, and not easily answered, question. In general, the more complex the system, the more observables that are required. The minimum number is two; at least two non-commuting observables must be measured (see Royer [3] or Ballentine [4] and the references therein.) This is related to the “Pauli Question.” Pauli asked in 1933 whether or not the wave function of a particle could be uniquely determined by measurements of its position and momentum distributions [11]. In three dimensions the answer turns out to be no. As pointed out by Gale et al., the probability distributions obtained from the position and momentum wave functions for the hydrogen atom  $\Psi(r, \theta, \varphi) = R_{n,l}(r)P_l^m(\theta)e^{im\varphi}$  are independent of the sign of  $m$  [12]. However, note that the only difference between the wave functions with different signs of  $m$  is that they are complex conjugates of each other. For pure states in one dimension, knowledge of the position and momentum distributions is sufficient to reconstruct the wave function up to a potential ambiguity in the sign of the complex phase function [13–15]. However, in the case of mixed states, even for one-dimensional systems, more than two distributions are needed for state reconstruction.

Because some of the first experimental work on quantum state measurement [16] analyzed the collected data using a mathematical technique that is very similar to the tomographic reconstruction technique used in medical imaging, and because all techniques are necessarily indirect, a generally accepted term for quantum state measurement has become quantum state tomography (QST.) Systems measured have included angular momentum states of electrons [17], the field [16, 18–20] and polarization [21] states of photon pairs, molecular vibrations [22], trapped ions [23], atomic beams [24], and

nuclear spins [25]. Indeed, QST has become so prevalent in modern physics that it has been given its own Physics and Astronomy Classification Scheme (PACS) code by the American Institute of Physics: 03.65.Wj, State Reconstruction and Quantum Tomography.

The purpose of this article is to review some of the theoretical and experimental work on quantum state measurement. Prior reviews can be found in [1, 8, 9, 26]. We will concentrate here on measurements of the state of the field of an optical beam with an indeterminate number of photons in one or more modes. Measurement of the polarization state of beams with fixed photon number is discussed in the article by Altepeter, James and Kwiat in this volume. Nearly all field-state measurements are based on the technique of balanced homodyne detection, and hence fall under the category of optical homodyne tomography (OHT.) This technique was first suggested by Vogel and Risken [10] and first demonstrated by Smithey et al. [16]. A reasonably complete list of experiments that have measured the quantum states (or related quantities) of optical fields is presented in Table 1.

Here we will describe balanced homodyne detection, and its application to OHT. We emphasize pulsed, balanced homodyne detection at zero frequency (DC) in order to model experiments on “whole-pulse” detection. Such detection provides information on, for example, the total number of photons in a particular spatial-temporal mode [27]. (This differs from the other commonly used technique of radio frequency (RF) spectral analysis of the photocurrent [28-30].) Discussion of theoretical issues will be given in Secs. 7.2 and 7.3, while experimental issues will be discussed in Sec. 7.4. We also discuss applications of state measurement techniques to an area of scientific and technological importance—the ultrafast sampling of time-resolved photon statistics. This new technique is called linear optical sampling.

Recently, array detectors have been used to perform QST [31–33]. One motivation for using arrays is to increase the effective detection efficiency, and hence the fidelity of the state reconstruction. Furthermore arrays allow measurements of the states of multiple modes, spatial or temporal, of the same beam. In Sec. 7.5 we review the use of array detection in QST.

### 7.1.1 Balanced Homodyne Detection and Optical Quantum State Tomography

Since the development of balanced optical homodyne detection (BHD) by Yuen and Chan in 1983 [58, 59], it has been used widely in both continuous-wave [28, 29] and pulsed-laser [60] applications for characterizing the quadrature-amplitude fluctuations of weak optical fields. By quadrature amplitudes one means the cosine and sine, or real and imaginary, components of the oscillating electric field associated with a particular spatial-temporal mode. Schematically we can write the (real) field as

**Table 7.1.** Experiments measuring quantum states and other related quantities.

Ref.	Measured Quantities	number of modes	Description
[16]	$W$	1	first implementation of OHT, squeezed vacuum
[34]	$W, \Pr(\phi)$	1	quantum phase distributions
[35]	$W, \Psi, \langle x \hat{\rho} x'\rangle, \Pr(\phi)$	1	first quantum wave function (vacuum)
[36]	<i>moments</i>	1	number-phase uncertainty relations
[37]	$W, \Psi, \Pr(n), \Pr(\phi)$	1	number-phase uncertainty relations, wave function of coherent state
[38]	<i>moments</i>	1	time resolved measurement
[39]	$\Pr(n)$	1	quantum sampling, time resolved measurement
[40]	$W$	1	first OHT with RF detection, <u>squeezed</u> vacuum
[41]	<i>moments</i>	1	time & frequency resolution
[42]	$\langle n \hat{\rho} m\rangle$	1	even-odd oscillations of $P(n)$
[43]	$g^{(2)}(t, t + \tau)$	2	first two-mode measurement
[48]	$W, \langle n \hat{\rho} m\rangle$	1	squeezed coherent state
[44]	$W$	1	modulated signal
[45]	$\Pr(n)$	1	self-homodyne tomography
[46]	$W, \Pr(n)$	1	squeezed light from cold atoms
[19]	$W$	1	first photon counting measurement
[47]	$\Pr(n, m)$	2	joint photon number distributions
[48]	$\Pr(n)$	1	application to telecommunications
[49]	$W$	1	self-homodyne tomography
[31]	$\langle n \hat{\rho} m\rangle$	2	first array detection experiment
[20]	$W, \Pr(n)$	1	one-photon state
[50]	<i>moments</i>	2	time-resolved polarization correlations
[51]	$W, \langle n \hat{\rho} m\rangle$	1	high repetition rate
[33]	<i>Q-moments</i>	many	Q-function, unbalanced array detection
[52]	$W, \Pr(n)$	1	very high repetition rate
[53]	$\langle n \hat{\rho} m\rangle W$ ,	1	displaced one-photon state
[54]	$W, \Pr(n)$	1	displaced one-photon state
[55]	$W$	1	self-homodyne tomography, amplitude squeezing
[56]	$W, \Pr(n)$	1	microwave one-photon state, Rydberg atoms
[57]	$\Pr(n, m)$	2	universal homodyne tomography
[32]	$\langle n \hat{\rho} m\rangle$	2	mode-optimized array detection

$$\begin{aligned} E &= E_{VAC} [q \cos(\omega t + \theta) + p \sin(\omega t + \theta)] \\ &= E_{VAC} \operatorname{Re} [(q + ip) \exp(-i\omega t - i\theta)], \end{aligned} \quad (7.1)$$

where  $E_{VAC}$  is an electric field corresponding to the vacuum (zero-point) strength of a single mode, and  $q, p$  are the dimensionless quadrature amplitudes associated with the carrier frequency  $\omega$  and reference phase  $\theta$ . The goal of BHD is to allow a quantum-limited measurement of these field amplitudes. This is a quite different strategy than the more common intensity or photon-flux detection that is accomplished by *directly* illuminating a photoemissive detector (such as a photodiode or a photomultiplier) with the light beam to be measured.

The idea of indirect quantum measurements is an outgrowth of the development of quantum-state reconstruction [39, 41, 61, 62]. Rather than targeting the quantum state itself as the object of measurement, one can target various statistical averages (moments) of the optical field, such as mean field, mean intensity, variance of field, or variance of intensity. When considering time-nonstationary or multimode fields, one can target various time-dependent correlation functions, which often are the quantities of greatest interest in quantum optics. Using BHD allows one to measure one type of quantity (field quadratures) and from their ensemble measurements infer the averages of other quantities. This offers a new way to view quantum measurements in general [1, 8].

There have been many theoretical discussions of balanced homodyne detection [63-77]. The measurement of the full probability distribution of the quadrature amplitudes was demonstrated in 1993 [16, 35], generalizing earlier experiments which measured only the variances of the quadrature amplitudes. The distributions were measured using pulsed, balanced homodyne detection with integrating (DC) photodetectors, a technique that allows the statistical characterization of weak, repetitive optical fields on ultrafast time scales [38, 39]. This technique detects optical field amplitudes in a particular spatial-temporal mode that is defined by a coherent local-oscillator (LO) pulse. Ideally the LO pulse should have a known spatial-temporal shape and should be phase-locked to the signal field, although useful information can be obtained even when there is no stable phase relation between LO and signal [39].

Combined with the data processing method of tomographically reconstructing the Wigner quasi-distribution for the two quadrature-field amplitudes, the method is called optical homodyne tomography [16, 35]. Alternatively one can reconstruct the density matrix directly from the measured quadrature distributions, as first pointed out by D'Ariano et al. [61] and Kuhn et al. [78].

The Wigner distribution  $W(q, p)$  is related to the density matrix in the quadrature or number representations by [79]

$$\begin{aligned} W(q, p) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \langle q + \frac{1}{2}q' | \hat{\rho}| q - \frac{1}{2}q' \rangle e^{-ipq'} dq' \\ &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \langle n | \hat{\rho} | m \rangle W_{nm}(q, p) \end{aligned} . \quad (7.2)$$

To obtain the form of the functions  $W_{nm}(q, p)$ , insert  $\hat{\rho} = \sum_{n,m} \rho_{nm} |n\rangle \langle m|$  into the above, giving

$$W_{nm}(q, p) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_n(q + \frac{1}{2}q') \psi_m(q - \frac{1}{2}q') e^{-ipq'} dq', \quad (7.3)$$

where  $\Psi_n(q) = \langle q | n \rangle$  are the real wave functions of the number states. These are well known to be Hermite-Gaussian functions for a harmonic oscillator, such as an optical mode. The explicit forms for  $W_{nm}(q, p)$  are Gaussian-Laguerre functions [1], pg. 129.

Clearly, knowing either the Wigner distribution or the density matrix in any representation is equivalent to knowing the quantum state of the ensemble of identically prepared systems (having a single degree of freedom) that have undergone measurement.

The Wigner distribution plays a natural role in the quantum theory of homodyne detection, since its marginal distributions give directly the measured quadrature-amplitude statistics [10] [see Eq. (7.33) below.] Because of this property, the Wigner distribution can be determined mathematically by tomographic inversion of the measured photoelectron distributions. In the case of a classical-like field (as from an ideal laser) the Wigner distribution can be interpreted simply as the joint distribution for the field quadrature amplitudes. In the case of a quantized field, it is uniquely related to the quantum state (density matrix or wave function) of a spatial-temporal mode of the signal field.

Mathematically, the quadrature probability distributions for different reference phases  $\theta$  are given by the density matrix in the number (Fock) basis according to [1], p.118

$$\Pr(q, \theta) = \sum_{\mu, \nu} \rho_{\mu\nu} \psi_{\mu}(q) \psi_{\nu}(q) \exp[i(\nu - \mu)\theta]. \quad (7.4)$$

To understand this formula intuitively, note that the phase  $\theta$  serves a role similar to a time variable in the Schrödinger time evolution of an energy eigenstate of a harmonic oscillator:

$$\psi_n(q, t) = \psi_n(q, 0) \exp[-in\omega t] \leftrightarrow \psi_n(q, \theta) = \psi_n(q, 0) \exp[-in\theta]. \quad (7.5)$$

Then using

$$\langle q, \theta | n \rangle = \Psi_n(q, \theta) = \Psi_n(q) \exp(-in\theta)$$

in  $\Pr(q, \theta) = \text{Tr}[\hat{\rho}|q, -\theta\rangle \langle q, -\theta|]$  gives Eq. (7.4).

The idea behind quantum state tomography for a harmonic oscillator (or optical homodyne tomography for a single field mode) is that complete knowledge of  $\text{Pr}(q, \theta)$ , through an ensemble of measurements on identically prepared systems, allows an accurate determination of the density matrix (or equivalently the Wigner distribution) – that is, the quantum state [1, 10, 16, 80]. This idea has also been extended to apply to systems other than harmonic oscillators [22, 81, 82].

The inversion of the raw measured data to arrive at a believable form of the quantum state is a delicate and interesting procedure. The inversion methods fall into two broad classes—deterministic and nondeterministic. In the deterministic methods an experimentally determined  $\text{Pr}(q, \theta)$  is used to determine, for example,  $\rho_{nm}$ , by a direct mathematical inversion of Eq. (7.4) [or Eq. (7.39) below.] But this approach begs the question, “How well can we determine or, more properly, infer the probability densities  $\text{Pr}(q, \theta)$  from a set of measurements of the quantities  $q$  for many values of  $\theta$ ? ” This is essentially a problem of classical statistics, and the fact that its answer plays a critical role in quantum state estimation is intriguing. This would seem to provide a link between the micro-quantum world and the macro-classical world, as emphasized in early discussions, notably by Niels Bohr. For a recent discussion see Caves et al. [83].

The nondeterministic approach recognizes this question from the beginning and targets the estimation of the quantum state directly, rather than the classical distributions as intermediate objects. A broad perspective on this approach has recently been given by Schack, Brun, and Caves [84]. These authors (following K.R.W. Jones before them [85, 86] consider a number  $N + M$  of like systems, all prepared by the same procedure, one at a time. They describe the situation this way:

“Suppose one is given a (prior) state  $\hat{\rho}^{(N+M)}$  on [Hilbert space]  $H^{(N+M)}$  and the results of measurements on  $M$  subsystems. The task is to find the (posterior) state of the remaining  $N$  subsystems conditioned on the measurement results.” [84]

Because all systems are prepared by the same (but unknown to us) procedure, the density operator of the entire ensemble prior to the measurements is entangled, and can be represented uniquely in the form

$$\hat{\rho}^{(N+M)} = \int d\hat{\rho} p(\hat{\rho}) \hat{\rho}^{\otimes(N+M)}, \quad (7.6)$$

where  $p(\hat{\rho})$  is the “probability” that all systems are prepared in the (common) state  $\hat{\rho}$ , and the functional integral is over all possible prior density operators (which can be parameterized by a set of density-matrix elements). Consider a measurement on the first subsystem that yields a result “ $k$ ,” with probability  $p_k$ . Schack et al. show that following such a measurement; the state of the remaining  $N + M - 1$  subsystems is conditioned to

$$\hat{\rho}^{(N+M-1)} = \int d\hat{\rho} p(\hat{\rho}|k) \hat{\rho}^{\otimes(N+M-1)}, \quad (7.7)$$

where

$$p(\hat{\rho}|k) = \frac{p(k|\hat{\rho})}{p_k}, \quad (7.8)$$

with  $p(k|\hat{\rho})$  being the conditional probability to observe  $k$  if the state is  $\hat{\rho}$ . Equation (7.8) has precisely the form of the Bayes Rule in classical probability or inference theory [87]. For this reason Schack et al. call this the Quantum Bayes Rule. The strategy in using this theorem for quantum state estimation is to iterate the procedure, updating the state after each measurement on a different system (which is then discarded), until only  $N$  systems remain. To continue their description:

"If the measurements on individual subsystems correspond to an informationally complete POVM (positive-operator-valued measure) or if they contain sequences of measurements of a tomographically complete set of observables [88], the posterior probability on density operators approaches a delta function in the limit of many measurements. This is the case of quantum state tomography [16], which can thus be viewed as a special case of quantum Bayesian inference. In this limit, the exact form of the prior probability on density operators becomes irrelevant." [84]

This result provides a satisfying link between the deterministic and non-deterministic methods for QST. It leads to the conclusion that measurements of quadrature probability densities for many (ideally all) phase values is a well-founded scheme for accurately inferring the state of a large ensemble of identical and similarly prepared single-mode EM fields.

Nevertheless, there still remains an issue about the most effective strategy for state reconstruction when the number of measurements is not large enough for the limit suggested above to be valid. In particular, it has been observed that when the sample size is small, so that statistical uncertainties prevent obtaining highly accurate values for  $\text{Pr}(q, \theta)$ , then the density matrix reconstructed using the deterministic inversion method can have nonphysical properties—for example, negative diagonal values (or probabilities). One satisfactory way to "cover up" these negative values is to calculate statistical error bars for each element of the density matrix [89]. Other methods rely on least-squares estimation [90], maximum-entropy estimation [91] (see the chapter by Bužek in this volume), or maximum-likelihood estimation of the density matrix elements [92]. We will not present a review of all of these methods here. For reviews see [1, 8, 9, 93] as well as other chapters in this volume.

### 7.1.2 Why Measure the State?

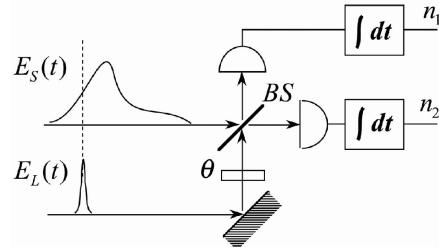
Aside from the question of how to measure the state of a quantum system, it is useful to ask why one would *want* to measure the state of a system. Of what use is knowing the state? An answer to this question is that once the state is obtained, distributions or moments of quantities can be calculated,

even though they have not been directly measured. Indeed, one can calculate moments of quantities that do not correspond to Hermitian operators, and thus cannot even in principle be directly measured.

For example, some early measurements demonstrated the ability to measure photon number distributions at the single-photon level, even though the detectors used had noise levels that were too large to directly measure these distributions [37]. Distributions of optical phase have been measured, even though there is no known experimental apparatus capable of directly measuring this phase directly [34, 37]. Expectation values of the number-phase commutator  $[n, \phi]$  have been measured, even though this operator is not Hermitian, and thus cannot be directly observed [37].

Furthermore, even if the full quantum state is not measured, the same basic idea of performing many measurements corresponding to different observables can yield important information about optical fields. For example, one can obtain information, with high time-resolution (on the time scale of 10's of fs), on the photon statistics of light propagation in scattering media [38], light emitted by edge-emitting lasers [39], or vertical-cavity surface-emitting lasers [50] and the performance optical communication systems [48, 94]. We will discuss some of this work here.

## 7.2 Balanced Homodyne Detection of Temporal-Spatial Modes



**Fig. 7.1.** BHD setup. A short local oscillator (LO) pulse  $E_L$ , after undergoing a phase shift  $\theta$ , interferes with a signal pulse  $E_S$  on a 50/50 beam splitter (or equivalent). The interfered pulses are detected by photodiodes and integrated to produce photoelectron numbers  $n_1$  and  $n_2$ .

Balanced homodyne detection, shown in Fig. 7.1, has the advantage that it rejects classical intensity fluctuations of the local oscillator (LO) field while measuring the signal quadrature-field amplitude in a particular spatial-temporal mode. Such a mode is defined by the space-time form of the LO pulse [64]. This provides a powerful technique for ultrafast time-gated detection (of field rather than intensity), recently developed into a practical scheme

called *Linear Optical Sampling* [38,41,94]. The time resolution can be as short as the LO pulse (down to a few fs), while providing spectral resolution consistent with the time-frequency uncertainty product [41]. Further, because BHD measures the field quadrature amplitudes, for various phase values, it provides a tomographically complete set of variables (a “quorum,” [95] for quantum state tomography).

In homodyne detection the signal and LO fields have spectra centered at the same, or nearly the same, optical frequency. The signal electric-field  $\hat{\underline{E}}_S^{(+)}$  interferes with the LO field  $\hat{\underline{E}}_L^{(+)}$  at a loss-less beam splitter to produce two output fields, represented by vector-field operators [96–98],

$$\begin{aligned}\hat{\underline{E}}_1^{(+)} &= t_1 \hat{\underline{E}}_S^{(+)} + r_2 \hat{\underline{E}}_L^{(+)} \\ \hat{\underline{E}}_2^{(+)} &= r_1 \hat{\underline{E}}_S^{(+)} + t_2 \hat{\underline{E}}_L^{(+)},\end{aligned}\quad (7.9)$$

where unitarity of the transformation requires that the transmission and reflection coefficients satisfy the Stokes relation,  $r_1 t_2^* + t_1 r_2^* = 0$ . This guarantees that the output field operators commute. We assume the phase convention  $t_1 = -t_2 = t$ ,  $r_1 = r_2 = r$ , and assume that  $t$  and  $r$  have values  $1/\sqrt{2}$ , corresponding to a 50/50 beam splitter (effects of imperfect balance are discussed in [27].)

The LO field is usually assumed to be a strong, pulsed coherent field from a laser (nonclassical LO’s can be treated [27].) The fields incident on the photodiode detectors generate photoelectrons with detection probability (quantum efficiency) equal to  $\eta$ , and the resulting current is integrated using low-noise charge-sensitive amplifiers [99, 100]. This provides the DC-detection, and differs from the often-used method of radio-frequency spectral analysis of the current to study noise at higher frequencies [59]. The values of the integrated photocurrents are recorded with analog-to-digital converters (ADC’s), to give the numbers of photoelectrons  $n_1$  and  $n_2$  per pulse, which are then subtracted, pulse-by-pulse. Alternatively the photocurrents may be subtracted using an analog circuit and then converted to digital form [51, 99]. As seen below, the difference of these numbers is proportional to the chosen electric-field quadrature amplitude of the signal, spatially and temporally averaged over the space-time window defined by the duration of the LO pulse.

The positive-frequency part of the electric field operator is (underbars indicate vectors and carets indicate operators)

$$\hat{\underline{E}}_S^{(+)}(\underline{r}, t) = i \sum_j \sqrt{\frac{\hbar\omega_j}{2\varepsilon_0}} \hat{b}_j \underline{u}_j(\underline{r}) \exp(-i\omega_j t). \quad (7.10)$$

Optical polarization is also indicated by the index  $j$ . In Dirac’s quantization scheme the modes are plane waves,  $\underline{u}_j(\underline{r}) = V^{-1/2} \underline{\varepsilon}^{(j)} \exp(i\mathbf{k}_j \cdot \underline{r})$ , defined in some large volume  $V$  (which may be taken to infinity later). Alternatively they can be taken as any complete set of monochromatic solutions to

the source-free Maxwell equations, assumed to be orthogonal and normalized over volume  $V$ . The annihilation operators obey the commutator

$$[\hat{b}_j, \hat{b}_k^\dagger] = \delta_{jk}. \quad (7.11)$$

We assume that the photodetectors respond to the incident photon flux. (This approach is different from the Glauber/Kelly-Kleiner formulation where the observed photoelectron current is in terms of the electromagnetic energy flux at the detector [101, 102].) The approach used here is considered to be more appropriate for broadband fields in the case that the detector's quantum efficiency is frequency independent, i.e., the detector is an ideal photoemissive, rather than an energy-flux, detector [63, 64, 66, 67, 70, 103, 104]. Appropriately, “photon-flux amplitudes”  $\hat{\Phi}_S^{(+)}(\underline{r}, t)$  and  $\hat{\Phi}_L^{(+)}(\underline{r}, t)$  can be defined for the signal and LO as

$$\begin{aligned} \hat{\Phi}_S^{(+)}(\underline{r}, t) &= i\sqrt{c} \sum_j \hat{b}_j \underline{u}_j(\underline{r}) \exp(-i\omega_j t), \\ \hat{\Phi}_L^{(+)}(\underline{r}, t) &= i\sqrt{c} \sum_j \hat{b}_{Lj} \underline{u}_j(\underline{r}) \exp(-i\omega_j t). \end{aligned} \quad (7.12)$$

The photon fluxes (photon/sec) of the fields ( $i = 1, 2$ ) at the detector faces at  $z = 0$  are then represented by

$$\hat{I}_i(t) = \int_{Det} d^2x \hat{\Phi}_i^{(-)}(\underline{x}, 0, t) \cdot \hat{\Phi}_i^{(+)}(\underline{x}, 0, t), \quad (7.13)$$

integrated over the detectors' faces, where  $\underline{x}$  denotes the transverse variables ( $x, y$ ). It is assumed that the quantities  $\hat{\Phi}^{(+)}(\underline{r}, t)$  obey the same transformations at the beam splitter as do the fields in Eq. (7.9). The photon number contained in a time interval  $(0, T)$  at each detector is represented by  $\hat{N}_i = \int_0^T \hat{I}_i dt$ . The interval duration  $T$  is assumed to be longer than the duration of the signal and LO pulses, so that all of the energy of each is detected. The electronic system, including detector and amplifiers used to process the photocurrent, is assumed to act as a low-pass filter with an integration time (inverse bandwidth) much larger than  $T$ . The difference-photon number contained in a time interval  $(0, T)$  is represented by

$$\hat{N}_- = \hat{N}_1 - \hat{N}_2 = \int_0^T (\hat{I}_1 - \hat{I}_2) dt, \quad (7.14)$$

and the total photon number is

$$\hat{N}_+ = \hat{N}_1 + \hat{N}_2 = \int_0^T (\hat{I}_1 + \hat{I}_2) dt. \quad (7.15)$$

In the case of perfect balancing ( $r = t$ ), the difference number is given by

$$\hat{N}_- = \int_0^T dt \int_{\text{Det}} d^2x \hat{\underline{\Phi}}_L^{(-)}(\underline{x}, 0, t) \cdot \hat{\underline{\Phi}}_S^{(+)}(\underline{x}, 0, t) + h.c. \quad (7.16)$$

As previously mentioned, this (DC) implementation of balanced homodyne detection is different than that in which a radio frequency (RF) spectrum analyzer is used to detect the oscillations of photocurrent at some frequency other than zero. The non-DC case can be analyzed by inserting into Eq. (7.14) and Eq. (7.15) a time-domain filter function centered at some nonzero frequency [67, 68, 71].

The quantum theory of BHD with pulsed LO's is closely related to the general theory of electromagnetic field quantization in terms of non-monochromatic, or temporal-spatial wave packet, modes, first discussed by Titulaer and Glauber [105], and later used in the theory of Raman scattering [106]. Here we apply this approach to the photon-flux amplitude rather than the EM field. This choice brings a simplification of the formalism regarding the orthogonality of the wave packets. Rewrite the signal photon-flux amplitude as

$$\hat{\underline{\Phi}}_S^{(+)}(\underline{r}, t) = i\sqrt{c} \sum_k \hat{a}_k \underline{v}_k(\underline{r}, t), \quad (7.17)$$

where the wave packet modes are given by the superpositions

$$\underline{v}_k(\underline{r}, t) = \sum_j C_{kj} \underline{u}_j(\underline{r}) \exp(-i\omega_j t), \quad (7.18)$$

and  $C_{kj}$  is a unitary matrix of coefficients. The new photon annihilation operators for the wave packet modes are related to the original operators by  $\hat{a}_j = \sum_m C_{jm}^* \hat{b}_m$ , and obey  $[\hat{a}_j, \hat{a}_k^\dagger] = \delta_{jk}$ . As an example of their use, a one-photon wave packet state is created by acting on the vacuum by  $\hat{a}_k^\dagger |vac\rangle$ .

The wave packet modes  $\underline{v}_k(\underline{r}, t)$  are orthonormal in the same large volume  $V$  as are the original modes,  $\int d^3r \underline{v}_k^*(\underline{r}, t) \cdot \underline{v}_m(\underline{r}, t) = \delta_{km}$ . But because photodetection takes place in a plane (the surface of the detector) and not an infinite volume, we find it useful to reformulate the orthogonality of the wave packet modes. Consider that the detector plane (assumed infinite) is at  $z = 0$ , with transverse variables  $\underline{x} = (x, y)$ . Then it can be shown [see proof in Appendix] that for large time  $T$  and paraxial (small-divergence) beams, the integral  $c \int_0^T dt \int_{\text{Det}} d^2x$  acts similarly to the spatial volume integral  $\int_0^{cT} dz \int_{\text{Det}} d^2x$ , since the beam sweeps out such a volume in the detector integration time  $T$ . Therefore if we choose the  $z$ -dimension of the quantization volume to have length equal to  $cT$ , (with  $T$  large), then we have orthogonality in the transverse-space-plus-time domain,

$$c \int_0^T dt \int_{\text{Det}} d^2x \underline{v}_k^*(\underline{x}, 0, t) \cdot \underline{v}_m(\underline{x}, 0, t) = \delta_{km}, \quad (7.19)$$

where  $\underline{v}_m(\underline{x}, 0, t)$  is the wave packet mode in the  $z = 0$  plane. This property makes it easy to analyze BHD with a pulsed LO.

Let us assume that the LO pulse is a strong coherent state of a particular localized wave packet mode  $\underline{v}_L(\underline{r}, t)$ . Then it is useful to separate the LO operator into a term for this mode and terms for all other (vacuum) modes,

$$\hat{\underline{\phi}}_L^{(+)}(\underline{r}, t) = i\sqrt{c} \hat{c}_L \underline{v}_L(\underline{r}, t) + \text{vacuum terms}, \quad (7.20)$$

where  $\hat{c}_L$  is the annihilation operator for LO mode.

Inserting this into Eq. (7.16) and dropping the vacuum terms, which are small compared to the terms involving the strong LO, gives for the difference number

$$\begin{aligned} \hat{N}_- &= -i\sqrt{c} \hat{c}_L^\dagger \int_0^T dt \int_{Det} d^2x \underline{v}_L^*(\underline{x}, 0, t) \cdot \hat{\underline{\phi}}_S^{(+)}(\underline{x}, 0, t) + h.c. \\ &= \hat{c}_L^\dagger \hat{a} + \hat{c}_L \hat{a}^\dagger, \end{aligned} \quad (7.21)$$

where

$$\hat{a} = \sum_k \hat{a}_k c \int_0^T dt \int_{Det} d^2x \underline{v}_L^*(\underline{x}, 0, t) \cdot \underline{v}_k(\underline{x}, 0, t). \quad (7.22)$$

This illustrates the spatial and temporal gating of the signal field, since it is multiplied by the LO field, which has some controlled shape—where the LO is zero, that portion of the signal is rejected.

If we assume that the detector is large enough to capture the whole transverse profile of the modes of interest, then the integral in Eq. (7.21) acts like an integral over the quantization volume, and orthogonality [Eq. (7.19)] shows that  $\hat{a}$  in Eq. (7.22) stands for  $\hat{a}_{k=L}$ , the photon operator for the mode of the signal beam that has the same spatial-temporal shape as does the LO mode. It is the operator for the detected part of the signal field—the part that is “mode-matched” to the local oscillator. We can represent  $\hat{a}$  by its “real” and “imaginary” parts given by  $\hat{q} = (\hat{a} + \hat{a}^\dagger)/\sqrt{2}$  and  $\hat{p} = (\hat{a} - \hat{a}^\dagger)/(i\sqrt{2})$ . These two variables are called quadrature-amplitudes for the detected spatial-temporal mode and obey  $[\hat{q}, \hat{p}] = i$ .

If the LO field is coherent and strong, and treated classically, with amplitude  $\hat{c}_L$  in Eq. (7.21) replaced by  $\alpha_L = |\alpha_L|e^{i\theta}$ , then the difference number becomes [107]

$$\hat{N}_- = \hat{N}_-(\theta) = |\alpha_L|(\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta}). \quad (7.23)$$

With the LO phase equal to zero, a balanced homodyne detector measures the real quadrature  $\hat{q}$ , while with the LO phase equal to  $\pi/2$  it measures the imaginary quadrature  $\hat{p}$ . For arbitrary phase it measures the generalized quadrature amplitude

$$\hat{q}_\theta = \hat{N}_-(\theta)/(|\alpha_L|2^{1/2}) = (\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta})/2^{1/2}. \quad (7.24)$$

A conjugate variable is defined by

$$\hat{p}_\theta = (\hat{a}e^{-i\theta} - \hat{a}^\dagger e^{i\theta})/(i2^{1/2}). \quad (7.25)$$

These variables may be expressed as a rotation of the original variables,

$$\begin{pmatrix} \hat{q}_\theta \\ \hat{p}_\theta \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \hat{q} \\ \hat{p} \end{pmatrix}. \quad (7.26)$$

The variables  $\hat{q}_\theta$  and  $\hat{p}_\theta$  do not commute, and so cannot be measured jointly with arbitrarily high precision. Their standard deviations obey the uncertainty relation,  $\Delta q_\theta \Delta p_\theta \geq 1/2$ .

The photoelectron counting distribution for the difference photoelectron number  $n_-$  in DC pulsed homodyne detector is [27]

$$\Pr(n_-) = \left\langle : \exp[-\eta(\hat{N}_1 + \hat{N}_2)] \left( \frac{\hat{N}_2}{\hat{N}_1} \right)^{n_-/2} I_{|n_-|}[(2\eta(\hat{N}_1 \hat{N}_2)^{1/2}] : \right\rangle_{S,L}, \quad (7.27)$$

where the double dots indicate normal operator ordering (annihilation operators to the right of creation operators), and  $I_n(x)$  is the modified Bessel function of  $n$ -th order. This general result incorporates multimode, pulsed signal and LO fields in arbitrary quantum states. It can also accommodate non-mode-matched background from the signal beam (see [27] for a more general discussion.)

We consider the LO field to be in an intense coherent state  $|\alpha_L\rangle$  of a single spatial-temporal-mode such that  $\hat{c}_L|\alpha_L\rangle = \alpha_L|\alpha_L\rangle$ . Then, the expression Eq. (7.27) can be evaluated and well approximated by a Gaussian function [27],

$$\Pr(n_-, \theta) = \left\langle : \frac{\exp[-(n_- - \eta\hat{N}_-(\theta))^2/(2\eta\hat{N}_+)]}{[\pi 2\eta\hat{N}_+]^{1/2}} : \right\rangle_S, \quad (7.28)$$

where the difference number  $\hat{N}_-(\theta)$  is given by Eq. (7.23) and the total number of photons hitting both detectors is

$$\hat{N}_+ = |\alpha_L|^2 + \hat{a}^\dagger \hat{a}. \quad (7.29)$$

The total number of photons sets the scale for the shot-noise level. The quantum expectation value in Eq. (7.28) is with respect to the signal field, denoted by  $S$ , i.e.,  $\langle \dots \rangle_S = \text{Tr}(\hat{\rho}_s \dots)$  where  $\hat{\rho}_s$  is the density operator for the signal mode(s).

We use the quadrature-amplitude operator  $\hat{q}_\theta = \hat{N}_-(\theta)/(|\alpha_L|^2)^{1/2}$  defined in Eq. (7.24), and define the corresponding real variable  $q_\theta = n_-/(\eta|\alpha_L|^2)^{1/2}$ , (accounting for detector efficiency  $\eta$ ). Then we can transform Eq. (7.28) into the probability density for quadrature amplitude of the mode-matched signal [27]

$$\Pr(q_\theta, \theta) = \left\langle : \frac{\exp[-(q_\theta - \hat{q}_\theta)^2/(2\sigma^2)]}{[\pi 2\sigma^2]^{1/2}} : \right\rangle_S, \quad (7.30)$$

where  $2\sigma^2 = 1/\eta$ .

The distributions  $\Pr(q_\theta, \theta)$  are experimentally estimated by repeatedly measuring values of  $q_\theta$  for various fixed LO phases  $\theta$  and building measured histograms  $\Pr_M(q_\theta, \theta)$  of relative frequencies of the occurrence of each quadrature value. From these measured quadrature distributions one reconstructs the quantum state of the mode-matched signal field.

### 7.3 Quantum State Reconstruction and Optical Mode Statistical Sampling

#### 7.3.1 Inverse Radon Reconstructions

It can be shown that the (exact) quadrature distributions are related to the Wigner distribution of the signal mode by [27]

$$\Pr(q_\theta, \theta) = \iint \frac{\exp[-\{q_\theta - \tilde{q}_\theta(q, p)\}^2/(2\varepsilon^2)]}{\sqrt{\pi 2\varepsilon^2}} W(q, p) dq dp, \quad (7.31)$$

where  $\tilde{q}_\theta(q, p) = q \cos \theta + p \sin \theta$  and  $2\varepsilon^2 = 2\sigma^2 - 1 = 1/\eta - 1$ . In the limit that  $\eta = 1$ , Eq. (7.31) becomes

$$\Pr(q_\theta, \theta) = \iint \delta(q_\theta - \tilde{q}_\theta(q, p)) W(q, p) dq dp. \quad (7.32)$$

Using  $q = q_\theta \cos \theta - p_\theta \sin \theta$ , and  $p = q_\theta \sin \theta + p_\theta \cos \theta$ , gives

$$\Pr(q_\theta, \theta) = \int_{-\infty}^{\infty} W(q_\theta \cos \theta - p_\theta \sin \theta, q_\theta \sin \theta + p_\theta \cos \theta) dp_\theta. \quad (7.33)$$

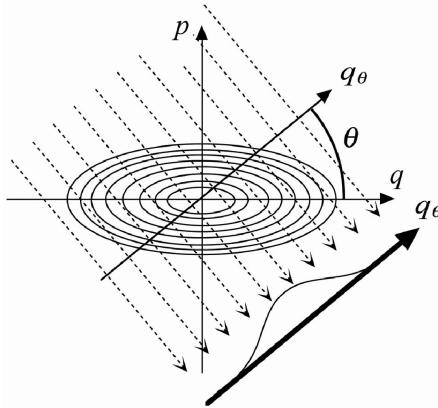
This integral has the form of a marginal distribution, that is, the joint distribution has been integrated over one independent random variable to yield the distribution for the other variable.

The integral Eq. (7.33) is known as the Radon transform [1, 108], and has the form of a projection of the  $W$  function onto the rotated  $q_\theta$  axis, as illustrated by the line integrals in Fig. 7.2.

As pointed out in the context of quantum mechanics first by Bertrand and Bertrand [80] and independently by Vogel and Risken [10], Eq. (7.33) can be inverted to yield  $W(q, p)$ , given a set of distributions  $\Pr(q_\theta, \theta)$  for all values of  $\theta$  between 0 and  $\pi$ . The formal inversion is

$$W(q, p) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dq_\theta \int_{-\infty}^{\infty} d\xi |\xi| \int_0^\pi d\theta \Pr(q_\theta, \theta) e^{i\xi(q_\theta - q \cos \theta - p \sin \theta)}. \quad (7.34)$$

In the case that one has measured a sufficient number of distributions  $\Pr(q_\theta, \theta)$  for a finite set of discrete  $\theta$  values, the inversion can be carried out



**Fig. 7.2.** Equal-value contours of a Gaussian-like Wigner distribution, showing squeezing. Line integrals perpendicular to an axis rotated by angle  $\theta$  create a projected, or marginal, distribution  $\text{Pr}(q_\theta, \theta)$ .

numerically (with a certain amount of smoothing of the final result) using the well-studied filtered back projection transformation familiar in medical tomographic imaging [1, 108, 109]. This is the origin of the expression *optical homodyne tomography* [16], and later *quantum state tomography*. Equation (7.34) explicitly demonstrates the validity of the statements made in Sec. 7.1; namely that many measurements of many observables ( $q_\theta$ , for many values of  $\theta$ ) enables one to determine the quantum mechanical state of a system [here the state is represented by  $W(q, p)$ .]

As mentioned above, if the inverse Radon transform Eq. (7.34) is applied to the measured histograms  $\text{Pr}_M(q_\theta, \theta)$ , which are only estimates of the true distributions, nonphysical forms of the reconstructed state may erroneously appear. Error bars can be estimated in this case [89], or nondeterministic methods can be adopted [90, 92].

If the detector efficiency  $\eta$  is less than unity, then the reconstruction does not yield the Wigner distribution of the signal mode [27, 78, 110]. It can be shown that the experimentally reconstructed Wigner distribution in this case is smoothed, and is given by

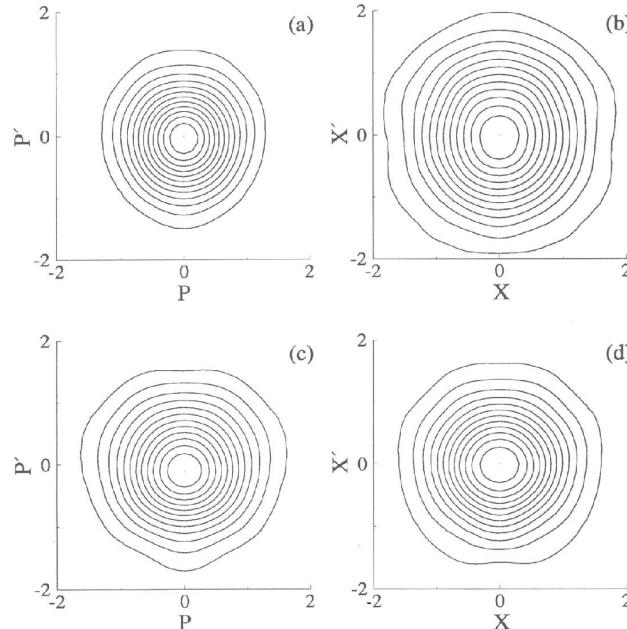
$$W_{\text{Exp}}(q, p) = \frac{1}{\pi 2\varepsilon^2} \iint e^{-(q-q')^2/(2\varepsilon^2) - (p-p')^2/(2\varepsilon^2)} W(q', p') dq' dp', \quad (7.35)$$

where again  $2\varepsilon^2 = 1/\eta - 1$ . In principle, the smoothing function could be de-convolved from the measured distribution  $W_{\text{Exp}}(q, p)$  to yield  $W(q, p)$ , but with experimental data having finite signal-to-noise ratio, and perhaps containing systematic errors, this is not usually practical. Although  $W(q, p)$  can be negative, for  $\eta \leq 1/2$  the integral in Eq. (7.35) is always positive definite, and so the measurement cannot show certain highly quantum effects.

If the signal field is excited in a *single* spatial-temporal mode  $\underline{w}_S(\underline{r}, t)$ , different than that of the LO,  $\underline{v}_L(\underline{r}, t)$ , then a part of it contributes to the detected amplitude, and the remainder to the background. In this case the detector efficiency  $\eta$  is replaced by the product  $\eta \eta_{LS}$ , where the complex mode overlap is

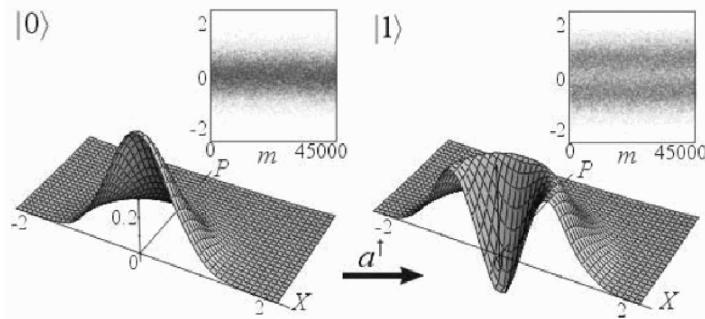
$$\eta_{LS} = \left| c \int_0^T dt \int_{Det} d^2x \underline{v}_L^*(\underline{x}, 0, t) \cdot \underline{w}_S(\underline{x}, 0, t) \right|, \quad (7.36)$$

and  $0 \leq \eta_{LS} \leq 1$ . The mode-overlap factor  $\eta_{LS}$ , which includes both spatial and temporal overlap, acts like an additional attenuation factor for the mode of interest.



**Fig. 7.3.** Measured density matrix for (a), (b) the squeezed vacuum and (c), (d) the vacuum state, in  $q$  (called here  $x$ ) or  $p$  representations: (a), (c)  $|\langle p + p' | \hat{\rho} | p - p' \rangle|$ , (b), (d)  $|\langle q + q' | \hat{\rho} | q - q' \rangle|$  [16].

The first such state reconstruction was carried out by Smithey et al. in 1993, for the case of a squeezed state created by optical parametric amplification of the vacuum [16,35]. Figure 7.3 shows the reconstructed density matrix in the quadrature representation, obtained by an inverse Fourier transform of the measured Wigner distribution [see Eq. (7.2)]. The overall efficiency  $\eta \eta_{LS}$  in this case was around 0.5, so the measured Wigner distribution Eq. (7.35) contained a fair amount of broadening and smoothing.



**Fig. 7.4.** Experimental reconstruction of a vacuum-state and a one-photon Wigner function. Insets show raw quadrature histogram data. (Ref. [20] and private communication, A. Lvovsky)

The first state reconstruction showing a negative Wigner distribution was performed by Lvovsky et al. [20]. Near-pure-state single-photon wave packets were created by parametric down-conversion combined with strong spatial and spectral filtering. An overall measurement efficiency of 0.55 allowed the one-photon component to dominate the vacuum component sufficiently to lead to a negative value of  $W_{Exp}(0,0)$ .

If the state determined by OHT is determined to be a pure state (by computing  $Tr[\hat{\rho}^2]$  and finding it equal to unity), then it is possible to reconstruct a Schrödinger wave function using

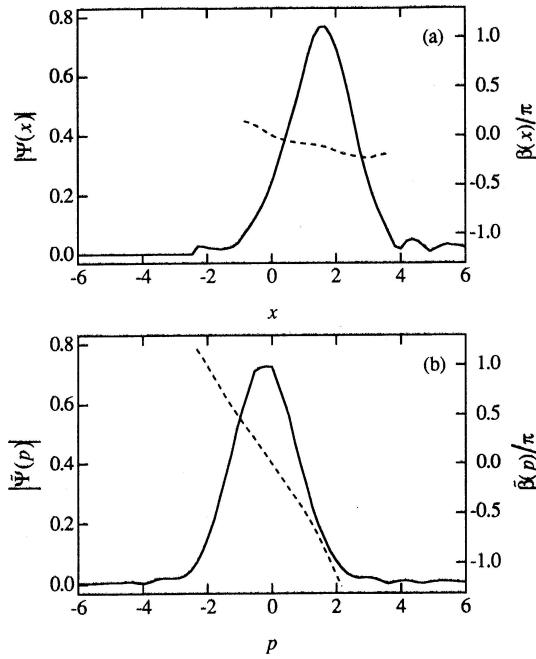
$$\psi(q) = \frac{\langle q|\hat{\rho}|q' = 0\rangle}{\sqrt{\langle q = 0|\hat{\rho}|q' = 0\rangle}} e^{i\beta_0}, \quad (7.37)$$

where  $\beta_0$  is a nonphysical constant phase. An example for a coherent-state optical mode is shown in Fig. 7.5. (Note that because coherent states are robust against detector losses, it is valid to consider the measured quadratures as simply scaled down versions of the original quadratures. In this case the state that is measured is interpreted as that *after* suffering detector losses.)

Given an experimentally reconstructed Wigner distribution  $W_{Exp}(q,p)$ , one can invert Eq. (7.2) to obtain the measured density matrix in the quadrature  $\langle q|\hat{\rho}|q'\rangle$  (as in Fig. 7.3) or number  $\langle n|\hat{\rho}|m\rangle$  representations. From this density matrix one can compute probability distributions and statistical averages of any quantity of interest pertaining to the system studied.

A significant example of this type of indirect measurement is the reconstruction of the photon-number probability distribution  $\text{Pr}(n) = \langle n|\hat{\rho}|n\rangle$ , which is obtained by measuring quadratures, not photon number directly. Also of interest is the distribution of the quantum phase  $\phi$  of the signal field.

Among the many useful definitions of quantum phase distributions [111], here we illustrate the idea using the London/Pegg-Barnett distribution [112],

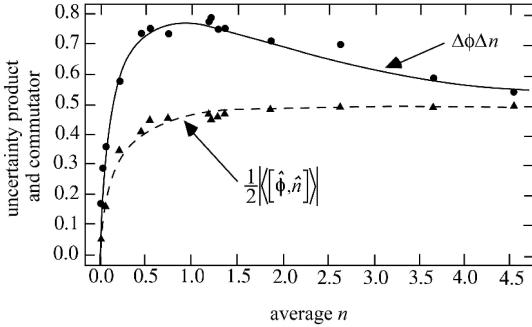


**Fig. 7.5.** Measured Schrödinger wave function of a coherent state of light with an average of 1.2 photons in (a) the  $q$  (or  $x$ )-quadrature representation and (b) the  $p$ -quadrature representation. [37] Solid line is the magnitude  $|\psi|$  plotted, against the left axis; dashed line is phase, plotted against the right axis.

$$\text{Pr}_{L/PB}(\phi) = \frac{1}{2\pi} \sum_{n,m=0}^s e^{i(m-n)\phi} \langle n|\hat{\rho}|m\rangle, \quad (7.38)$$

where  $s$  is a truncation parameter. Our group demonstrated the reconstruction of photon-number [37] and phase distributions [34, 35] for coherent and quadrature-squeezed states. We further demonstrated how to reconstruct the expectation value of the commutator of the number operator and the Pegg-Barnett phase operator,  $[\hat{\phi}, \hat{n}]$  [36, 37]. This is interesting because this commutator is not a Hermitian operator and so does not correspond to a conventional “observable” in quantum theory; nevertheless its measurement (or rather the measurement of its moments) is amenable to the OHT technique.

Figure 7.6 shows the number-phase uncertainty product and the expectation value of the number-phase commutator, both determined from experiment for a coherent state with varying mean photon number. The uncertainty principle is of course satisfied. The interesting features are the degree to which the product does not equal the commutator, and the amount by which the product is less than 0.5, the limit that is expected in the limit of large photon numbers.



**Fig. 7.6.** Measured number-phase uncertainty product and expectation value of the number-phase commutator for a coherent state with mean number  $n$ . Continuous curves are theoretical predictions [37].

### 7.3.2 Pattern Function Reconstructions

The path just described for obtaining the density matrix  $\rho_{mn} = \langle m|\hat{\rho}|n\rangle$  and photon number distribution  $\text{Pr}(n)$  from the raw quadrature data is rather tortuous, and can introduce unnecessary error, in part as a result of the smoothing that is necessary in the inverse Radon transform. It was first pointed out by D'Ariano et al. that it is possible, and perhaps preferable, to bypass the inverse Radon transform [61]. Why not invert Eq. (7.4) directly to obtain the number-basis density matrix directly for the homodyne data? It was independently proposed that the quadrature-basis density matrix could be similarly obtained [78]. Although these methods do not use the Radon inverse, they have nevertheless come to be called quantum state “tomography” (from the original *optical homodyne tomography* [16]), as now are all schemes that use raw statistical data of one form or another to infer a quantum state.

The idea behind inverting Eq. (7.4) is rather simple when viewed from the perspective of dual bases. Rewrite Eq. (7.4) as

$$\text{Pr}(q, \theta) = \sum_{\mu, \nu} \rho_{\mu\nu} G_{\mu\nu}(q, \theta), \quad (7.39)$$

where  $G_{\mu\nu}(q, \theta) = \psi_\mu(q)\psi_\nu(q) \exp[i(\nu - \mu)\theta]$  (with  $\psi_\mu$  being harmonic oscillator wavefunctions) serves as a basis for expanding the function  $\text{Pr}(q, \theta)$ . The basis  $\{G_{\mu\nu}(q, \theta)\}$  is comprised of linearly independent but nonorthogonal functions. Their linear independence guarantees that there exists a dual basis  $\{F_{mn}(q, \theta)\}$ , also nonorthogonal, with the bi-orthogonality property

$$\int_0^{2\pi} \frac{d\theta}{2\pi} \int_{-\infty}^{\infty} dq F_{mn}^*(q, \theta) G_{\mu\nu}(q, \theta) = \delta_{m\mu} \delta_{n\nu}. \quad (7.40)$$

Using this, the relation Eq. (7.39) can easily be inverted to yield

$$\rho_{mn} = \int_0^{2\pi} \frac{d\theta}{2\pi} \int_{-\infty}^{\infty} dq F_{mn}^*(q, \theta) \Pr(q, \theta). \quad (7.41)$$

To find the form of the dual-basis functions, make the ansatz  $F_{mn}(q, \theta) = M_{mn}(q) \exp[-i(m-n)\theta]$ , and define  $\chi_{\mu\nu} = \psi_\mu(q)\psi_\nu(q)$ , which, when inserted into Eq. (7.40), leads to the sole requirement [113]

$$\int_{-\infty}^{\infty} M_{n+D,n}(q) \chi_{\nu+D,\nu}(q) dq = \delta_{n\nu}, \quad (7.42)$$

where  $D = \mu - \nu = m - n$  is the common difference between the indices. This shows that the problem breaks up into a set of independent problems—one for each value of  $D$ . The dual functions can be found for each value of  $D$  by defining the matrix of overlap integrals,  $\phi_{\mu\nu}^{(D)} = \int_{-\infty}^{\infty} \varphi_\mu^{(D)}(q) \chi_{\nu+D,\nu}(q) dq$ , where  $\varphi_\mu^{(D)}(q)$  is a set of bases (one for each  $D$ ), which can be nonorthogonal and may be chosen for convenience. It is easy to see that the dual-basis vectors are constructed as

$$M_{\mu+D,\mu}(q) = \sum_{\nu=\Omega}^{\infty} [\phi^{(D)-1}]_{\mu\nu} \varphi_\nu^{(D)}(q), \quad (7.43)$$

where  $\phi^{(D)-1}$  is the inverse of the matrix  $\phi^{(D)}$  and  $\Omega$  is the greater of 0 and  $-D$ . A convenient choice for the numerical basis is given by a product of the Hermite-Gaussian functions and a Gaussian function,  $\varphi_\mu^{(D)}(q) = (2\nu + 1)^L \psi_{m(\nu)}(q) \exp(-q^2/2)$ , where  $m(\nu)$  is constrained to be even (odd) if  $D$  is even (odd).  $L$  is a parameter of order unity that can be adjusted to enhance numerical stability. This form allows the overlap integrals to be carried out in closed form.

The prescription just given does not provide the best numerical algorithm for finding the pattern functions, but it does prove that a linear-transform inverse of Eq. (7.39) exists. This is written in full as (in the notation of [114])

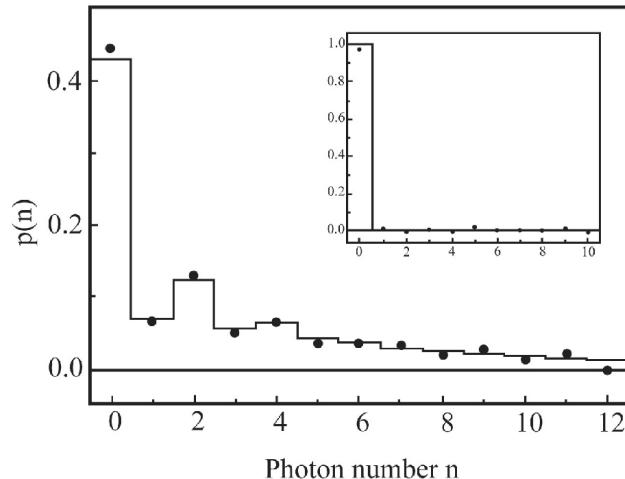
$$\rho_{mn} = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta \int_{-\infty}^{\infty} dq \exp[i(m-n)\theta] M_{mn}(q) \Pr(q, \theta), \quad (7.44)$$

where the functions  $M_{mn}(q)$  are the elements of the dual basis, and are often called *pattern functions* and given in the form  $f_{mn}(q) = (1/\pi)M_{mn}(q)$  [89]. Efficient algorithms for their numerical construction have been given [1, 89]. Several examples of the pattern functions are given in chapter by D'Ariano, Paris and Sacchi of this volume.

In experiments the continuous integral over phase in Eq. (7.44) is replaced by a sum over discrete, equally spaced phase values,  $\theta_k$ ,

$$\rho_{mn}^{(d)} = \frac{1}{d} \sum_{k=1}^d \int_{-\infty}^{\infty} dq \exp[i(m-n)\theta_k] M_{mn}(q) \Pr(q, \theta_k). \quad (7.45)$$

In general, one would think the larger the number of phases  $d$ , the better. An analysis by Leonhardt and Munroe determined the minimum number of LO phases needed for a faithful state reconstruction [114]. They proved that if the field is known a priori to contain at most  $n$  photons, then the number of LO phases needed equals simply  $n + 1$ , the dimension of the Hilbert space containing the state. Equation (7.4) shows that the quadrature distribution contains oscillations as a function of phase that are determined by the occupied photon-number states; there is no further information to be had by using more phase values than  $n + 1$  since those higher phase “frequencies” would be aliased to lower frequencies, as is familiar in the Nyquist/ Shannon sampling concept [115]. Leonhardt and Munroe also show how to estimate the error incurred by choosing too few phases in the case that the maximum photon number is not known ahead of time.



**Fig. 7.7.** Photon number distribution of squeezed vacuum and vacuum (insert). Continuous curves are theory. [42]

A beautiful demonstration of the reconstruction of a density matrix in the number basis was carried out by Schiller et al. [42], and a comprehensive and very instructive report was given by Breitenbach [93]. A quadrature-squeezed field produced by optical parametric oscillation was detected by BHD to produce quadrature histograms for 128 phase values. The density matrix elements up to  $n = 6$  were reconstructed using Eq. (7.45). The diagonal elements, or probabilities, which are plotted in Fig. 7.7, show clearly that even photon numbers (0, 2, 4) predominate, since photon-pair production is the origin of the quadrature-squeezed vacuum field. This was the first observation of the even-odd-number oscillations of squeezed light, an effect that could not be measured by using direct photon-number detection due to

a lack (at that time) of photon-number-discriminating detectors. This shows one of the unique capabilities of the indirect measurement idea. Even stronger number oscillations were reconstructed from the same measured data by using non-deterministic, least-squares methods [93].

### 7.3.3 Ultrafast Photon Number Sampling using Phase-Averaged BHD

An especially useful application of the pattern function idea is the indirect measurement of photon-number statistics by the use of a phase-random or phase-swept local oscillator (LO) [39,116]. It is clear from Eq. (7.44) that the photon-number probability  $\rho_{nn}$  is given by

$$\rho_{nn} = \int_{-\infty}^{\infty} d\xi M_{nn}(\xi) \overline{\Pr(\xi)}, \quad (7.46)$$

where the phase-averaged quadrature distribution is

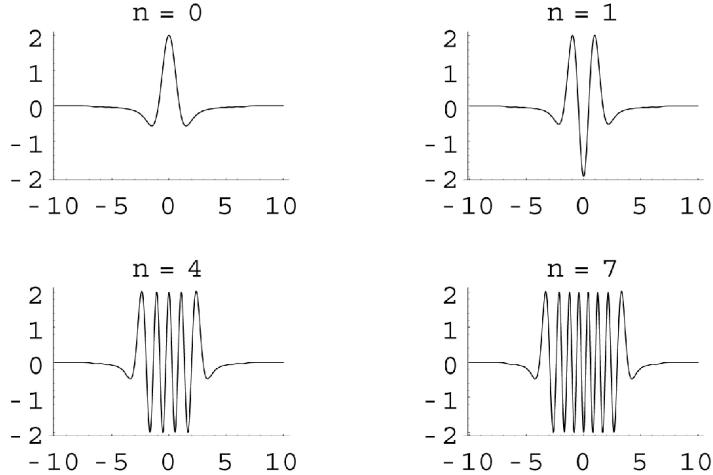
$$\overline{\Pr(\xi)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta \Pr(q = \xi, \theta), \quad (7.47)$$

where  $\xi$  represents the phase-independent quadrature variable. This means that in an experiment, only a *single* quadrature distribution needs to be measured, while randomizing or sweeping the phase uniformly in the interval  $-\pi$  to  $\pi$ . This greatly reduces the amount of data needed to obtain the number statistics using OHT and also removes artifacts that may arise from the discrete stepping of the phase as is usually done in OHT.

The diagonal pattern (or sampling) functions are generated numerically most efficiently using the algorithm in [89]. Several examples are shown Fig. 7.8.

An interesting feature of the functions  $M_{nn}(\xi)$  is that their maxima and minima correspond to the maxima and minima of the quadrature distribution for a Fock state  $\overline{\Pr(\xi)} = \psi_n^*(\xi)\psi_n(\xi)$ , where  $\psi_n(\xi)$  is the wave function in the Fock basis.

This leads to the interesting and useful point that in many cases the best way to measure the statistical properties of photon numbers is to *not* measure photon number directly. Rather measure the quadrature-field amplitudes and infer the photon statistics. (See also the discussion in Sec. 7.1.) The latter has the advantages of high-quantum-efficiency detection (approaching 100%), good discrimination between probabilities of zero, one, two, etc. photon numbers, tens of fs time resolution, temporal and spatial mode selectivity. A complementary disadvantage is that if the average number of photons in the selected spatial-temporal mode is far less than unity, then the amount of signal averaging required to reconstruct the field and number statistics can make the technique impractical. Nevertheless, even in the case of photon numbers much less than one, the mean photon number (if not its statistics)



**Fig. 7.8.** Sampling functions  $M_{n,n}(\xi)$  versus  $\xi$  for  $n = 0, 1, 4$ , and  $7$  [62].

in the detected mode can be measured using OHT if high pulse rates ( $> 1$  MHz) are used.

Another drawback to the indirect measurement method is that it may lead to a lower precision of measurement than a direct method (if one is available) [117, 118]. This arises because a tomographic measurement is capable of yielding information about all of the system variables and, consistent with the uncertainty principle, one would expect a degrading of the precision in estimating a given observable's value from a finite data set. On the other hand, for determining distributions of some variables, such as photon number on fs time scales (see below), or the London/ Pegg-Barnett phase (see above), there are no direct measurement techniques known. Then indirect measurement provides a useful path, as illustrated further in the following.

Perhaps the most interesting properties of the photon number of a given mode are its mean value and its fourth-order statistics. A nice property of *phase-averaged homodyne detection* is its ability to extract these quantities from the quadrature data without the need to first reconstruct the quantum state or even the photon number distribution. Munroe et al. showed that the statistical moments of photon number can be computed directly from the raw quadrature data [39, 62].

For example, the mean photon number in the detected signal mode is proportional to the expectation value of the square of the quadrature amplitude, averaged over all phase values,

$$\langle \hat{n} \rangle = \langle \hat{a}^\dagger \hat{a} \rangle = \frac{1}{2\pi} \int_0^{2\pi} \langle \hat{q}_\theta^2 \rangle d\theta - \frac{1}{2} = \langle\langle q^2 \rangle\rangle - \frac{1}{2}. \quad (7.48)$$

The double bracket in the last term indicates an experimental average over LO phase and measured quadrature values  $q$  (i.e.,  $\xi$ ). The subtraction of  $1/2$  removes the vacuum (zero-point) contribution.

Munroe also derived the minimum number  $N_{\min}$  of measurements needed to reliably determine the mean photon number [62],

$$N_{\min} = \frac{3}{2} \frac{\langle \hat{n}^2 \rangle + \langle \hat{n} \rangle + 1/2}{\langle \hat{n} \rangle^2}. \quad (7.49)$$

For example, a field that is coherent or thermal with mean number  $10^{-3}$  would require 750,000 measurements to reach a signal-to-noise ratio equal to one for the mean photon number. Munroe compared this to a similar quantity for standard photon counting in the presence of a background count,  $N_{\min}^{PC} = [\langle \hat{n}^2 \rangle + 2\langle \hat{n} \rangle \langle \hat{n}_B \rangle + \langle \hat{n}_B^2 \rangle]/\langle \hat{n} \rangle^2$  [119]. He pointed out that the two nearly coincide in the case that the background has a mean value of  $1/2$  photon. This arises because BHD detects the vacuum field, showing why standard counting can be superior to BHD for very weak signals in the absence of significant background.

Formulas for higher-order moments of the photon-number distribution can be calculated similarly to Eq. (7.48), using Richter's formula for the number factorial moment [120],

$$\begin{aligned} \langle n^{(r)} \rangle &= \sum_{n=0}^{\infty} [n(n-1)\dots(n-r+1)]p(n) = \langle (\hat{a}^\dagger)^r (\hat{a})^r \rangle \\ &= \frac{(r!)^2}{2^r (2r)!} \int_0^{2\pi} \frac{d\theta}{2\pi} \langle H_{2r}(\hat{q}_\theta) \rangle, \end{aligned} \quad (7.50)$$

where  $H_j(x)$  is the Hermite polynomial. An important quantity that can be derived from this is the “second-order coherence,” defined as the normalized, normally ordered number-squared,

$$g^{(2)}(t, t) = \frac{\langle : \hat{n}^2 : \rangle}{\langle \hat{n} \rangle^2} = \frac{\langle \hat{n}^2 \rangle - \langle \hat{n} \rangle}{\langle \hat{n} \rangle^2}, \quad (7.51)$$

where the time argument indicates that the sampling takes place using an LO pulse centered at time  $t$ . This quantity is computed from the data using [62, 120]

$$g^{(2)}(t, t) = \frac{(2/3) \langle \langle q^4 \rangle \rangle - 2 \langle \langle q^2 \rangle \rangle + 1/2}{\langle \langle q^2 \rangle \rangle^2 - \langle \langle q^2 \rangle \rangle + 1/4}. \quad (7.52)$$

Munroe has also derived a scheme for computing the statistical uncertainties of such moments directly from the raw data, making it possible to put error bars on the measured values [62, 89]. For example, the variance in the measured mean photon number  $\langle n \rangle$  is estimated by  $N^{-1} \langle \langle \xi^4 \rangle \rangle$ , where  $N$  is the total number of pulses sampled. And the variance in the measured photon number probability  $p(n)$  is estimated by  $N^{-1} \langle \langle M_{nn}^2(\xi) \rangle \rangle \leq 4/N$ .

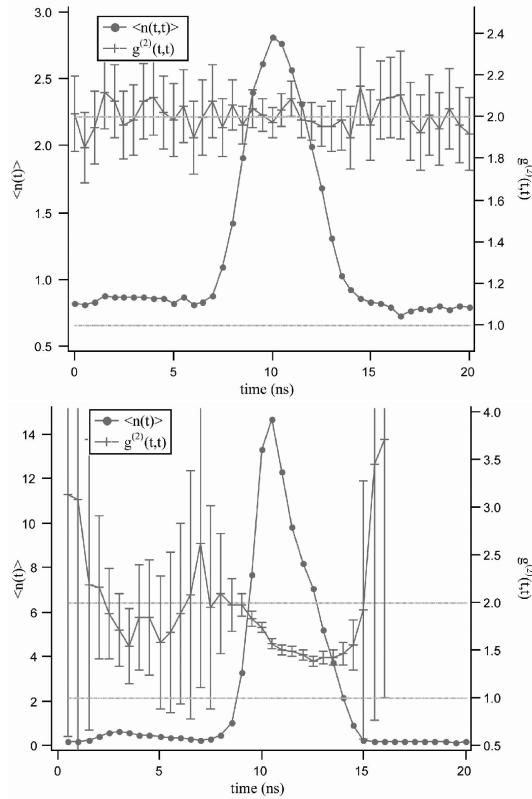
In the case that the LO has the form of an ultrashort pulse, BHD provides a fast time gating or windowing capability, as pointed out in connection with Eq. (7.21) above. This was first demonstrated by Munroe et al. for the 5-ns signal pulse from a semiconductor diode laser below threshold [39]. The diode laser was pumped electrically by a voltage pulse synchronized to the mode-locked pulse train from the fs laser that served as the LO. This is the first example of OHT being applied to a signal that was not optically pumped by light derived from the LO laser. In this case the LO phase needs not be swept, since the signal source is intrinsically phase random.

Munroe also carried out detailed studies of number statistics of light emitted by a traveling-wave semiconductor amplifier, or super luminescent diode (SLD) [62]. This is a diode with gain but virtually no cavity feedback. Amplified spontaneous emission (ASE) from amplifiers plays an important role as noise in optical communication systems. When electrically pulsed synchronously with the LO clock, the SLD emits ASE as a 5-ns pulse which is sampled using the random-phase BHD technique. The SLD could be operated either in a single-pass configuration, with no feedback, or with one-sided feedback (but no cavity) being provided by a diffraction grating. After 10,000 samples were collected using a 150-fs LO pulse that could be varied in its delay, the mean photon number and second-order coherence were computed as in Eqs. (7.48) and (7.52). These are shown in Fig. 7.9.

A value of  $g^{(2)}(t, t) = 2$  corresponds to thermal light, i.e. light produced primarily by spontaneous emission, and a value of  $g^{(2)}(t, t) = 1$  corresponds to light with Poisson statistics, i.e., light produced by stimulated emission in the presence of gain saturation. Figure 7.10 shows plots of the photon number distribution, with error bars, of the SLD emission determined from the measured quadrature distributions.

As we have seen, BHD with a pulsed LO provides a powerful technique for ultrafast time-gated detection of the signal field rather than its intensity. It has recently been developed into a practical scheme called *Linear Optical Sampling*, useful in, for example, testing fast optical telecommunications hardware [41, 48, 94]. It is useful to ask what fundamentally limits the time resolution of this sampling technique. It is clear from the above discussion that this resolution can be as short as the duration of the LO pulse. But it is easy to see that there are special cases in which the time resolution can be far better than would appear to be set by the LO duration itself. Consider that the LO temporal mode can be written as  $\underline{\Phi}_L^{(+)}(\underline{x}, 0, t) = i\sqrt{c} \hat{c}_L v_L(\underline{x}) f_L(t - \tau)$ , that is, a product of a normalized transverse spatial part and a normalized temporal part  $f_L(t - \tau)$  that is delayed by a variable time  $\tau$ . (This can be accomplished experimentally in a given plane, where the detector is located, while at any other plane the field will in general not factor.) Then the difference number in Eq. (7.16) can be written

$$\hat{N}_-(\tau) = -i\sqrt{c} \alpha_L^* \int_0^T dt f_L^*(t - \tau) \phi_S(t) + h.c., \quad (7.53)$$



**Fig. 7.9.** Plots of the average photon number  $\langle n(t) \rangle$  versus time ( $\bullet$ ) and the second-order coherence  $g^{(2)}(t,t)$  versus time (+) for the SLD in (a) the single-pass configuration and (b) double-pass with grating configuration. From [62].

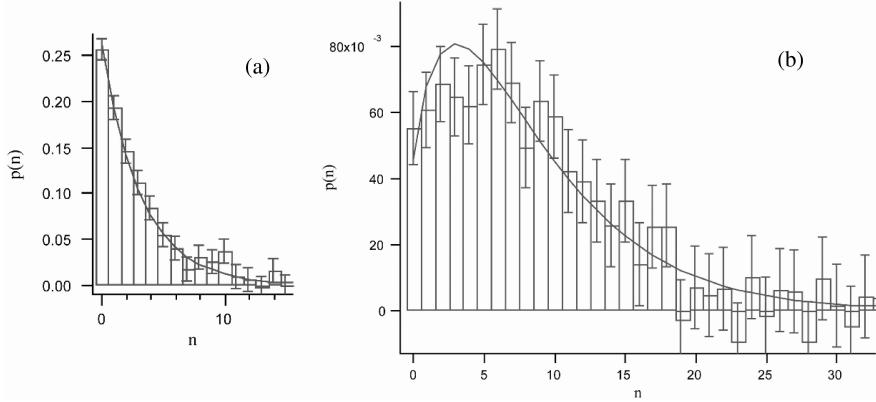
where the signal-field time dependence is given by

$$\phi_S(t) = \int_{Det} d^2x \underline{v}_L^*(\underline{x}) \cdot \Phi_S^{(+)}(\underline{x}, 0, t). \quad (7.54)$$

This can be written in the Fourier domain as

$$\hat{N}_-(\tau) = -i\sqrt{c}\alpha_L^* \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega\tau} \tilde{f}_L^*(\omega) \tilde{\phi}_S(\omega) + h.c. \quad (7.55)$$

Now consider the special case that the signal is band-limited with bandwidth  $B$ , that is, it equals zero outside of a spectral interval  $[\nu - B/2, \nu + B/2]$  around a central frequency  $\nu$ . Further assume that the LO field is constant throughout this interval. Then we have (exactly) [94],



**Fig. 7.10.** Measured photon number distributions at the pulse peaks for SLD data in Fig. 7.9, (a) single-pass and (b) double-pass with grating. The solid curves are (a) Bose-Einstein distribution, and (b) negative-binomial distribution. From [62]

$$\begin{aligned} \hat{N}_-(\tau) &= -i\sqrt{c}\alpha_L^*\tilde{f}_L^*(\nu) \int_{\nu-B/2}^{\nu+B/2} \frac{d\omega}{2\pi} e^{-i\omega\tau} \tilde{\phi}_S(\omega) + h.c. \\ &= -i\sqrt{c}\alpha_L^*\tilde{f}_L^*(\nu)\phi_S(\tau) + h.c. \end{aligned} \quad (7.56)$$

This remarkable result shows that, by using an LO pulse having a finite duration and a Fourier spectrum  $\tilde{f}_L(\nu)$  that is constant in the interval containing the signal spectrum, a band-limited signal field can be sampled exactly, with a resolution that is not degraded by the finite duration of the LO pulse. An example of such an LO pulse is  $f_L(t) \propto (1/t) \sin(Bt/2)$ . This result is closely related to the Whittaker-Shannon sampling theorem [115], which is used here in an inverse manner. Our result is most useful for measuring repetitive signals whose form is the same from pulse to pulse.

### 7.3.4 Dual-LO BHD and Two-Mode (or Two-Time) Tomography

There is no reason that the LO pulse needs to have a simple single-pulse form. If the LO instead is comprised of two well isolated short pulses (with the same or differing spatial forms), then we may choose to view it as a linear superposition of two distinct wave packet modes. (The same idea applies to two polarization modes; see below.) In general the LO field is written as [9, 121–123]

$$\hat{\underline{f}}_L^{(+)}(\underline{r}, t) = i\sqrt{c}|\alpha_L| \exp(i\theta) [\underline{v}_1(\underline{r}, t) \cos \alpha + \underline{v}_2(\underline{r}, t) \exp(-i\zeta) \sin \alpha], \quad (7.57)$$

where  $\alpha$  and  $\zeta$  are parameters that can be varied, and the two mode functions are orthonormal. When this is inserted into Eq. (7.22) we find that the detected quadrature corresponds to the mode operator

$$\begin{aligned}
\hat{a} &= \sum_k \hat{a}_k c \int_0^T dt \int_{D_{et}} d^2x \underline{v}_k(\underline{x}, 0, t) \\
&\quad \times [\underline{v}_1^*(\underline{x}, 0, t) \cos \alpha + \underline{v}_2^*(\underline{x}, 0, t) \exp(i\zeta) \sin \alpha] \\
&= \hat{a}_1 \cos \alpha + \hat{a}_2 \exp(i\zeta) \sin \alpha,
\end{aligned} \tag{7.58}$$

where  $\hat{a}_1$  and  $\hat{a}_2$  are the operators for the components of the signal field in each of the modes of interest. Using  $\hat{a}_1 = (\hat{q}_1 + i\hat{p}_1)/\sqrt{2}$  and  $\hat{a}_2 = (\hat{q}_2 + i\hat{p}_2)/\sqrt{2}$ , we find that the measured quadrature  $\hat{Q} = (\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta})/\sqrt{2}$  is

$$\hat{Q} = \cos(\alpha) [\hat{q}_1 \cos \theta + \hat{p}_1 \sin \theta] + \sin(\alpha) [\hat{q}_2 \cos(\theta - \zeta) + \hat{p}_2 \sin(\theta - \zeta)]. \tag{7.59}$$

If we define  $\theta - \zeta = \beta$  then the bracketed terms are recognized as the phase-dependent quadratures defined in Eq. (7.26) (one for each mode), so we can write

$$\hat{Q} = \cos(\alpha) \hat{q}_{1\theta} + \sin(\alpha) \hat{q}_{2\beta}. \tag{7.60}$$

This shows that we can measure, in a single event, an arbitrary linear superposition of the quadratures of two modes using dual-LO BHD (but, of course, not the variable conjugate to  $\hat{Q}$ ). The phases  $\theta$  and  $\beta$  of the two LO fields, as well as  $\alpha$ , are independently variable by the experimenter.

An intriguing application of the two-mode tomography concept is in the reconstruction of two-time statistics of a signal field [43, 124]. It should be pointed out that each subsystem is measured only once and then discarded, so measurement-induced dynamics cannot be measured by this technique – the system evolves as if no measurements were made. In this sense the situation is the same as in the well-known Hanbury Brown-Twiss correlation measurements, but in principle we can measure the complete two-time state evolution, not just the correlation of two particular observables. Then, from the reconstructed two-time (i.e., two-mode) state one could calculate the correlation function of any two variables at the two times selected.

It has been shown that it is possible to reconstruct the joint quantum state of the combined two-mode system by measuring probability histograms for the combined quadrature  $Q$  for many values of the parameters  $\alpha$ ,  $\theta$  and  $\beta$  [121, 122]. Reconstruction in the number basis uses two-mode pattern functions [121].

If only the two-mode joint photon statistics are desired, then the two components of the combined LO field can have uniformly random phases, and a simpler two-mode pattern function can be used [121]. The theoretical scheme has also been generalized to the case of many modes [125].

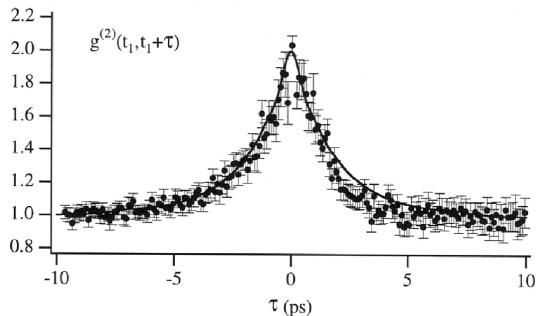
A complete two-mode state reconstruction, which uses two-mode pattern functions, has not been performed in a laboratory to our knowledge, illustrating the challenge this poses regarding the amount of data required and the technical difficulty of controlling all parameters in the experiment well enough during the data collecting time [126]. Nevertheless, the technique has been

applied to full reconstruction of two-mode photon-number statistics and correlations [57], as well as two-time number correlations on ps time scales [43]. Here we review the latter application.

The first experimental demonstration of ultrafast two-time number correlation measurements using dual-LO BHD was carried out by McAlister [43], who reconstructed the two-time second-order coherence defined (for a quasi-monochromatic field) as

$$g^{(2)}(t_1, t_2) = \frac{\langle : \hat{n}(t_1) \hat{n}(t_2) : \rangle}{\langle \hat{n}(t_1) \rangle \langle \hat{n}(t_2) \rangle}, \quad (7.61)$$

which generalizes Eq. (7.51). The signal source studied was a super luminescent laser diode (SLD, see above). While the LO was temporally synchronized with the 4-ns SLD pulse, there was no need for phase coherence between the signal and the LO since the signal field was intrinsically phase-random. In this special case the relative phase between the two LO pulses also is not important. The 150 fs LO pulses from a Ti:sapphire laser were split and, after a variable delay, recombined to make a dual-pulse LO in the form of Eq. (7.57) with adjustable  $\alpha$ , and random (or arbitrary)  $\theta$  and  $\beta$ . The second-order coherence  $g^{(2)}$  is reconstructed from combinations of second and fourth-order moments of  $Q$  obtained at three different values of  $\alpha$ : 0,  $\pi/4$ , and  $\pi/2$ . These three values correspond simply to a measurement of the first mode by itself, a measurement of the second mode by itself, and a measurement of an equal combination of both modes. Formulas are known also for estimating the statistical errors of the  $g^{(2)}$  measurement [124].



**Fig. 7.11.** Measured two-time photon number correlation from a pulsed SLD at low power. The solid curve is obtained from the Fourier transform of the measured optical spectrum of the source, under the assumption that the field is quasi-stationary and the statistics are thermal-like. [109]

In Fig. 7.11 we show the SLD data from the thesis of McAlister [109].

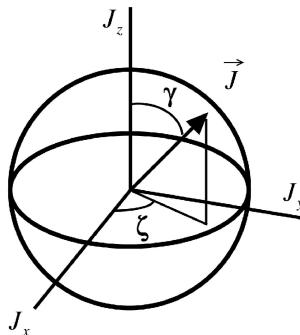
### 7.3.5 Optical Polarization Tomography

A dual-LO scheme for the purpose of Optical Polarization Tomography was proposed and analyzed in [127]. This can reconstruct the polarization state of a beam-like spatial mode comprised of two orthogonal polarization modes. Rather than requiring a full reconstruction of the two-mode state (density matrix), the measurement of the quantum state of polarization requires reconstruction only of a subset of the density matrix that we called the “polarization sector.” This is defined to be the number-basis elements  ${}_1\langle n_1|{}_2\langle n_2|\hat{\rho}|n'_1\rangle_1|n'_2\rangle_2$ , restricted to  $n_1 + n_2 = n'_1 + n'_2$ , where  $n_1$  and  $n_2$  refer to the number of photons in each polarization. Knowledge of this portion of  $\hat{\rho}$  is sufficient to calculate any statistical moments of the polarization Stokes operators, which are defined by  $\hat{J}_1 = (\hat{a}_1^\dagger \hat{a}_1 - \hat{a}_2^\dagger \hat{a}_2)/2$ ,  $\hat{J}_2 = (\hat{a}_1^\dagger \hat{a}_2 + \hat{a}_2^\dagger \hat{a}_1)/2$ ,  $\hat{J}_3 = (\hat{a}_1^\dagger \hat{a}_2 - \hat{a}_2^\dagger \hat{a}_1)/(2i)$ . In terms of the angular momentum eigenstates  $|j, m\rangle$  of  $\hat{J}^2$  and  $\hat{J}_1$ , the polarization sector corresponds to the elements  $\langle j, m|\hat{\rho}|j, m'\rangle$  for all  $j, m, m'$ .

The SU(2) group of unitary transformations on the  $\hat{J}_i$  operators is equivalent to the general two-mode transformation of the mode operators  $\hat{a}_1, \hat{a}_2$  to give new operators  $\hat{a}_3, \hat{a}_4$ ,

$$\begin{aligned}\hat{a}_3 &= \hat{a}_1 \cos(\gamma/2) + \hat{a}_2 \exp(i\zeta) \sin(\gamma/2), \\ \hat{a}_4 &= -\hat{a}_1 \sin(\gamma/2) + \hat{a}_2 \exp(i\zeta) \cos(\gamma/2),\end{aligned}\quad (7.62)$$

which preserves the commutator,  $[\hat{a}_3, \hat{a}_4^\dagger] = 0$ . Note that, with the identification  $\gamma/2 = \alpha$ , the operator  $\hat{a}_3$  is the same as  $\hat{a}$  defined in Eq. (7.58) and measured using dual-LO BHD. On the Poincaré sphere representation of the  $\hat{J}_i$  operators, the  $\hat{J}$  vector is rotated by angles  $\gamma$  and  $\zeta$  when the  $\hat{a}_i$  operators undergo the transformation Eq. (7.62), as illustrated in Fig. 7.12.



**Fig. 7.12.** Poincaré sphere representation of the optical polarization state.

It was shown in [127] that full reconstruction of the polarization sector can be accomplished by using a dual-polarization-mode LO of the form

Eq. (7.57), with the LO components having identical spatial-temporal forms but orthogonal polarizations. In this case the overall (common-mode) phase  $\theta$  can be randomized or swept uniformly over  $0 - 2\pi$ . This arises because polarization depends only on the relative phase between two modes, and significantly reduces the amount of data collection required for a reconstruction, compared to a full two-mode state reconstruction.

As shown by Bushev et al., a complete characterization of the quantum state of optical polarization (but not the complete two-mode state) can also be obtained by using direct photoelectric detection rather than homodyne detection [128]. One does this by measuring joint photoelectron statistics of the photon numbers of the two polarization modes following many different polarization-state transformations. From measured histograms the state is extracted in the form of a polarization Wigner distribution on the Poincaré sphere, whose marginal distributions give probabilities for a general polarization (Stokes) variable, from which arbitrary moments can be computed [129, 130]. Experimental results for polarization-squeezed light were reported [128].

If only first and second moments of the Stokes operators are desired, a simpler scheme can be used, in which only three mode transformations are needed before photoelectron statistics are collected [131]. This method is employed in a modified manner in the experiments discussed below in Sec. 7.3.7.

A smaller polarization sector can be used if it is known *a priori* (or determined by post selection) that the number of photons in a given set of modes is limited to a fixed number. For example, the recent experiments by White et al. [21], reconstructing the quantum state of polarization of pairs of photons, correspond to measuring a small subset of a polarization sector for a four-mode density matrix (two polarization modes for each photon). See this volume chapter by Altepeter, James and Kwiat.

### 7.3.6 Two-Mode Tomography by Generalized Rotations in Phase Space (GRIPS)

It was pointed out in [132] and Richter [133] that in some cases it may be easier to obtain tomographic state information about a pair of modes by using a fixed, single-mode (single-wave packet) LO field and implementing a two-mode transformation on the signal field [rather than on the LO field as in Eq. (7.57)]. The most general unitary transformation [SU(2)] on a pair of mode operators  $\hat{a}_1, \hat{a}_2$  defines new operators  $\hat{a}_3, \hat{a}_4$  as given in Eq. (7.62) above. The idea is to measure quadrature statistics for operator  $\hat{a}_3$  under many different transformation conditions. This yields enough information to reconstruct the state. Note again that the operator  $\hat{a}_3$  is the same as  $\hat{a}$  defined in Eq. (7.58).

For our purposes only  $\hat{a}_3$  needs to be measured. This is done using a linear-optical device, consisting of a pair of controllable birefringent phase retarders followed by a polarizing beam splitter (PBS) to separate the resulting modes

$\hat{a}_3$  &  $\hat{a}_4$ . Mode  $\hat{a}_3$  is created such that it has the same polarization as the LO. It and the LO enter the DC balanced-homodyne detector, which linearly combines (interferes) mode  $\hat{a}_3$  and the LO mode having phase  $\theta$ . After subtraction and scaling of the photodetectors' signals, the quantity measured on a single trial is the “combined quadrature amplitude,”

$$\begin{aligned}\hat{Q}(\gamma/2, \theta, \zeta) &= [\hat{a}_3 \exp(-i\theta) + \hat{a}_3^\dagger \exp(i\theta)]/2^{1/2} \\ &= \cos(\gamma/2)\hat{q}_{1,\theta} + \sin(\gamma/2)\hat{q}_{2,\theta-\zeta}.\end{aligned}\quad (7.63)$$

This shows that by varying the parameters  $\gamma$ ,  $\theta$  and  $\zeta$  the experimenter can measure the combined quadrature amplitude of the two-mode signal following a rotation by arbitrary angles on the Poincaré sphere. This provides a tomographically complete set of observables, whose statistical characterization allows reconstruction of the full quantum state by using the two-mode pattern functions mentioned in Sec. 7.3.4 above.

As before, it is possible to obtain directly certain field moments or number moments by measuring quadrature data at only a small set of well chosen values of  $\gamma$ ,  $\theta$  and  $\zeta$  [43]. For a summary of various moment formulas, see [109].

### 7.3.7 Ultrafast Optical Polarization Sampling by GRIPS Tomography

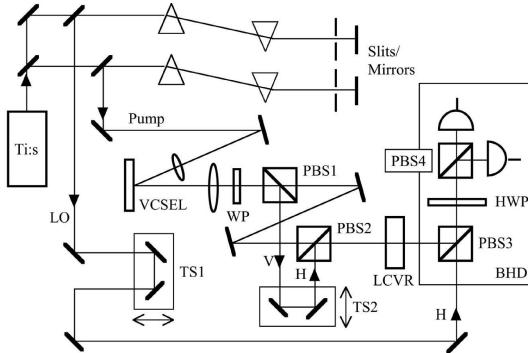
The statistical correlation between two orthogonal polarization modes ( $i$  and  $j$ ) with photon numbers  $n_i(t_i)$  and  $n_j(t_j)$  at two times  $t_1$  and  $t_2$  is characterized by the normalized two-mode, two-time, second-order coherence (correlation) function

$$g_{i,j}^{(2)}(t_1, t_2) = \frac{\langle : \hat{n}_i(t_1) \hat{n}_j(t_2) : \rangle}{\langle \hat{n}_i(t_1) \rangle \langle \hat{n}_j(t_2) \rangle}. \quad (7.64)$$

(The normal ordering is unnecessary unless  $i = j$ .) A value  $g_{i,j}^{(2)} = 1$  indicates uncorrelated fluctuations in  $n_i(t_i)$  and  $n_j(t_j)$ , and a value above (below) 1 indicates positive (negative) correlations. Such a quantity can be measured by the sampling schemes discussed above—either the dual-LO or the GRIPS scheme.

A version of the GRIPS scheme allows two-mode correlations to be measured without any assumption about the photon number. The analysis of optical polarization follows an earlier proposal by Karassiov and Masalov (KM) [131], and was implemented by Blangett, et al., for the purpose of studying ultrafast polarization dynamics of 30-ps pulses from a vertical-cavity surface-emitting semiconductor laser (VCSEL) [50, 134].

In the resulting “two-time optical polarization sampling method,” shown in Fig. 7.13, Blangett employed the idea of KM to measure the signal beam separately in three distinct polarization bases—R/L, H/V, or +45/-45, where R (L) means right (left) circular, H (V) means horizontal (vertical), and vertical +45 (-45) means + (-) 45 degree linear polarizations. To accomplish this



**Fig. 7.13.** Setup for sampling scheme for one- or two-time optical polarization correlations. TS—translation stages; WP—wave plate (optional); HWP—half-wave plates; LCVR—liquid-crystal-variable-retarder; BHD—balanced-homodyne detector. [134]

basis resolution, the signal beam emitted by the VCSEL may pass through one or two waveplates (or it may pass through no waveplates.) The first is a quarter-wave plate (QWP), which converts R circularly polarized light into V linearly polarized light, and L polarized light into H polarized light. The second is a half-wave plate (HWP) which converts +45 deg polarization into V polarization, and -45 deg linear into H. KM pointed out that measuring means and fluctuations of intensity separately in all three bases allows one to characterize fully the polarization statistics up to second order, that is, the first and second moments of the Stokes operators. Blangett's generalized scheme also allows measuring the correlations between orthogonally polarized modes, which cannot be extracted from moments of the Stokes operators alone [134].

Whereas in the KM scheme fast photodetectors would ordinarily be used to make time-resolved measurements of intensity in the three bases, Blangett adopted ultrafast, phase-averaged BHD (as described in section 7.3.3) to sample these statistics on sub-ps time scales. This is accomplished by splitting the signal after the WP using a polarizing beam splitter (PBS1) into V and H-polarized beams. These two are recombined at PBS2 with near-zero (few fs) time delay after the V component has had its phase shifted by a movable mirror (TS2) whose purpose is to sweep uniformly over  $0 - 2\pi$  during the data acquisition of quadrature histograms (typically requiring 20,000 pulses). The LO overall phase is also uniformly swept by moving translation stage TS1. The phase sweeping allows the use of the simpler two-mode tomographic reconstruction scheme mentioned above.

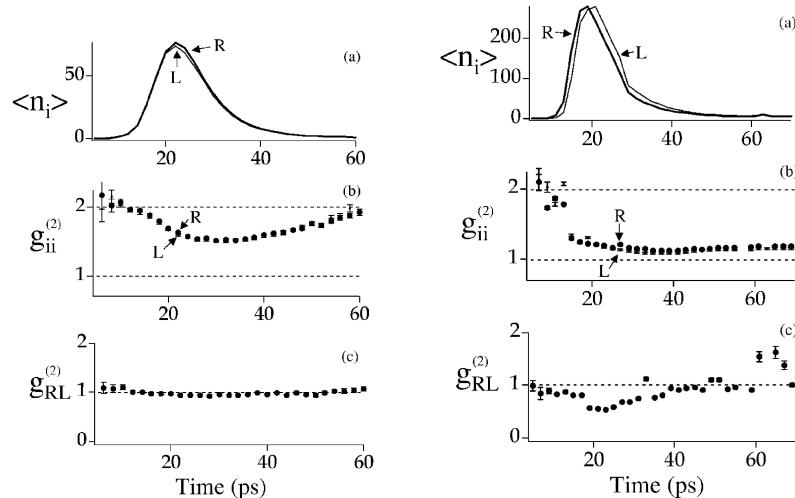
Note that for this experiment the VCSEL emission has a longer wavelength than does the laser pulse that pumps the VCSEL. The pump pulse and the LO pulse, which must have the same wavelength as the VCSEL signal,

are derived from the wide-band Ti:sapphire pulse by use of prism-slit-based spectral filters, and both had durations of 300 fs.

In order to select the two-mode quadrature to be measured in the BHD, a device (liquid-crystal variable-phase retarder, LCVR) is used as a computer-controllable half-wave plate to rotate the linear polarizations exiting PBS2 by 90 deg, 45 deg, or 0 deg. Then PBS3 reflects only V polarization into the path of the H-polarized LO beam. Finally, a HWP rotates by 45 deg and PBS4 projects out + and - linear superpositions of LO and combined-signal beam. The choice of rotation angle induced by the LCVR sets the value of  $\gamma$  (0,  $\pi/2$ , or  $\pi$ ), in the combined quadrature

$$Q = \cos(\gamma/2)q_{1,\theta} + \sin(\gamma/2)q_{2,\theta-\zeta}, \quad (7.65)$$

and the phases  $\theta$  and  $\zeta$  are swept (and averaged over) in the manner described above. The moments are calculated using the formulas given in [109]. Statistical error bars are calculated according to formulas in [124].



**Fig. 7.14.** Mean photon number  $\langle n \rangle$  and correlations, with H-polarized pump, for (left) low-temperature (10 K) VCSEL and (right) room-temperature VCSEL [134].

In the measurements shown in Fig. 7.14, Blansett used a near-zero time delay in TS2 so the scheme measures a one-time, two-polarization correlation and corresponds to the GRIPS method described above. For the low-temperature VCSEL emission shown on the left, the R and L emission modes show  $g_{RR}^{(2)} = g_{LL}^{(2)} \cong 2$  at early and late times, corresponding to spontaneous emission, and  $g_{RR}^{(2)} = g_{LL}^{(2)} \cong 1.5$  at the emission peak, indicating lasing or stimulated emission, with photon statistics tending toward a Poisson number distribution.  $g_{RL}^{(2)} \cong 1$  shows that the R and L modes emit in uncorrelated

fashion. This is in contrast to results obtained with a room-temperature VC-SEL, shown on the right, which shows anticorrelated R and L intensities  $g_{RL}^{(2)} < 1$ . Theoretical modeling reveals that this anticorrelation is caused by a higher spin-flip rate at the higher temperature, leading to gain competition between modes. This new measurement capability allows unprecedented time resolving of such lasing dynamics and statistics, described in detail in [50].

In the second set of measurements (not shown) Blansett used a non-zero time delay in TS2, so the scheme measures a two-time, two-polarization correlation.

### 7.3.8 Simultaneous Time and Frequency Measurement

An important property of the dc-balanced homodyne technique is that it provides spectral as well as temporal information about the signal field. This arises because if the LO field is frequency-tuned [by harmonically varying the function  $f_L(t)$ ] away from the spectral region of the signal, the integral in Eq. (7.22) defining the mode-matched amplitude  $\hat{a}$  will decrease. To analyze this, define the LO temporal function in Eq. (7.53) to be

$$f_L(t) = e^{-i\omega_L t} h_L(t - t_L), \quad (7.66)$$

where  $\omega_L$  is the LO's center frequency and  $h_L(t - t_L)$  is a real function with maximum at  $t = t_L$ . One way to achieve this would be by generating an ultrashort pulse (e.g., 30fs) and passing it through a tunable bandpass filter, followed by a time-delay. Using the reconstructed joint statistics of quadrature amplitudes, the mean number of mode-matched signal photons  $\bar{N}_{MMS}(\omega_L, t_L) = \langle \hat{a}^\dagger \hat{a} \rangle$  can be determined for a set of  $\omega_L$  and  $t_L$  values, as demonstrated in [41]. In the semiclassical case, this quantity, for a given LO center frequency  $\omega_L$ , is equal to

$$\bar{N}_{MMS}(\omega_L, t_L) = \left\langle \left| \int_0^T dt e^{i\omega_L t} h_L(t - t_L) \phi^{(+)}(t) \right|^2 \right\rangle, \quad (7.67)$$

where  $\phi^{(+)}(t)$  is the spatial-mode-matched signal,

$$\phi_S(t) = \int_{Det} d^2x v_L^*(x) \cdot \Phi_S^{(+)}(\underline{x}, 0, t). \quad (7.68)$$

and the brackets indicate an average over multimode coherent-state amplitudes. Equation (7.67) is identical to the general form for time-dependent spectra [135], with  $h_L(t - t_L)$  acting as a time-gate function. For example, it can be put into the same form as appearing in the “time-dependent physical spectrum” [136–138] if we specialize to

$$h_L(t - t_L) = \begin{cases} 0, & t > t_L \\ e^{-\gamma(t-t_L)}, & t < t_L \end{cases}. \quad (7.69)$$

Thus by scanning the LO center frequency  $\omega_L$  and arrival time  $t_L$  independently, and measuring  $\bar{N}_{MMS}(\omega_L, t_L)$ , one obtains both time and frequency information, within the usual time-frequency bandwidth limitations. This method provides an alternative to the nonlinear optical upconversion technique which has been used to measure time-frequency information for light emitted by vibrational molecular wave packets [22].

## 7.4 Experimental Techniques

Detection circuits for balanced detection fall into two major classes: radio frequency (RF) detection and charge-sensitive whole-pulse detection (which we refer to as DC detection). Recently array detection has been used for quantum state measurements [31–33]; array detection is a generalized version of DC detection. We will discuss each of these detection technologies, with an emphasis on DC detection.

We also note that quantum-state measurement of electromagnetic fields has been performed with single-photon counting detectors [19], and by probing an optical field in a cavity with Rydberg atoms [56]. These state measurement techniques are very different from OHT, and we will not discuss them in detail.

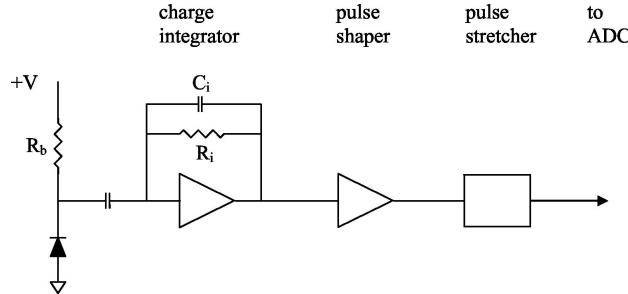
### 7.4.1 DC Detection

DC detection was used by Smithey et al. in the first demonstration of OHT [16]. It has been the workhorse for our experiments ever since, and it is the system we have the most familiarity with. To our knowledge, Guéna et al. were the first to operate such a detector at the shot-noise level (SNL) [100], while the first detection of nonclassical light using this technology was performed by Smithey et al. [99]. Other implementations of this technology have been performed by Hansen et al. [51] and Zavatta et al. [52].

In this detection technique a short pulse of light (usually much shorter than the response time of the detector) is incident on a photodiode and produces a charge pulse. On average, the number of photoelectrons in the output pulse is equal to the number of incident photons times the quantum efficiency  $\eta$  of the photodiode. The goal is to measure the number of photoelectrons produced by each pulse as precisely as possible.

In order to be digitized by an analog-to-digital converter (ADC) and stored in a computer, a charge pulse must be converted to a voltage pulse and then amplified. A charge-sensitive preamplifier (also known as a charge integrator) followed by pulse-shaping electronics does this. The electronics used are common in nuclear spectroscopy or x-ray detection. We refer to this detection technique as DC because the preamplifiers integrate the total charge per pulse, and we sample the data synchronously with the pulse repetition frequency. Every photoelectron entering the amplifier is counted; we

do not measure the current within some bandwidth about a nonzero offset frequency.



**Fig. 7.15.** Circuit diagram for a DC detection system.

A schematic of the electronics is shown in Fig. 7.15. The photodiode is reverse biased with a DC voltage; the input coupling capacitor blocks this DC bias and passes the short (a few ns) current pulse produced when an optical pulse hits the photodiode. The charge integrator integrates the input charge, and puts out a voltage pulse with a rise time on the order of a few nanoseconds, and a fall time on the order of 100's of microseconds. The peak voltage of this output pulse is proportional to the total input charge; the fall time is determined by the time constant of the integrator  $R_i C_i$ . The pulse shaper (also known as a spectroscopy amplifier) converts this highly asymmetrical pulse to a nearly Gaussian-shaped output pulse having a width on the order of 1  $\mu$ s. The shaper also further amplifies the pulse; the peak voltage of the output pulse from the shaper is proportional to the input charge (calibration of the proportionality constant will be discussed below.) If a fast ADC is available then the ADC can directly sample the pulse from the shaper; otherwise the output of the shaper goes to a stretching amplifier, which stretches the Gaussian pulse into a rectangular-shaped output pulse, having a width on the order of 10  $\mu$ s, which can then be easily sampled by an ADC.

Charge-sensitive preamplifiers, spectroscopy amplifiers and pulse stretchers are commercially available from vendors that manufacture electronics for nuclear spectroscopy. Vendors we have used include:

- Amptek Inc; Bedford, MA; [www.amptek.com](http://www.amptek.com).
- eV Products Inc.; Saxonburg, PA; [www.evproducts.com](http://www.evproducts.com).
- Ortec; Oak Ridge, TN; [www.ortec-online.com](http://www.ortec-online.com).
- Canberra Industries; Meriden, CT; [www.canberra.com](http://www.canberra.com).

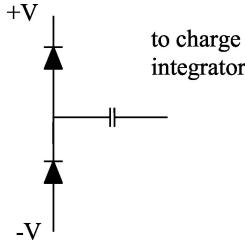
A primary objective in building a system is to achieve low electronic noise – a target of 10 dB below the shot noise is suitable. One source of noise is Johnson noise in the resistors. Another is series resistor noise in the field-effect

transistor (FET) channel of the charge-sensitive preamp. This noise depends on properties of the FET (such as the carrier transit time through the channel), the input capacitance (usually dominated by the intrinsic capacitance of the detector), and the shaping time of the amplifier. For detailed discussions see [139, 140]. For a fixed FET and shaping time, lowering the input capacitance lowers the noise. For a fixed capacitance, it is possible to choose the proper FET and shaping time to lower the noise. Consulting the specifications of the charge-sensitive preamp can help make these choices. Depending on the detector capacitance (typically 10 pF), amplifier gain (typically  $10^{-6}$  V/e $^-$ ) and shaping time (typically 1  $\mu$ s), typical noise levels range from 200–900 electrons rms (root-mean-square). To achieve such performance for single-pulse measurements, the pulse repetition rate must be less than the inverse of the shaping time, preventing operation faster than about 1 MHz with current technology.

The system shown in Fig. 7.15 contains a single photodiode. Since two detectors are needed for balanced detection, an identical second system is needed. When using two separate detectors once can measure the number of photoelectrons in each detector,  $N_1$  and  $N_2$ , and then use these to calculate the photoelectron sum  $N_+ = N_1 + N_2$  and difference  $N_- = N_1 - N_2$ . As discussed in Sec. 7.2 above, the scaled difference number yields the quadrature amplitude, which is the quantity of interest in performing OHT. The scaling factor is proportional to the average of the total number of photoelectrons, which is simply given by  $\langle N_+ \rangle$ . Thus, using two separate detectors makes determination of the scaling factor quite easy; it is determined *in situ* as the quadrature amplitude data is collected.

The price paid for this convenience is that the ADC must be capable of simultaneously recording the outputs from two detectors. It must also be capable of resolving the number of photoelectrons on a given channel, e.g.  $N_1$ , with a resolution better than the standard deviation of the fluctuations in that channel, (i.e., to better than the square-root of the SNL  $\langle (\Delta N_1)^2 \rangle^{1/2} = \langle N_1 \rangle^{1/2}$ , where  $\Delta N_1 \equiv N_1 - \langle N_1 \rangle$ .) For example, suppose a detector has an electronic noise level of 300 electrons rms. In order that the signal dominate this noise, we must have the shot noise fluctuations of the signal much larger than this;  $\langle N_1 \rangle^{1/2} = 1200$  photoelectrons rms is sufficient, as this corresponds to a shot-noise variance 12 dB above the electronic noise variance. The average number of photoelectrons needed to achieve this is  $\langle N_1 \rangle = 1.44 \times 10^6$ . Thus, the resolution of the ADC must be greater than  $\langle N_1 \rangle / \langle N_1 \rangle^{1/2} = \langle N_1 \rangle^{1/2}$ . In this example a 12-bit digitizer having a resolution of 1 part in 4,096 is not adequate. Typically 16-bit digitizers are required when independent measurements of two detectors are performed.

An alternative to the simultaneous sampling of two channels is first to subtract the photodiode currents, and then amplify and digitize. In this case the single reverse-biased photodiode shown in Fig. 7.16 is replaced with two photodiodes as shown in Fig. 7.16. The difference is that instead of detecting



**Fig. 7.16.** Photodiode electrical configuration for subtraction before integration.

two large numbers and then subtracting, the difference number  $N_-$  is directly detected, so lower-resolution digitizers are sufficient (12-bit resolution is more than adequate.)

The disadvantage of subtracting first is that it is no longer possible to calibrate the SNL *in situ*. It is necessary first to block the signal beam and measure the variance of the difference number  $\langle(\Delta N_-)^2\rangle$  with only the LO incident on the photodetectors. If the (classical) noise fluctuations of the LO are not too large, and the precision of subtraction is adequate to suppress them, then  $\langle(\Delta N_-)^2\rangle$  should be equal to the SNL. To verify that this is the case, it is standard practice to block one of the detectors and measure the average number of photoelectrons incident on the other detector  $\langle N_1 \rangle$ . If the detectors are well balanced, the average number of photons hitting the second detector should be the same, so the total number of detected photoelectrons is  $2\langle N_1 \rangle$ . As long as  $\langle(\Delta N_-)^2\rangle = 2\langle N_1 \rangle$  one is confident that the detectors are operating at the shot-noise limit.

In the above we have been discussing measurements of photoelectron number  $N$ , however the quantity measured directly by the ADC is a voltage  $V$ . In order to obtain  $N$  from  $V$  it is necessary to measure the gain  $g$  of the detection system, such that

$$N = gV. \quad (7.70)$$

There are two methods to determine the gain, and it is standard practice to use both methods and to ensure that they are consistent.

In the first method the input to the charge-sensitive preamp is replaced by a test capacitor connected to a voltage pulser. By applying a voltage pulse of known height  $V_p$ , a known charge  $Q = CV_p$  is delivered to the amplifier and its output voltage  $V$  is measured. The gain of the amplifier is then given simply by

$$g = \frac{Q}{V} = \frac{CV_p}{V}. \quad (7.71)$$

The accuracy of this calibration is limited to about 5 or 10% by the accuracy of the calibration of the capacitance.

The second method simultaneously provides measures of the gain and the electronic noise, as well as verifying that the detector is operating at the SNL. If a balanced homodyne detector makes measurements at the SNL, then

$$\langle (\Delta N_-)^2 \rangle = \langle N_+ \rangle + \sigma_e^2, \quad (7.72)$$

where we have added the variance of the electronic noise  $\sigma_e^2$ .

The ADC measures the voltage outputs from the two detectors,  $V_1$  and  $V_2$ . We define the sum and difference voltages as  $V_+ = V_1 + V_2$  and  $V_- = V_1 - V_2$ . Using the fact that the measured voltage  $V$  is proportional to the photoelectron number, Eq. (7.70), we can rewrite Eq. (7.72) as

$$g^2 \langle (\Delta V_-)^2 \rangle = g \langle V_+ \rangle + \sigma_e^2, \quad (7.73)$$

or

$$\langle (\Delta V_-)^2 \rangle = \frac{1}{g} \langle V_+ \rangle + \frac{1}{g^2} \sigma_e^2. \quad (7.74)$$

Thus, by varying the LO pulse energy and plotting  $\langle (\Delta V_-)^2 \rangle$  vs.  $\langle V_+ \rangle$  we should obtain a linear plot. The slope of the line determines the gain, and the intercept determines the electronic noise. If the plot is not linear then the detector is not operating at the SNL. Data illustrating this calibration procedure can be found in Ref. [62].

#### 7.4.2 RF Detection

The majority of balanced detection circuits fall into the class of radio frequency RF detectors. This class was developed for the detection of squeezed light, and have been widely used ever since [28,58]. Here the current fluctuations from the photodetectors are electrically filtered to within some bandpass about a nonzero center frequency. The filtering is frequently done using either an RF spectrum analyzer or an RF mixer in combination with a filter, and the center frequency is typically in the range of 1 to 10's of MHz. For discussions of some of the technical aspects of operating an RF detector, the reader is referred to articles by Machida and Yamamoto [30], and Wu et al. [141].

One difference between DC and RF detection is in the quantity that is measured. As discussed above, the quantity measured, after scaling, in DC balanced detection is the rotated quadrature amplitude of a particular mode of the signal field

$$\hat{q}_\theta = \frac{1}{\sqrt{2}} (\hat{a} e^{-i\theta} + \hat{a}^\dagger e^{i\theta}). \quad (7.75)$$

In RF detection the LO has an optical frequency  $\omega$ , and the detector current is analyzed with a bandpass filter centered at the radio frequency  $\Omega$ . The detected RF can come from beating between the LO and signal fields at optical frequencies of  $\omega \pm \Omega$ . Thus, the detected signal arises from combinations of optical modes at two different frequencies. The measured quadrature operator is then [42]

$$\hat{q}_\theta = (\hat{a}_{\omega+\Omega} e^{-i\theta} + \hat{a}_{\omega-\Omega}^\dagger e^{i\theta}) + (\hat{a}_{\omega+\Omega}^\dagger e^{i\theta} + \hat{a}_{\omega-\Omega} e^{-i\theta}). \quad (7.76)$$

If the LO is pulsed, and therefore not monochromatic, this formula must be generalized by integrating over  $\omega$  [67].

### 7.4.3 Balancing the BHD system

One of the more difficult, but important, aspects of performing OHT is achieving balanced detection at the SNL. If the detector noise is not shot-noise limited then one cannot measure the true quantum state. An excellent test for this is to block the signal beam completely, thus creating a signal in a vacuum state; then experimentally reconstruct this vacuum state and see how well it corresponds to a true vacuum state.

Fluctuations of the LO intensity (beyond the expected Poisson fluctuations) cause the detector noise to rise above the SNL. Balanced detection can largely eliminate the deleterious effects of these fluctuations if everything is working properly. Noise levels above the SNL are often due to poor subtraction between the two detectors. This can be caused by poor alignment of the beams on the detector faces, or by unequal splitting of the LO on the beam splitter. One must ensure that all of the light leaving the beam splitter is collected by the photodiodes. Nearly perfect 50/50 splitting of the LO beam is also necessary to achieve high subtraction efficiency. One method for achieving this is to use a combination of a  $\lambda/2$ -plate and a polarizing beam splitter as a 50/50 beam splitter [142]. Rotation of the  $\lambda/2$ -plate allows one to adjust the splitting ratio of the beam splitter. This issue is more critical for DC detection than for RF detection.

As discussed in Sec. 7.2, detection quantum efficiency (QE) plays an important role in OHT, as losses degrade information about the measured state. One wants to use the highest-QE detectors possible; photodiodes with efficiencies of over 85% are easily available in the near-IR portion of the spectrum. Furthermore, the homodyne efficiency of the system must be large. Homodyne efficiency is a measure of the overlap of the spatial-temporal field modes of the signal and LO beams [see Eq. (7.36)]. This can be maximized by careful alignment, however, for signal fields created by nonlinear processes it is nearly impossible to perfectly overlap the signal and LO modes [143–145]. One method of getting around this is to generate an LO using the same nonlinear process, so that it is matched to the signal mode [146]. Array detection can also be used to circumvent some of the losses due to mode-mismatch; we discuss this alternative in the following section.

When studying light with a nonzero mean amplitude (such as coherent states or bright squeezed states) using the DC whole-pulse technique, it is crucial to achieve extremely precise balancing between the two detection channels. (This is less important when using RF techniques.) In order to do this, both stages of the two detection chains need to be equalized—the effective quantum efficiency (QE) of the two photodiodes, and the overall electronic gain must be the same for each channel. The effective QE's can be matched by introducing a small variable optical loss in front of the detector with the highest intrinsic QE (e.g., by using a rotated glass Brewster plate).

A procedure for setting the gain of the amplifiers used with the detectors is as follows [147]. As described above, voltage pulses of known amplitude

are used in conjunction with capacitors to introduce a known charge into the input of the amplifiers. Since two detectors are being calibrated simultaneously, an input pulse is sent through a  $50\ \Omega$  voltage divider with a nominal splitting of 50% to produce two outputs. The voltage divider can be adjusted to change the ratio of the outputs from 1:1 to slightly more or less. High quality ceramic chip microwave grade capacitors are used to obtain proper operation over large bandwidths. Semi-rigid SMA cables and connectors are used to provide stability of the intrinsic capacitance of the connectors themselves when the connectors are subjected to mechanical strains such as when the cables are connected and disconnected. The intrinsic capacitance of the SMA cables and connectors changes much less than the capacitance of the test capacitors. Since the gain of the preamps change with the capacitance at the input, the photodiodes are left in place during the electronic calibration.

If we have two charge pulses, one from each test capacitor, with charges  $Q_1$  and  $Q_2$ , the voltage measured on each channel at the computer will be proportional to the charge contained in the pulse. Therefore,  $V_1 = \alpha Q_1$ ,  $V_2 = \beta Q_2$ , where  $V_1(V_2)$  is the voltage measured by the computer on channel 1(2) with  $\alpha(\beta)$  the overall conversion gain from the charge to the voltage for channel 1(2). Note that  $Q_1$  and  $Q_2$  are derived from the same voltage pulse. If we interchange the inputs to the charge preamplifiers, then the voltages at the computer will be given by  $V_1 = \alpha Q_2$ ,  $V_2 = \beta Q_1$ . If in both cases the difference between the voltages is equal to zero, i.e.  $V_1 - V_2 = \alpha Q_1 - \beta Q_2 = \alpha Q_2 - \beta Q_1 = 0$ , then it is necessary that  $\alpha = \beta$  and  $Q_1 = Q_2$ .

Experimentally, to set the conditions such that  $V_1 - V_2 = \alpha Q_1 - \beta Q_2 = \alpha Q_2 - \beta Q_1 = 0$  are satisfied, the pulse generator is set to a voltage that provides approximately  $10^6$  electrons to each preamplifier, and the gains are adjusted so that the difference between the two channels at the computer is zero. At this point, we do not know if  $\alpha = \beta$ , because we do not know if the input charges are the same.

Next the inputs to the preamplifiers are switched so that the difference measured at the computer is  $V_1 - V_2 = \alpha Q_2 - \beta Q_1$ . Assuming that for the previous step,  $Q_1 \neq Q_2$ , we then have that  $V_1 - V_2 \neq 0$ . The  $50\Omega$  variable voltage splitter is now adjusted, changing the amount of charge sent to each preamplifier until the difference measurement is zero,  $V_1 - V_2 = 0$ . The inputs to the preamplifiers are switched back to the original configuration and the voltage difference is once again measured. The process of adjusting the gain, switching the inputs, adjusting the voltage splitter, and switching the inputs back is iterated until the difference measurement is as close to zero as possible for the configuration of both inputs, with no adjustments to the gains or the voltage splitter.

We can calculate how closely the gains can be set equal to each other given a finite (but small) difference number instead of a difference number of zero as assumed above. The above relations are replaced by  $V_1 - V_2 = \alpha Q_1 - \beta Q_2 = n_{diff1}$  and  $V_1 - V_2 = \alpha Q_2 - \beta Q_1 = n_{diff2}$  for the two connection

configurations. It is straightforward to show that for small difference numbers,  $\alpha$  and  $\beta$  can be set equal to within a precision  $(n_{diff1} + n_{diff2})/n_{tot}$  where  $n_{tot}$  is the total number of photoelectrons. Experimentally, for a total number of electrons  $n_{tot} = 10^6$  it is relatively easy to achieve difference numbers  $n_{diff1}, n_{diff2} \approx 10^2$ . Thus it is possible to achieve  $\alpha = \beta$  to within 1 part in  $10^4$ .

## 7.5 Array Detection

The use of array detectors for measuring quantum phase distributions was suggested by Raymer et al [148]. Their use for measuring quantum states was analyzed by Beck [149] and by Iaconis et al. [150]. Experiments to measure density matrices [31, 32] and Husimi distributions (Q-functions) [33] using arrays have been performed. We note that the use of single-photon counting arrays has been suggested for state measurement [151], but this falls under the category of photon-counting methods [152, 153], not OHT; the arrays required for this are quite different from the arrays used in experiments to date.

Array detection is a form of DC detection technology in which the individual detectors used in the balanced homodyne detector are replaced by array detectors. These arrays have many pixels, and resolve the transverse intensity profiles of the beams that illuminate them. The ability to resolve the transverse structure offers several advantages: increased effective detection efficiency, the ability to measure many spatial modes simultaneously, the ability to find the mode that meets a particular definition of “optimal”, and even the ability to perform unbalanced homodyne measurements at the SNL (i.e., to use a single output port from a beam splitter instead of subtracting the outputs of two ports.) A disadvantage is that array detection is slow using current technology.

### 7.5.1 Array Detection of Spatial Modes

In Sec. 7.2 we considered an expansion of the signal and local oscillator fields in terms of spatial-temporal modes  $v_k(\underline{x}, 0, t)$ . Here we are mainly interested in the spatial part of the mode function, and for simplicity will assume that the spatial and temporal parts of the mode function factorize. We’ll furthermore assume that the signal and LO temporal mode functions, as well as their polarizations, are perfectly matched. In this case we can perform the time integral in Eq. (7.21), and with proper normalization and scalar, co-polarized modes, we are left with

$$\hat{N}_- = -i \int_{Det} d^2x v_L^*(\underline{x}, 0) \hat{\Phi}_S^{(+)}(\underline{x}, 0) + h.c. \quad (7.77)$$

After time integration the orthogonality condition for the mode functions [Eq. (7.19)] becomes

$$\int_{\text{Det}} d^2x v_k^*(\underline{x}, 0) v_m(\underline{x}, 0) = \delta_{km}. \quad (7.78)$$

In Sec. 7.2 we expanded the LO and signal modes in the same set of mode functions, in order to demonstrate that the signal is projected onto the mode of the LO when using single detectors. This is not the case with array detection, however, so it is convenient to expand the LO and signal using separate mode functions. For the LO modes we'll use the mode functions  $v_k(\underline{x}, 0)$  described above, while for the signal we'll use  $w_k(\underline{x}, 0)$ . The  $w_k$ 's satisfy the same orthogonality condition as the  $v_k$ 's:

$$\int_{\text{Det}} d^2x w_k^*(\underline{x}, 0) w_m(\underline{x}, 0) = \delta_{km}. \quad (7.79)$$

While the  $v_k$ 's are orthogonal and the  $w_k$ 's are orthogonal, the  $v_k$ 's are not in general orthogonal to the  $w_k$ 's. In the new basis, the mode expansion of the photon flux of the signal field [Eq. (7.17)] becomes

$$\hat{\Phi}_S^{(+)}(\underline{x}, 0) = i \sum_k \hat{a}_k w_k(\underline{x}, 0). \quad (7.80)$$

In balanced array detection the individual detectors shown in Fig. 7.1 are replaced by arrays. The arrays are made up of individual detectors (pixels) that spatially resolve the intensity of the light that illuminates the array. Note that the area of integration in Eqs. (7.77)- (7.79) is over the entire area of the array,  $A_a$ . Thus, the difference number in Eq. (7.77) refers to the difference between the total number of photons striking array 1 and the total number striking array 2.

We now need to consider what is measured at each individual pixel of the array. The difference number  $\hat{N}_{-j}$  is obtained by subtracting the output from pixel  $j$  of array number 2 from the corresponding pixel on array 1. This difference number is found by replacing the integration over the entire array, Eq. (7.77), by integration only over the area of pixel  $j$

$$\hat{N}_{-j} = -i \int_{\text{pixel } j} d^2x v_L^*(\underline{x}, 0) \hat{\Phi}_S^{(+)}(\underline{x}, 0) + h.c. \quad (7.81)$$

Substituting Eq. (7.80) into Eq. (7.81) we obtain

$$\hat{N}_{-j} = \sum_k \hat{a}_k \hat{c}_L^\dagger \int_{\text{pixel } j} d^2x v_L^*(\underline{x}, 0) w_k(\underline{x}, 0) + h.c. \quad (7.82)$$

We now assume that the LO is in a plane-wave coherent state. The properly normalized plane-wave mode is

$$v_L(\underline{x}, 0) = (A_a)^{-1/2}. \quad (7.83)$$

As in Sec. 7.2, the fact that the LO is a large-amplitude coherent state means that we can replace the amplitude  $\hat{c}_L$  by  $|\alpha_L|e^{i\theta}$ . For this approximation to be valid  $|\alpha_L|$  must be sufficiently large that each pixel in the array is illuminated by a large-amplitude coherent state. With this assumption, Eq. (7.82) becomes

$$\hat{N}_{-j}(\theta) = \frac{|\alpha_L|}{(A_a)^{1/2}} \sum_k \int_{pixel\ j} d^2x \hat{a}_k w_k(\underline{x}, 0) e^{-i\theta} + h.c. \quad (7.84)$$

If the spatial variations of the signal field are well resolved by the array, then the relevant mode amplitudes are approximately constant over the dimensions of a pixel. We can then integrate over the pixel area and obtain

$$\hat{N}_{-j}(\theta) = \frac{|\alpha_L| A_p}{(A_a)^{1/2}} \sum_k \left( \hat{a}_k w_k(\underline{x}_j, 0) e^{-i\theta} + \hat{a}_k^\dagger w_k^*(\underline{x}_j, 0) e^{i\theta} \right), \quad (7.85)$$

where  $A_p$  is the area of an individual pixel and  $\underline{x}_j$  is the location of pixel  $j$ .

Since discrete pixels are being used, it is convenient to express the normalization condition Eq. (7.79) in terms of a discrete sum as

$$A_p \sum_j w_k^*(\underline{x}_j, 0) w_m(\underline{x}_j, 0) \cong \delta_{km}. \quad (7.86)$$

Furthermore, if at least one of the mode functions [e.g.  $w_m(\underline{x}_j, 0)$ ] is real, then taking the complex conjugate of Eq. (7.86) yields

$$A_p \sum_j w_k(\underline{x}_j, 0) w_m(\underline{x}_j, 0) \cong \delta_{km} (w_m \text{ real}). \quad (7.87)$$

Multiplying both sides of Eq. (7.85) by  $w_m(\underline{x}_j, 0)$ , summing over  $j$ , and using Eqs. (7.86) and (7.87) yields

$$\sum_j \hat{N}_{-j}(\theta) w_m(\underline{x}_j, 0) = \frac{|\alpha_L|}{(A_a)^{1/2}} \sum_k \left( \hat{a}_k e^{-i\theta} \delta_{km} + \hat{a}_k^\dagger e^{i\theta} \delta_{km} \right). \quad (7.88)$$

Performing the sum over  $k$ , and rearranging demonstrates that the quadrature amplitude for the signal in mode  $m$  is given by [149]

$$\hat{q}_{m\theta} = \frac{1}{\sqrt{2}} (\hat{a}_m e^{-i\theta} + \hat{a}_m^\dagger e^{i\theta}) = \frac{1}{|\alpha_L|} \left( \frac{A_a}{2} \right)^{1/2} \sum_j \hat{N}_{-j}(\theta) w_m(\underline{x}_j, 0). \quad (7.89)$$

The detector itself yields measurements of  $\hat{N}_{-j}(\theta)$ , while according to Eq. (7.89) the quadrature amplitude  $\hat{q}_{m\theta}$  corresponding to the measured mode  $m$  is determined after all the data has been collected by summing the measured values of  $\hat{N}_{-j}(\theta)$  with a weighting factor given by the mode function  $w_m(\underline{x}_j, 0)$ . Since an array detector is capable of making measurements

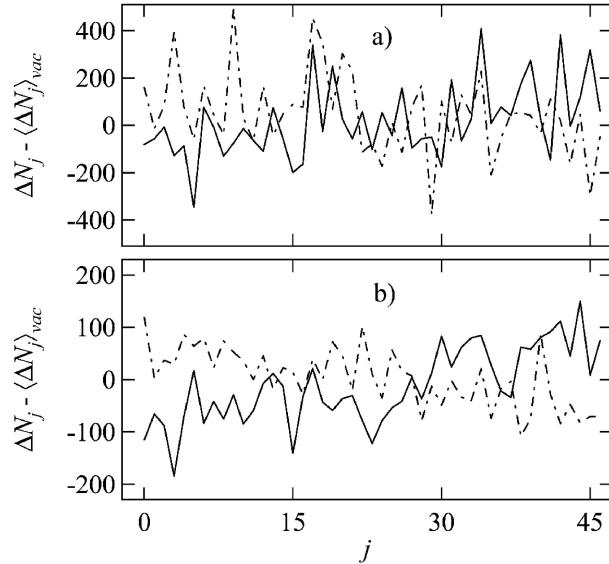
of the quadrature amplitude, these measurements can be used to reconstruct the quantum state of the signal in mode  $m$ , as described above in Sec. 7.3.

The fact that an array detector can measure the state of an optical field is not surprising. What probably is surprising is that in this detection scheme the mode functions of the measured mode  $w_m(\underline{x}_j, 0)$  and the plane-wave LO mode are not the same, but this mode-mismatch does *not* reduce the effective detection efficiency of the measurements as it would when using standard point-like detectors. Any properly normalized, real mode function  $w_m(\underline{x}_j, 0)$  can be used in Eq. (7.89), and the measured quadrature amplitude is not reduced by a factor proportional to the overlap of the signal and LO modes [as in Eq. (7.36)]. In Ref. [31] array detection was found to be over 40 times more efficient than standard detection for a particularly poorly matched set of LO and signal modes. One limitation is that the measured mode  $w_m(\underline{x}_j, 0)$  must be real, that is, must have a constant phase across its profile. This ensures that the quadrature amplitude in Eq. (7.89) is hermitian.

The mode function enters into the determination of the quadrature amplitudes during the data analysis, after all the data has been collected. Thus, by choosing different mode functions it is possible to determine the quadrature amplitudes of many different spatial modes for any given set of measurements  $\hat{N}_{-j}(\theta)$  [31]. Despite the fact that the states of many modes may be measured simultaneously, it is not possible to use this technique directly to measure the *joint* quantum state of these modes. This is because all of the modes are measured with the same rotation angle (phase shift)  $\theta$ ; to determine the joint quantum state each mode must have its own independently adjustable phase [121, 154].

In Fig. 7.17 we plot data from Ref. [31] showing the corrected difference number  $\hat{N}_{-j}(\theta) - \langle \hat{N}_{-j}(\theta) \rangle_{vac}$  as a function of pixel number observed across a one-dimensional array detector; details about subtraction of the vacuum average are deferred until Sec. 7.5.4. In Fig. 7.17 the signal field is in a weak coherent state having a mean of approximately 1 photon; Fig. 7.17(a) shows data collected on a single exposure, while Fig. 7.17(b) shows data averaged over 200 exposures. The two curves in each figure differ in that each curve corresponds to a different value of the LO phase; the phase difference between them is  $\pi$ . For this experiment the weak signal field occupied a field mode in which amplitude varied linearly with position, and the data in Fig. 7.17 confirm this. Notice that there is a  $\pi$  phase shift in the middle of the beam (the difference counts tend to be negative for half the beam, and positive for the other half). Changing the LO phase by  $\pi$  causes the slope of the curves in Fig. 7.17 to invert (positive difference counts become negative and vice versa), as expected.

Figure 7.17 is a dramatic illustration of interference at the single-photon level. While these curves contain large noise (due to the shot noise of the LO and imperfect subtraction of the vacuum difference level), they can clearly be seen to have opposite slopes. An average of one photon in the signal beam,



**Fig. 7.17.** The corrected difference number is plotted as a function of pixel number for a signal mode in a coherent state with a mean of approximately 1 photon: a) shows data for a single exposure, while b) shows an average of 200 exposures. The two curves in each figure correspond to two values of the local oscillator phase that differ by  $\pi$ . From Ref. [31]

even on single shots as shown in Fig. 7.17(a), can lead to macroscopic differences in the detected signal across *many* pixels of the array. The single signal photon acts as a “traffic cop,” determining where the millions of photons in the LO beam strike the array.

### 7.5.2 Optimization of the Measured Mode

Since it is possible to measure many different modes for a given set of data, it is natural to consider whether there is a procedure for finding the mode that optimizes the measurement of a particular quantity. Dawes et al. have shown that it is indeed possible to optimize the measurement of quantities that are quadratic in the field operators [32].

For example, consider the case of trying to find the mode that maximizes the average number of detected signal photons. In terms of the quadrature amplitudes, the average photon number of mode  $m$  is given by

$$\langle \hat{N}_m \rangle = \frac{1}{2\pi} \int_0^{2\pi} d\theta \langle \hat{q}_{m\theta}^2 \rangle - \frac{1}{2}. \quad (7.90)$$

Substituting Eq. (7.89) into Eq. (7.90) leads to

$$\langle \hat{N}_m \rangle = \frac{A_a}{2|\alpha_L|^2} \underline{\mathbf{w}}^T \cdot \underline{\underline{\mathbf{M}}} \cdot \underline{\mathbf{w}} - \frac{1}{2}, \quad (7.91)$$

where we have introduced vector notation. Here  $\underline{\underline{\mathbf{M}}}$  is the correlation matrix for the difference photocounts, averaged over the phase of the local oscillator:

$$\underline{\underline{\mathbf{M}}}_{jj'} = \frac{1}{2\pi} \int_0^{2\pi} d\theta \langle \hat{N}_{-j}(\theta) \hat{N}_{-j'}(\theta) \rangle, \quad (7.92)$$

and  $\underline{\mathbf{w}}_j = w_m(x_j, 0)$  is a vector composed from the values of the mode function taken at the pixels of the array detector.

Our goal is to find the vector  $\underline{\mathbf{w}}$  that maximizes  $\langle \hat{N}_m \rangle$ . As the second term on the right-hand side of Eq. (7.90) is constant, this task is in turn equivalent to finding the eigenvector of  $\underline{\underline{\mathbf{M}}}$  corresponding to its maximum eigenvalue. Once the optimal mode has been determined it can then be substituted for  $w_m(x_j, 0)$  in Eq. (7.89), and the quadrature amplitudes corresponding to this mode can be computed. These amplitudes can then be used to determine the quantum state of the field. This technique was experimentally demonstrated by Dawes et al. in Ref. [32].

We note that we are able to use standard numerical methods for solving symmetric eigenvalue problems because the quantity used as the optimization criterion is quadratic in the quadrature amplitude operators. Similar methods can be used to optimize other quadratic operators (e.g., it is possible to find the mode that maximizes the amount of detected squeezing.) In a general case the optimization criterion can be a highly nonlinear function, which makes the optimization problem significantly more complicated.

### 7.5.3 Joint Q-Function of Many Modes

In the above-described experiments with arrays it was possible to obtain the full quantum state of many modes simultaneously, but it was not possible to obtain the joint state (i.e., the correlations between modes were lost.) However, it is possible to use array detection to obtain information about correlations between the modes. Indeed, the experimental procedure for extracting this information is less complicated than the previously described array experiments, as the use of balanced detection is not necessary (only a single array is needed.) The price paid for joint information is that one does not measure the Wigner function or the density matrix of the field modes, but instead the joint Q-function of the modes.

The Q-function, or Husimi function, is a quantum mechanical, phase-space, quasi-probability distribution; it is positive definite and may be used to calculate quantum expectation values of antinormally ordered operators [1, 155]. It is equal to the state's Wigner function convolved with the Wigner function of the vacuum state, that is Eq. (7.35) with  $2\varepsilon^2 = 1$ . In principle

the Q-function contains all information about the quantum mechanical state of a system. However, to extract the density matrix from the Q-function it is necessary to perform a numerical deconvolution, which is impractical with real experimental data. Despite this limitation, it is possible to calculate low-order moments of anti-normally ordered operators using the Q-function (e.g., to obtain moments of photon number and a suitably defined phase.)

Measurement of the joint Q-function of many temporal modes was demonstrated in Ref. [33]. In this experiment it was possible to obtain shot-noise limited operation without balancing because the signal and LO occupied temporally distinct modes, so that classical noise on the LO could be separated out. Needed interference between the LO and signal was obtained by making measurements in the frequency domain where they overlapped, and the noise was eliminated during data processing by Fourier transforming back into the temporal domain.

Following the analysis in Refs. [33] and [150], we consider a time window consisting of  $2M+1$  temporal modes  $\hat{b}_k$ . The signal and LO fields are separated by a time delay. For the purposes of this analysis we will thus assume that the LO occupies the  $2J+1$  temporal modes near the center of our time window ( $J < M/2$ ). The signal occupies the temporal modes after the LO, and the modes before the LO are empty. In order to make this distinction between these modes more clear, we rewrite the mode operators as

$$\hat{b}_k = \begin{cases} \hat{b}_k^{(vac)} & -M \leq k < -J \\ \hat{b}_k^{(LO)} & -J \leq k \leq J \\ \hat{b}_k^{(S)} & J < k \leq M \end{cases}, \quad (7.93)$$

where the superscripts refer to vacuum, LO, or signal mode operators.

In the experiment the LO and signal pulses are measured by an array detector at the back focal plane of a grating spectrometer. The number operator for the measured spectral modes is  $\hat{N}_j = \hat{a}_j^\dagger \hat{a}_j$ , where annihilation operators  $\hat{a}_j$  can be expressed as the Fourier transform of the temporal mode operators as

$$\hat{a}_j = \frac{1}{\sqrt{(2M+1)}} \sum_k \exp[i2\pi jk/(2M+1)] \hat{b}_k. \quad (7.94)$$

The quantity of primary interest in the experiments corresponds to the Fourier transform of  $\hat{N}_j$ ,

$$\hat{K}_l \equiv \sum_j \exp[-i2\pi lj/(2M+1)] \hat{N}_j. \quad (7.95)$$

Equations (7.93)- (7.95) can be combined to express  $\hat{K}_l$  in terms of temporal mode operators  $\hat{b}_k$ . Terms of second order in operators corresponding to the weak fields  $\hat{b}_k^{(S)}$  and  $\hat{b}_k^{(vac)}$  are discarded. Furthermore, since the modes of the LO pulse are in large-amplitude coherent states, the dominant contributions are retained if we replace the LO mode operators  $\hat{b}_k^{(LO)}$  by their

corresponding coherent-state amplitudes  $\beta_k$ . The terms that contribute to the summations in the expression for  $\hat{K}_l$  depend on the value of  $l$ ; for  $l > 2J$  we find

$$\hat{K}_l = \sum_{k=-J}^J (\beta_k^* \hat{b}_{k+l}^{(S)} + \beta_k \hat{b}_{k-l}^{\dagger(vac)}). \quad (7.96)$$

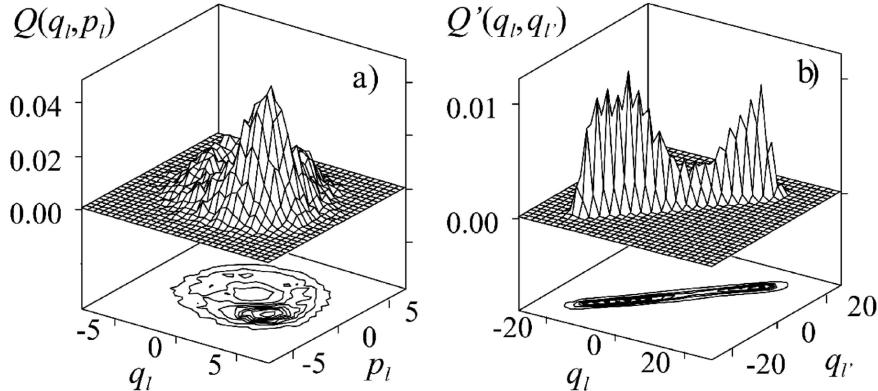
If the LO occupies only a single ( $k = 0$ ) temporal mode, then Eq. (7.96) simplifies to

$$\hat{K}_l = \beta_0^* \hat{b}_l^{(S)} + \beta_0 \hat{b}_{-l}^{\dagger(vac)}. \quad (7.97)$$

Notice that terms quadratic in the LO amplitude (i.e., the terms normally eliminated by subtraction when using balanced homodyne detection) are absent from Eqs. (7.96) and (7.97). This is because these terms are located near  $l=0$ , not near the terms of interest,  $l > 2J$ . Thus, LO noise is effectively removed, without the need to perform balanced detection.

Measurement of  $\hat{K}_l$  returns a complex number, which can be interpreted as a measurement of the signal mode plus an added vacuum noise contribution [Eq. (7.97)]. Real and imaginary parts of the measurement correspond to simultaneous measurement of the quadrature amplitudes  $q_l$  and  $p_l$ . The price paid for simultaneous measurement of noncommuting observables is the presence of the additional vacuum noise, as was first pointed out by Arthurs and Kelly [1, 156]. A similar situation arises in heterodyne detection [103].

By histogramming the measured values of  $q_l$  and  $p_l$  one creates a joint probability distribution, which in the limit of a large number of samples tends to the Q-distribution for the field quadratures  $Q(q_l, p_l)$ . Since the quadrature amplitudes for all values of  $l$  are measured simultaneously, joint Q-distributions for multiple modes can be created.



**Fig. 7.18.** The measured  $Q$ -functions of a chirped signal pulse with a random phase: a) shows the  $Q$ -function of a single mode, while b) shows correlations between 2 modes. [33]

Figure 7.18(a), taken from Ref. [33], shows the measured Q-function of a single temporal mode of an optical pulse. This is a two-dimensional histogram of measured quadrature amplitudes on 14,000 shots. The signal and LO beams came from the two arms of a Michelson interferometer; the signal arm also contained 1.5 cm of glass that added dispersion to the signal pulse; the signal pulse was thus stretched and chirped. The path-length difference between the two arms was not stabilized; this randomized the signal phase producing non-Gaussian Q-functions. The Q-function of Fig. 7.18(a) is largely circular, however there is still a peak in the distribution indicating that the phase was not completely randomized.

Two-mode distributions have the form  $Q(q_l, p_l, q_{l'}, p_{l'})$ ; these four-dimensional distributions are difficult to display graphically, so correlations between modes are usually displayed in terms of the joint distribution of the  $q$  quadratures

$$Q'(q_l, q_{l'}) = \iint Q(q_l, p_l, q_{l'}, p_{l'}) dp_l dp_{l'}. \quad (7.98)$$

Figure 7.18(b) shows correlations between the  $q$ -quadratures of two modes in terms of joint distributions  $Q'(q_l, q_{l'})$ . This figure shows correlations between two temporal modes that were both near the peak of the pulse, and consequently had nearly the same amplitude and phase. This joint distribution lies mainly along a line whose slope is 1, indicating strong positive correlations between the  $q$ -quadratures. This is what one would expect for two modes whose relative phases are nearly the same, but whose absolute phases are random.

In the experiment of Ref. [33], the exposure time of the array was 300 ms, so each “shot” was actually composed of millions of pulses, since the laser had a repetition rate of 82 MHz. Any noise at frequencies of  $1/(300 \text{ ms})$  and higher was then integrated and averaged over. The CCD array used had a very long read time, so the repetition rate of the measurements was approximately  $1/2 \text{ Hz}$ , yielding an experimental run of nearly eight hours. Noise due to slow drift of the interferometer phase over this time was the dominant contribution to the shapes of the measured distributions in Fig. 7.18. Furthermore, since many pulses were integrated on each shot, the experiment did not measure the statistics of temporal slices of an ensemble of single pulses. The measured temporal mode was a “super”-mode, representing the corresponding time slices of millions of pulses. In principle it is not necessary to average over millions of pulses; a laser pulse repetition rate slower than the inverse CCD exposure time would lead to measurements of an ensemble of single pulses.

#### 7.5.4 Technical Considerations for Array Detection

As with ordinary single detectors, one wants array detectors to have high quantum efficiency and low electronic noise. Since arrays are made of many adjacent pixels, one factor that contributes to the efficiency is the “fill factor,”

which is a measure of the fraction of the detector area that is sensitive to light. Fill factors of less than 100% can be due to gaps between individual pixels. For example, interline-transfer charge-coupled device (CCD) detectors have alternating rows of photosensitive pixels and rows of nonphotosensitive areas that are used to read the charge out of the array. Such CCD's have fill factors of only 25-75%, and hence are not suitable for quantum optics applications. Full-frame CCD's, on the other hand, have fill factors of 100%. The trade off here is that since charge is transferred from one pixel to the next in order to be read out, charge smearing occurs if the array is exposed to light during readout. This means that an external shutter is necessary to block the light during readout.

CCD's come in two basic types: front-illuminated and back-illuminated. The polysilicon gate structure used to read the charge out of the array is placed on the front surface of the array. In front-illuminated devices light is incident through this gate structure. Light is absorbed and reflected by the polysilicon, reducing the quantum efficiency. In back-illuminated CCD's the backside of the silicon wafer is thinned to a thickness of about  $20\ \mu\text{m}$ , and light is incident through the back. There is no gate structure in the way of the light, so the quantum efficiency of back-illuminated devices can exceed 90%. However, the  $20\ \mu\text{m}$  layer of silicon can act as an etalon, producing unwanted fringes in the acquired image; this effect is especially pronounced in the near IR. The latest generation of back-illuminated CCD's has been engineered to greatly reduce this etaloning effect.

Electronic noise in CCD's comes in the form of dark noise and readout noise. Scientific grade CCD cameras can be cooled to temperatures of  $-100^\circ\text{C}$  using either thermoelectric or liquid nitrogen cooling, and at this temperature dark noise is essentially non-existent (on the order of  $1\ \text{e}^-/\text{pixel/hr}$ ). Readout noise can be less than  $10\ \text{e}^-/\text{pixel rms}$ , depending on the readout rate—higher readout rates have larger noise.

Most scientific grade CCD's have 16-bit resolution ADC's. These CCD's are frequently used for Raman or fluorescence spectroscopy, and can be purchased from most vendors that sell spectrometers. Some vendors that we are aware of are:

- Roper Scientific Inc., Trenton, NJ; [www.roperscientific.com](http://www.roperscientific.com)
- Andor Technology, South Windsor, CT; [www.andor-tech.com](http://www.andor-tech.com)

As with single detectors, it is necessary to ensure that the detection is shot-noise limited. This is done in the same manner as described in Sec. 7.4.1. The detector is illuminated by an LO beam of varying intensity and the variance of the measured difference number between pairs of pixels is plotted versus the mean. Detection at the SNL yields a linear plot. The slope and intercept determine the gain and electronic noise, and these should agree with the specifications of the manufacturer. This calibration can be done either pixel-by-pixel, or for the sum of the outputs from a large number of pixels.

In balanced array detection it is extremely important to register properly the individual pixels detecting the two beams illuminating the array to obtain high-efficiency subtraction and hence good classical noise reduction. This is one of the more difficult parts of array experiments. If the outputs are not properly registered, operation at the SNL is not obtained. A detailed description of one convenient registration procedure is given in Ref. [31].

Once the pixels have been registered the detector needs to be balanced as well as possible. If the signal field is blocked (i.e., the signal mode entering the detector is in the vacuum state) then the average difference number for each pixel should be zero:  $\langle N_{-j}(\theta) \rangle_{vac} = 0$ , where the subscript indicates that the signal is in the vacuum state. This is extremely difficult to achieve for every pixel simultaneously. To eliminate the effects of offsets in the measured difference number for individual pixels, it is necessary in practice to subtract these offsets. Thus, in Eq. (7.89) one uses the corrected difference number  $N_{-j}(\theta) - \langle N_{-j}(\theta) \rangle_{vac}$  in place of  $N_{-j}(\theta)$  when calculating quadrature amplitudes. The measured background and signal levels are obtained in the experiment by alternately blocking and unblocking the signal with a shutter.

In unbalanced array detection there is no need to register or balance pixels, and operation at the SNL is easily obtained, even with LO fluctuations of 15% peak-to-peak [33].

## 7.6 Conclusions

Quantum state tomography of optical fields has come of age. Many theoretical algorithms exist for converting measured quadrature amplitudes into information about the quantum state. Numerous experiments have been performed which demonstrate the utility of these algorithms (see Table 1.) These experiments have used several different detection technologies (DC, RF, array), and have measured quantities ranging from Wigner functions to photon number and phase distributions. Techniques have been developed to measure two or more mode systems, allowing for the measurement of temporal or polarization correlations between modes.

Ultrafast linear optical sampling has been a spin off of state measurement technology. The ability of a balanced homodyne detector to perform time-resolved measurements of weak fields is very important from a practical perspective. Work in this area is really just beginning, and we expect it to have a bright future.

## Acknowledgements

We wish to thank the many collaborators who contributed so much to our work in the area of QST: Matt Anderson, Ethan Blangett, Howard

Carmichael, Jinx Cooper, Adel Faradani, Andy Funk, Tamas Kiss, Ulf Leonhardt, Dan McAlister, Mike Munroe, Thomas Richter, Dan Smithey, Ian Walmsley, Andrew Dawes, Konrad Banaszek, and Christophe Dorrer. We thank A. Funk for writing the latter half of Sec. 7.4.3. The work in our groups has been supported by the National Science Foundation and by the Army Research Office.

## A Spatial-Temporal Orthogonality

Define

$$\begin{aligned} S_{ml} &= c \int_0^T dt \int_{-\infty}^{\infty} d^2x \underline{v}_m^*(\underline{x}, 0, t) \cdot \underline{v}_l(\underline{x}, 0, t) \\ &= c \int_0^T dt \int_{-\infty}^{\infty} d^2x \sum_j C_{mj}^* \underline{u}_j^*(\underline{x}, 0) e^{i\omega_j t} \cdot \sum_i C_{li} \underline{u}_i(\underline{x}, 0) e^{-i\omega_i t}. \end{aligned} \quad (7.99)$$

Using plane-wave modes,  $\underline{u}_j(\underline{r}) = V^{-1/2} \underline{\varepsilon}^{(j)} \exp(i\underline{k}_j \cdot \underline{r})$ , with  $V = L_z L_y L_x$ , gives

$$S_{ml} = \sum_{j,i} C_{mj}^* C_{li} \delta_K(\underline{k}_{Tj} - \underline{k}_{Ti}) \frac{c}{L_z} \int_0^T dt \exp[i(\omega_j - \omega_i)t], \quad (7.100)$$

where  $\delta_K(\underline{k}_{Tj} - \underline{k}_{Ti})$  is a Kronecker delta indicating that the transverse components of  $\underline{k}_j$  and  $\underline{k}_i$  must be equal. In the paraxial approximation ( $k_x, k_y \ll k_z$ ) the frequencies are  $\omega_j/c \cong k_{zj} + (k_{Tj}^2)/(2k_{zj})$  and  $\omega_i/c \cong k_{zi} + (k_{Ti}^2)/(2k_{zi})$ , where  $k_{zj} = j2\pi/L_z$  ( $j = 1, 2, 3 \dots$ ) and  $cT = L_z$ . This gives (with the constraint  $\underline{k}_{Tj} = \underline{k}_{Ti}$ ),

$$\frac{c}{L_z} \int_0^T dt \exp[i(\omega_j - \omega_i)t] \cong \begin{cases} 1 & (j = i) \\ (\theta_0/2)^2 & (j \neq i) \end{cases}, \quad (7.101)$$

where  $\theta_0 = k_{Tj}/k_z \ll 1$  is the angle of the modes' propagation vector from the normal to the  $z = 0$  plane. For typical laser beam divergences of  $\theta_0 \approx 10^{-3}$  this deviation from zero is small, making the time integral (combined with the transverse integral) behave like a Kronecker delta  $\delta_{ji}$ . This then leads to

$$S_{ml} = \sum_j C_{mj}^* C_{lj} = \delta_{ml}. \quad (7.102)$$

## References

1. U. Leonhardt: *Measuring the Quantum State of Light* (Cambridge University Press, Cambridge, UK, 1997).

2. J. Polchinski: Phys. Rev. Lett. **66**, 397 (1991).
3. A. Royer, Found. Phys. **19**, 3 (1989).
4. L. E. Ballentine: *Quantum Mechanics: a Modern Development* (World Scientific, Singapore, 1998).
5. O. Alter and Y. Yamamoto: Phys. Rev. Lett. **74**, 4106 (1995).
6. O. Alter and Y. Yamamoto: *Quantum Measurement of a Single System* (Wiley, New York, 2001).
7. G. M. D'Ariano and H. P. Yuen: Phys. Rev. Lett. **76**, 2832 (1996).
8. G. M. D'Ariano: In *Quantum Optics and the Spectroscopy of Solids*, ed by T. Hakioglu and A. S. Shumovsky (Kluwer, Dordrecht, 1997), p 139.
9. D.-G. Welsch, W. Vogel, and T. Opatrny: In *Progress in Optics*, vol XXXIX, ed by E. Wolf (North Holland, Amsterdam, 1999), p. 63.
10. K. Vogel and H. Risken: Phys. Rev. A **40**, R2847 (1989).
11. W. Pauli: *General Principles of Quantum Mechanics* (Springer-Verlag, Berlin, 1980).
12. W. Gale, E. Guth, and G. T. Trammel: Phys. Rev. **165**, 1434 (1968).
13. R. W. Gerchberg and W. O. Saxon: Optik **35**, 237 (1972).
14. A. M. J. Huiser, A. J. J. Drenth, and H. A. Ferwerda: Optik **45**, 303 (1976).
15. A. Orlowski and H. Paul: Phys. Rev. A **50**, R921 (1994).
16. D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani: Phys. Rev. Lett. **70**, 1244 (1993).
17. J. R. Ashburn, R. A. Cline, P. J. M. van der Burgt, W. B. Westerveld, and J. S. Risley: Phys. Rev. A **41**, 2407 (1990).
18. G. Breitenbach, S. Schiller, and J. Mlynek: Nature **387**, 471 (1997).
19. K. Banaszek, C. Radzewicz, K. Wodkiewicz, and J. S. Krasinski: Phys. Rev. A **60**, 674 (1999).
20. A. I. Lvovsky, H. Hansen, T. Aichele, O. Benson, J. Mlynek, and S. Schiller: Phys. Rev. Lett. **87**, 050402 (2001).
21. A. G. White, D. F. V. James, P. H. Eberhard, and P. G. Kwiat: Phys. Rev. Lett. **83**, 3103 (1999).
22. T. J. Dunn, I. A. Walmsley, and S. Mukamel: Phys. Rev. Lett. **74**, 884 (1995).
23. D. Leibfried, D. M. Meekhof, B. E. King, C. Monroe, W. M. Itano, and D. J. Wineland: Phys. Rev. Lett. **77**, 4281 (1996).
24. C. Kurtsiefer, T. Pfau, and J. Mlynek: Nature **386**, 150 (1997).
25. I. L. Chuang, N. Gershenfeld, M. G. Kubinec, and D. W. Leung: Proc. R. Soc. London A **454**, 447 (1998).
26. M. G. Raymer: Contemp. Phys. **38**, 343 (1997).
27. M. G. Raymer, J. Cooper, H. J. Carmichael, M. Beck, and D. T. Smithey: J. Opt. Soc. Am. B **12**, 1801 (1995).
28. R. E. Slusher, L. W. Hollberg, B. Yurke, J. C. Mertz, and J. F. Valley: Phys. Rev. Lett. **55**, 2409 (1985).
29. L.-A. Wu, H. J. Kimble, J. L. Hall, and H. Wu: Phys. Rev. Lett. **57**, 2520 (1986).
30. S. Machida and Y. Yamamoto: IEEE J. Quantum Elect. **QE-22**, 617 (1986).
31. A. M. Dawes and M. Beck: Phys. Rev. A **63**, 040101 (2001).
32. A. M. Dawes, M. Beck, and K. Banaszek: Phys. Rev. A **67**, 032102 (2003).
33. M. Beck, C. Dorner, and I. A. Walmsley: Phys. Rev. Lett. **87**, 253601 (2001).
34. M. Beck, D. T. Smithey, and M. G. Raymer: Phys. Rev. A **48**, R890 (1993).
35. D. T. Smithey, M. Beck, J. Cooper, M. G. Raymer, and A. Faridani: Physica Scripta **T48**, 35 (1993).

36. M. Beck, D. T. Smithey, J. Cooper, and M. G. Raymer: Opt. Lett. **18**, 12 (1993).
37. D. T. Smithey, M. Beck, J. Cooper, and M. G. Raymer: Phys. Rev. A **48**, 3159 (1993).
38. M. Beck, M. E. Anderson, and M. G. Raymer: In *OSA Proceedings on Advances in Optical Imaging and Photon Migration*, vol 21, ed by R. R. Alfano (Optical Society of America, Washington, DC, 1994), p 257.
39. M. Munroe, D. Boggavarapu, M. E. Anderson, and M. G. Raymer: Phys. Rev. A **52**, R924 (1995).
40. G. Breitenbach, T. Muller, S. F. Pereira, J.-P. Poizat, S. Schiller, and J. Mlynek: J. Opt. Soc. Am. B **12**, 2304 (1995).
41. M. E. Anderson, M. Munroe, U. Leonhardt, D. Boggavarapu, D. F. McAlister, and M. G. Raymer: In *Generation, Amplification, and Measurement of Ultrashort Laser Pulses III*, vol 2701, ed by W. E. White and D. H. Reitze (SPIE, Bellingham, WA, 1996), p 142.
42. S. Schiller, G. Breitenbach, S. F. Pereira, T. Muller, and J. Mlynek: Phys. Rev. Lett. **77**, 2933 (1996).
43. D. F. McAlister and M. G. Raymer: Phys. Rev. A **55**, R1609 (1997).
44. J. W. Wu, P. K. Lam, M. B. Gray, and H.-A. Bachor: Opt. Express **3**, 154 (1998).
45. M. Vasilyev, S. K. Choi, P. Kumar, and G. M. D'Ariano: Opt. Lett. **23**, 1393 (1998).
46. T. Coudreau, L. Vernac, A. Z. Khoury, G. Breitenbach, and E. Giacobino: Europhys. Lett. **46**, 722 (1999).
47. M. Vasilyev, S. K. Choi, P. Kumar, and G. M. D'Ariano: Phys. Rev. Lett. **84**, 2354 (2000).
48. P. Voss, M. Vasilyev, D. Levandovsky, T. G. Noh, and P. Kumar: IEEE Photon. Tech. Lett. **12**, 1340 (2000).
49. M. Fiorentino, A. Conti, A. Zavatta, G. Giacomelli, and F. Marin: J. Opt. B: Quantum Semiclassical Opt. **2**, 184 (2000).
50. E. L. Blangett, M. G. Raymer, G. Khitrova, H. M. Gibbs, D. K. Serkland, A. A. Allerman, and K. M. Geib: Opt. Express **9**, 312 (2001).
51. H. Hansen, T. Aichele, C. Hettich, P. Lodahl, A. I. Lvovsky, J. Mlynek, and S. Schiller: Opt. Lett. **26**, 1714 (2001).
52. A. Zavatta, M. Bellini, P. L. Ramazza, F. Marin, and F. T. Arecchi: J. Opt. Soc. Am. B **19**, 1189 (2002).
53. A. I. Lvovsky and J. Mlynek: Phys. Rev. Lett. **88**, 250401 (2002).
54. A. I. Lvovsky and S. A. Babichev: Phys. Rev. A **66**, 011801 (2002).
55. A. Zavatta, R. Marin, and G. Giacomelli: Phys. Rev. A **66**, 043805 (2002).
56. P. Bertet, A. Auffeves, P. Maioli, S. Osnaghi, T. Meunier, M. Brune, J. M. Raimond, and S. Haroche: Phys. Rev. Lett. **89**, 200402 (2002).
57. P. Voss, T. G. Noh, S. Dugan, M. Vasilyev, P. Kumar, and G. M. D'Ariano: J. Mod. Opt. **49**, 2289 (2002).
58. H. P. Yuen and V. W. S. Chan: Opt. Lett. **8**, 177 (1983).
59. G. L. Abbas, V. W. S. Chan, and T. K. Yee: Opt. Lett. **8**, 419 (1983).
60. R. E. Slusher, P. Grangier, A. La Porta, B. Yurke, and M. J. Potasek: Phys. Rev. Lett. **59**, 2566 (1987).
61. G. M. D'Ariano, C. Macchiavello, and M. G. A. Paris: Phys. Rev. A **50**, 4298 (1994).

62. M. Munroe: Ultrafast Photon Statistics of Cavityless Laser Light, Ph.D. Dissertation, University of Oregon, Eugene, OR, (1996).
63. H. P. Yuen and J. H. Shapiro: IEEE Trans. Inform. Theory **IT-24**, 657 (1978).
64. J. H. Shapiro, H. P. Yuen, and J. A. Machado Mata: IEEE Trans. Inform. Theory **IT-25**, 179 (1979).
65. H. P. Yuen and J. H. Shapiro: IEEE Trans. Inform. Theory **IT-26**, 78 (1980).
66. B. Yurke: Phys. Rev. A **32**, 300 (1985).
67. B. Yurke: Phys. Rev. A **32**, 311 (1985).
68. B. Yurke, P. Grangier, R. E. Slusher, and M. J. Potasek: Phys. Rev. A **35**, 3586 (1987).
69. P. D. Drummond: In *Quantum Optics V* ed. by J. D. Harvey and D. F. Walls (Springer-Verlag, Berlin, 1989).
70. K. J. Blow, R. Loudon, S. J. D. Phoenix, and T. J. Shepherd: Phys. Rev. A **42**, 4102 (1990).
71. B. Huttner, J. J. Baumberg, J. F. Ryan, and S. M. Barnett: Opt. Commun. **90**, 128 (1992).
72. R. Loudon: In *Coherence, Cooperation and Fluctuations*, ed by F. Haake, L. M. Narducci and D. F. Walls (Cambridge Univ. Press, Cambridge, 1986), p 240.
73. B. Yurke and D. Stoler: Phys. Rev. A **36**, 1955 (1987).
74. M. J. Collett, R. Loudon, and C. W. Gardiner: J. Mod. Opt. **34**, 881 (1987).
75. S. L. Braunstein: Phys. Rev. A **42**, 474 (1991).
76. W. Vogel and J. Grabow: Phys. Rev. A **47**, 4227 (1993).
77. W. Vogel and W. Schleich: Phys. Rev. A **44**, 7642 (1991).
78. H. Kuhn, D.-G. Welsch, and W. Vogel: J. Mod. Opt. **41**, 1607 (1994).
79. M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner: Phys. Rep. **106**, 121 (1984).
80. J. Bertrand and P. Bertrand: Found. Phys. **17**, 397 (1987).
81. U. Leonhardt and M. G. Raymer: Phys. Rev. Lett. **76**, 1985 (1996).
82. G. Drobny and V. Bužek: Phys. Rev. A **65**, 053410 (2002).
83. C. M. Caves, C. A. Fuchs, and R. Schack: Phys. Rev. A **65**, 022305 (2002).
84. R. Schack, T. A. Brun, and C. M. Caves: Phys. Rev. A **6401**, 014305 (2001).
85. K. R. W. Jones: Ann. Phys. **207**, 140 (1991).
86. K. R. W. Jones: Phys. Rev. A **50**, 3682 (1994).
87. E. T. Jaynes: In *Maximum Entropy and Bayesian Methods*, ed by J. Skilling (Kluwer Academic Publishers, Dordrecht, 1989), p 1.
88. V. Bužek, R. Derka, G. Adam, and P. L. Knight: Ann. Phys. **266**, 454 (1998).
89. U. Leonhardt, M. Munroe, T. Kiss, T. Richter, and M. G. Raymer: Opt. Commun. **127**, 144 (1996).
90. S. M. Tan: J. Mod. Opt. **44**, 2233 (1997).
91. V. Bužek, G. Adam, and G. Drobny: Ann. Phys. **245**, 37 (1996).
92. K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi: Phys. Rev. A **6001**, 010304 (2000).
93. G. Breitenbach: Quantum state reconstruction of classical and nonclassical light and a cryogenic opto-mechanical sensor for high-precision interferometry, UFO Dissertation, Universität Konstanz, Konstanz, Germany, (1998).
94. C. Dorrrer, D. C. Kilper, H. R. Stuart, G. Raybon, and M. G. Raymer: Photon. Tech. Lett. (to be published, 2003).
95. U. Fano: Rev. Mod. Phys. **29**, 74 (1957).

96. For a tutorial derivation see Z. Y. Ou and L. Mandel: Am. J. Phys. **57**, 66 (1989).
97. For review and references see U. Leonhardt: Phys. Rev. A **48**, 3265 (1993).
98. These relations are most appropriate in the context of pulsed fields with a local description, as considered here, rather than for monochromatic plane waves which strictly speaking extend throughout space. See W. Vogel and D.-G. Welsch: Lectures on Quantum Optics (Akademie Verlag, Berlin, 1994), Sec. 6.3.1; L. Knoll, W. Vogel, and D.-G. Welsch: Phys. Rev. A **36**, 3803 (1987).
99. D. T. Smithey, M. Beck, M. Belsley, and M. G. Raymer: Phys. Rev. Lett. **69**, 2650 (1992).
100. J. Guena, P. Jacquier, M. Lintz, L. Pottier, M. A. Bouchiat, and A. Hrisohou: Opt. Commun. **71**, 6 (1989).
101. R. J. Glauber: In *Quantum Optics and Electronics*, ed by C. DeWitt-Morette, A. Blandin and C. Cohen-Tannoudji (Gordon and Breach, New York, 1965), p 63.
102. P. L. Kelly and W. H. Kleiner: Phys. Rev. A **136**, 316 (1964).
103. J. H. Shapiro and S. S. Wagner: IEEE J. Quantum Elect. **QE-20**, 803 (1984).
104. P. D. Drummond: Phys. Rev. A **35**, 4253 (1987).
105. U. M. Titulaer and R. J. Glauber: Phys. Rev. **145**, 1041 (1966).
106. M. G. Raymer and I. A. Walmsley: In *Progress in Optics*, vol XXVIII, ed by E. Wolf (North Holland, Amsterdam, 1990), p 181.
107. This replacement can be shown to be rigorous if the LO is in a coherent state; M. G. Raymer, J. Cooper, H. J. Carmichael, M. Beck, and D. T. Smithey: J. Opt. Soc. Am. B **12**, 1801 (1995).
108. G. T. Herman: *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography* (Academic Press, New York, 1980).
109. D. F. McAlister: Measuring the Classical and Quantum States and Ultrafast Correlations of Optical Fields, Ph.D. Dissertation, University of Oregon, Eugene, OR, (1999).
110. U. Leonhardt and H. Paul: Phys. Rev. A **48**, 4598 (1993).
111. R. Lynch: Phys. Rep. **256**, 367 (1995).
112. D. T. Pegg and S. M. Barnett: Phys. Rev. A **39**, 1665 (1989).
113. T. Richter: Phys. Rev. A **61**, 063819 (2000).
114. U. Leonhardt and M. Munroe: Phys. Rev. A **54**, 3682 (1996).
115. R. J. Marks: *Introduction to Shannon Sampling and Interpolation Theory* (Springer-Verlag, New York, 1991).
116. H. Paul, U. Leonhardt, and G. M. D'Ariano: Acta Phys. Slovaca **45**, 261 (1995).
117. G. M. D'Ariano, C. Macchiavello, and M. G. A. Paris: Phys. Lett. A **195**, 31 (1994).
118. G. M. D'Ariano and M. G. A. Paris: Phys. Lett. A **233**, 49 (1997).
119. D. V. O'Connor and D. Phillips: *Time-Correlated Single Photon Counting* (Academic Press, London, 1984).
120. T. Richter: Phys. Rev. A **53**, 1197 (1996).
121. M. G. Raymer, D. F. McAlister, and U. Leonhardt: Phys. Rev. A **54**, 2397 (1996).
122. T. Opatrny, D.-G. Welsch, and W. Vogel: Acta Phys. Slovaca **46**, 469 (1996).
123. The article in Progress in Optics by Welsch corrects a sign error for  $\xi$  in M. G. Raymer, D. F. McAlister, and U. Leonhardt: Phys. Rev. A **54**, 2397 (1996).

124. D. F. McAlister and M. G. Raymer: *J. Mod. Opt.* **44**, 2359 (1997).
125. G. M. D'Ariano, M. F. Sacchi, and P. Kumar: *Phys. Rev. A* **60**, 013806 (2000).
126. Although the full joint Q-function, which contains much of the information about the state, of a multiple-mode beam has been reconstructed using array detection; see Sec. 7.5.3 and M. Beck, C. Dorner, and I. A. Walmsley: *Phys. Rev. Lett.* **87**, 253601 (2001).
127. M. G. Raymer, D. F. McAlister, and A. C. Funk: In *Quantum Communication, Computing and Measurement 2*, ed by P. Kumar, G. M. D'Ariano and O. Hirota (Kluwer Academic/Plenum Publishers, New York, 2000), p 147.
128. P. A. Bushev, V. P. Karassiov, A. V. Masalov, and A. A. Putilin: In *Quantum and Atomic Optics, High-Precision Measurements in Optics, and Optical Information Processing, Transmission, and Storage*, vol 4750, ed by S. N. Bagayev, S. S. Chesnokov, A. S. Chirkin and V. N. Zadkov (SPIE, Bellingham, WA, 2002), p 36.
129. V. P. Karassiov and A. V. Masalov: *Laser Phys.* **12**, 948 (2002).
130. V. P. Karassiov and A. V. Masalov: *J. Opt. B: Quantum Semiclassical Opt.* **4**, S366 (2002).
131. V. P. Karasev and A. V. Masalov: *Optics and Spectroscopy* **74**, 551 (1993).
132. M. G. Raymer and A. C. Funk: *Phys. Rev. A* **60**, 015801 (2000).
133. T. Richter: *J. Mod. Opt.* **44**, 2385 (1997).
134. E. L. Blangett: Picosecond Polarization Dynamics and Noise of Vertical-Cavity Surface-Emitting Lasers, Ph.D. Dissertation, University of Oregon, Eugene, (2002).
135. K.-H. Brenner and K. Wodkiewicz: *Opt. Commun.* **43**, 103 (1982).
136. J. H. Eberly and K. Wodkiewicz: *J. Opt. Soc. Am.* **67**, 1252 (1977).
137. G. Nienhuis: *Physica* **96C**, 391 (1979).
138. G. Nienhuis: *J. Phys. B* **16**, 2677 (1983).
139. P. W. Nicholson: *Nuclear Electronics* (Wiley, New York, 1974).
140. V. Radeka: *Ann. Rev. Nucl. Part. Sci.* **38**, 217 (1988).
141. L.-A. Wu, M. Xiao, and H. J. Kimble: *J. Opt. Soc. Am. B* **4**, 1465 (1987).
142. P. Grangier, R. E. Slusher, B. Yurke, and A. La Porta: *Phys. Rev. Lett.* **1987**, 2153–2156 (1987).
143. A. La Porta and R. E. Slusher: *Phys. Rev. A* **44**, 2013 (1991).
144. S. K. Choi, R. D. Li, C. Kim, and P. Kumar: *J. Opt. Soc. Am. B* **14**, 1564 (1997).
145. J. H. Shapiro and A. Shakeel: *J. Opt. Soc. Am. B* **14**, 232 (1997).
146. C. Kim and P. Kumar: *Phys. Rev. Lett.* **73**, 1605 (1994).
147. A. C. Funk: Univ. of Oregon (Personal Communication, 2003).
148. M. G. Raymer, J. Cooper, and M. Beck: *Phys. Rev. A* **48**, 4617 (1993).
149. M. Beck: *Phys. Rev. Lett.* **84**, 5748 (2000).
150. C. Iaconis, E. Mukamel, and I. A. Walmsley: *J. Opt. B: Quantum Semiclassical Opt.* **2**, 510 (2000).
151. T. Opatrny, D.-G. Welsch, S. Wallentowitz, and W. Vogel: *J. Mod. Opt.* **44**, 2405 (1997).
152. S. Wallentowitz and W. Vogel: *Phys. Rev. A* **53**, 4528–4533 (1996).
153. K. Banaszek and K. Wodkiewicz: *Phys. Rev. Lett.* **76**, 4344 (1996).
154. H. Kuhn, D.-G. Welsch, and W. Vogel: *Phys. Rev. A* **51**, 4240 (1995).
155. Y. Lai and H. A. Haus: *Quantum Opt.* **1**, 99 (1989).
156. E. Arthurs and J. L. Kelly, Jr.: *Bell Syst. Tech. J.* **44**, 725 (1965).

## 8 Characterization of Quantum Devices

Giacomo Mauro D'Ariano and Paoloplacido Lo Presti

QUIT Group of INFM, Dipartimento di Fisica "A. Volta", via Bassi 6, Pavia, Italy  
[dariano@unipv.it](mailto:dariano@unipv.it), [lopresti@unipv.it](mailto:lopresti@unipv.it), [www.qubit.it](http://www.qubit.it)

### Introduction

The new field of quantum information has opened the way to a new kind of astonishingly efficient information processing achieved by physical transformations. This new kind of processing will be performed by a radically new generation of quantum devices, and this will make the design of characterization tools for such devices of paramount importance, besides being already of foundational interest by themselves, for the obvious possibility of experimental determination of the dynamics of a quantum system.

Quantum devices can perform either deterministic or probabilistic transformations of a quantum state. The transformations of the deterministic class are generally referred to as "processes" or "channels", and describe the evolution of closed systems or of open systems undergoing an irreversible dynamics, such as due to an interaction with a bath. The class of probabilistic transformations, on the other hand, typically describe the so called "state reduction" occurring in a quantum measurement. Both types of transformations can be described in the language of quantum operations (QO) [1,2], and, within this common mathematical structure, both deterministic and non-deterministic transformations can be characterized by the same means.

At the root of the characterization problem, there is the need of finding a way to imprint the description of the QO of the device on a suitable input state that is processed by the device, and is then characterized at its output by some quantum tomographic means. Linearity of QO's is the first key ingredient for solving this "quantum black box" problem. In Refs. [3,4] it was shown that for a "complete" set of input states, i.e. for a set of generators of the space of states, the "transfer matrix" of the device remains encoded in the input/output correlations in the same way as for any classical linear system, the only difference being that in the quantum case one needs many copies of the outputs to perform their quantum tomography. Quantum process tomography was achieved by this method in liquid nuclear magnetic resonance systems [5–7], and for processes on qubits encoded in the polarization of a radiation mode [8,9]. Unfortunately, this method needs the preparation of an orthogonal set of input states along with some relative superpositions, and such sets of states are very seldom available in the lab: for example, they are not achievable in quantum optics.

Quantum mechanics, however, offers a unique opportunity to achieve our goal by using a composite system. In fact, in Refs. [10] and [11] it was shown that the action of a quantum process on one system of an entangled pair produces a joint output state containing a complete description of the process itself, a result also known as the Jamiolkowsky isomorphism [12,13]. In simple words, a fixed maximally entangled input state supports the imprinting of any QO, as if it was effectively running all possible input states in parallel, and in this way the determination of the process is achieved by simply performing the tomography of the joint state at the output, with the device acting on one of the two entangled systems only. Experiments of process tomography using entangled input probes have been recently implemented [14–16] for optical qubits, and proposed for optical “continuous variables” systems using homodyne tomography [10].

In Ref. [17], the two methods—“many inputs” versus “single entangled input”—have been bridged together in a complete classification of all states (and/or all ensembles of states) that support a complete imprinting of a generic QO, thereafter named “faithful states”. There, the existence of separable faithful states has been established, thus clarifying that for the “quantum black box” problem the only thing that matters is the use of composite systems (i. e. with the tensor product rule), more than entanglement itself. Among such faithful separable states there are also the Werner states used in the process tomography experiment of Ref. [15]. In Ref. [17], a measure of the “faithfulness” of the state has also been given, which measure in some way the precision of the tomographic characterization, showing that maximally entangled states offer the best performance.

Once the information on the process is encoded on the quantum state, all known techniques of state-tomography and state-discrimination [18] can be applied. Such techniques will allow, in the future, a precise characterization of any kind of quantum device, from an optical fiber for a quantum communication channel, to an NMR qubit gate, from a parametric amplifier to a photon-counting detector.

The present chapter is aimed at a complete and self-contained presentation of the theoretical basis of the methods for imprinting quantum operations on quantum states, also providing concrete examples of experimental setups based on homodyne tomography, to be used for tomography of either quantum processes or detectors. In Sec. 8.1 we introduce the formalism of quantum operations (QO) and positive operator-valued measures (POVM) for describing the state transformations operated by a quantum device and the statistics of the outcomes of a quantum measurement, respectively. The properties of these two mathematical objects are derived as necessary consequences of the definition and interpretation of quantum state, and the definition of the states of composite systems. These properties are the starting point for constructing powerful representations of QO's that will be used first for illustrating the relation between QO's and POVM with customary unitary evolutions and

projective measurements, and then to analyze the problem of the characterization of a device. This mathematical framework is then employed in Sec. 8.2 for giving the complete classification of the *faithful states*— i.e. the input states that can be used for the characterization of a quantum device—and to address the problem of quantifying their degree of faithfulness. Finally, Sec. 8.3 is devoted to the exposition of a quantum optical setup for performing a device characterization by homodyne tomography using an entangled input state from parametric down-conversion of vacuum. Finally, we report numerical simulations of experimental results that can be obtained with the current technology for the homodyne tomography of an amplitude displacing device and of an On/Off photo-detector, using either the averaging or the maximum-likelihood strategies.

## 8.1 Quantum operations and quantum measurements

Quantum operations (QO), introduced for the first time in Refs. [1, 2], describe all possible transformations—either deterministic or probabilistic—of the state  $\rho$  of a quantum system. Mathematically, QO’s are completely positive (CP) linear maps from the set of (trace-class) operators on  $H$  to itself, and are trace preserving when deterministic and trace-decreasing when probabilistic, with the probability of occurrence given by the output trace.

In this section we’ll show how these properties for QO’s can be traced back to the indistinguishability of different preparations of the same ensemble of systems and to the tensor product structure of composite systems. After these observations, we will introduce a one-to-one correspondence between CP maps and positive operators on  $H^{\otimes 2}$ , which provides the easiest framework to proof most relevant results concerning quantum operations.

At the end, we will review the concept of POVM and its connection with quantum operations, for representing the probability distribution of the outcomes of a quantum measurement.

### 8.1.1 Properties of quantum operations

Let’s consider a system in the state  $\rho$ , and suppose it enters a device in which a physical transformation described by the map  $\mathcal{T}$

$$\rho \mapsto \mathcal{T}(\rho) \tag{8.1}$$

occurs with a probability  $p(\rho)$ , in such a way that we know whether the transformation has occurred or not. This situation describes a general quantum measurement, in which an “occurrence flag” for the transformation represents the “outcome”, and the dependence of  $p(\rho)$  on  $\rho$  will give us some information on the state of the system. Since  $\mathcal{T}(\rho)$  is a quantum state, then the map  $\mathcal{T}$  satisfies for all  $\rho$

$$\mathcal{T}(\rho) \in \mathbf{T}(\mathbf{H}) , \quad \mathcal{T}(\rho) \geq 0 , \quad \text{and} \quad \text{tr } \mathcal{T}(\rho) = 1 , \quad (8.2)$$

where  $\mathbf{T}(\mathbf{H})$  denotes trace-class operators on  $\mathbf{H}$ . Consider now an ensemble of systems prepared as  $\{(p_i, \rho_i)\}$ . After the action of the device, the portion of systems having undergone the transformation is  $\sum_i p_i \mathcal{T}(\rho_i)$ , and the selection of these systems yields an ensemble described by the state

$$\rho' = \frac{\sum_i p_i \mathcal{T}(\rho_i)}{\sum_i p_i} .$$

On the other hand, the initial ensemble is also represented by the state  $\rho = \sum_i p_i \rho_i$ , so that the final post-selected ensemble will correspond to the state  $\mathcal{T}(\rho)$ , with a fraction of transformed systems equal to  $p(\rho)$ . The two descriptions must be consistent, because of the indistinguishability of two different preparations of the same ensemble, thus the fraction of transformed systems and the final state must be the same in both cases, namely

$$p(\rho) = \sum_i p_i p(\rho_i) , \quad (8.3)$$

$$\mathcal{T}(\rho) = \frac{\sum_i p_i p(\rho_i) \mathcal{T}(\rho_i)}{\sum_i p_i} . \quad (8.4)$$

The first equation implies that  $p(\rho)$  is a *linear* function of  $\rho$ , and, as we shall see later, this holds for any probability distribution of the outcomes of a quantum measurement, and it is unrelated to the details of the state transformation corresponding to each outcome: this will allow us to introduce the concept of POVM, which gives only the probability distribution of the outcomes as a function of the state. In the present context we are actually describing a “yes/no” measurement, i.e. our transformation  $\mathcal{T}$  “has” or “has not” occurred.

If we now introduce the map  $\mathcal{E}(\rho) \doteq p(\rho)\mathcal{T}(\rho)$ , Eq. (8.4) tells us that  $\mathcal{E}$  is a *linear* function of  $\rho$ . Taking the trace of the above definition of  $\mathcal{E}$ , and remembering that  $\text{tr } \mathcal{T}(\rho) = 1$ , we find  $p(\rho) = \text{tr } \mathcal{E}(\rho)$ , so that the transformation  $\mathcal{T}$  and the probability  $p(\rho)$  can be written in terms of  $\mathcal{E}$  as follows

$$\mathcal{T}(\rho) = \frac{\mathcal{E}(\rho)}{\text{tr } \mathcal{E}(\rho)} , \quad p(\rho) = \text{tr } \mathcal{E}(\rho) . \quad (8.5)$$

From Eq. (8.2), and from the fact that  $p(\rho)$  is a probability, one argues that also the following properties must hold for  $\mathcal{E}$

$$\begin{aligned} \mathcal{E}(\rho) &\geq 0 \quad (\textit{positivity}), \\ \text{tr } \mathcal{E}(\rho) &\leq 1 \quad (\textit{trace decreasing or preserving}) . \end{aligned} \quad (8.6)$$

The more stringent property of complete positivity for  $\mathcal{E}$  follows from the tensor-product structure of composite systems in Quantum Mechanics. In

fact, when  $\mathcal{T}$  acts only on a single subsystem of a bipartite quantum system, the joint state  $R$  of the system transforms according to

$$\rho \mapsto (\mathcal{E} \otimes \mathcal{I})(R),$$

and thus not only  $\mathcal{E}$  but also its extension  $\mathcal{E} \otimes \mathcal{I}$  must be positive, in such a way that the result of the local transformation is still a quantum state. This must hold for all possible extensions to larger composite systems. This property is called *complete positivity* and it is not equivalent to positivity, as counterexamples exist. For example, the transposition of the state with respect to a given basis  $\rho \mapsto \rho^T$  is a linear, positive, trace preserving map, but generally gives a non positive operator when acting on a system of an entangled pair, whence it is not completely positive and it can't be achieved physically. In the following we will refer to completely positive linear maps simply as *CP maps*, quantum operations corresponding to the class of trace non-increasing CP maps.

Up to now, we have shown that any transformation of the state of a quantum system is described by a *quantum operation* (QO), namely a linear, completely positive, trace non increasing map  $\mathcal{E} : \mathcal{T}(\mathcal{H}) \rightarrow \mathcal{T}(\mathcal{H})$ , with the state transformation given by  $\rho \mapsto \mathcal{E}(\rho) / \text{tr } \mathcal{E}(\rho)$ , and occurring with probability  $p(\rho) = \text{tr } \mathcal{E}(\rho)$ . Trace preserving QO's describe deterministic transformations—also called *quantum processes*, or *channels*)—namely with  $p(\rho) = 1$ , whereas trace decreasing QO's describe the transformation of the state of a system undergoing a quantum measurement for a given outcome occurring with a probability  $p(\rho) = \text{tr } \mathcal{E}(\rho) \leq 1$ .

### 8.1.2 Representing CP maps

In the following we will suppose  $\dim(\mathcal{H}) < \infty$ , whence we will generically denote trace-class, Hilbert-Schmidt and bounded operators on  $\mathcal{H}$  simply as  $\mathcal{B}(\mathcal{H})$ . CP maps are nothing but a special subset of the set of linear maps from  $\mathcal{B}(\mathcal{H})$  to  $\mathcal{B}(\mathcal{H})$ , and therefore they can be represented by means of their “matrix elements”

$$\mathcal{E}_{ij}^{lm} = \langle i | \mathcal{E}(|l\rangle\langle m|) | j \rangle, \quad (8.7)$$

so that, once defined  $\rho_{lm} = \langle l | \rho | m \rangle$ ,  $\mathcal{E}(\rho)$  can be evaluated as

$$\mathcal{E}(\rho) = \sum_{ijlm} \mathcal{E}_{ij}^{lm} \rho_{lm} |i\rangle\langle j|. \quad (8.8)$$

However, to have more insight into the structure of linear maps, it is preferable to reorganize the set of matrix elements  $\mathcal{E}_{ij}^{lm}$  into an operator on  $\mathcal{H} \otimes \mathcal{H}$ , aiming that the properties of the map (being CP, trace decreasing, invertible, etc.) have a simple translation into properties of the associated operator.

The following notation will be useful to simplify calculations by avoiding the use of a lot of indices in our equations, thus making them more insightful.

Fixed an orthonormal basis  $\{|m\rangle\}$  for the Hilbert space  $\mathsf{H}$ , we identify any vector  $|\Psi\rangle\rangle \in \mathsf{H} \otimes \mathsf{H}$ ,

$$|\Psi\rangle\rangle = \sum_{m,n} \Psi_{mn} |m\rangle \otimes |n\rangle , \quad (8.9)$$

with the operator  $\Psi \in \mathcal{B}(\mathsf{H})$  whose matrix elements on the chosen basis are  $\Psi_{mn}$ . For example, the vector  $|I\rangle\rangle$  represents the maximally entangled unnormalized vector  $\sum_m |m\rangle \otimes |m\rangle$ . It is easy to check that

$$\begin{aligned} A \otimes B |C\rangle\rangle &= |ACB^T\rangle\rangle , & \langle\langle A|B\rangle\rangle &= \text{tr}[A^\dagger B] , \\ \text{tr}_2[|A\rangle\rangle\langle\langle B|] &= AB^\dagger , & \text{tr}_1[|A\rangle\rangle\langle\langle B|] &= A^T B^* , \end{aligned} \quad (8.10)$$

where  $O^T$  and  $O^*$  denote respectively the transposition and the complex conjugation of the operator  $O$  with respect to the chosen basis.

Focusing our attention on the linearity of  $\mathcal{E}$ , with the notation introduced in Eq. (8.9), we notice that the vector  $|\mathcal{E}(\rho)\rangle\rangle$  is a linear transformation of  $|\rho\rangle\rangle$ , and thus the relation between the two vectors can be expressed by means of an operator  $\check{S}_{\mathcal{E}} \in \mathcal{B}(\mathsf{H} \otimes \mathsf{H})$  such that

$$|\mathcal{E}(\rho)\rangle\rangle = \check{S}_{\mathcal{E}} |\rho\rangle\rangle . \quad (8.11)$$

The map is faithfully represented by  $\check{S}_{\mathcal{E}}$ , since the previous relation defines its action on any state. By substituting in the above equation the definition of  $\mathcal{E}_{ij}^{lm}$  given in Eq. (8.7) one finds

$$\check{S}_{\mathcal{E}} = \sum_{ijlm} \mathcal{E}_{ij}^{lm} |i\rangle\langle l| \otimes |j\rangle\langle m| . \quad (8.12)$$

The power of this representation of linear maps resides in the fact that it translates the composition of two maps into the multiplication of their related operators, as one can easily verify from the following identity

$$|\mathcal{E}_1 \circ \mathcal{E}_2(\rho)\rangle\rangle = \check{S}_{\mathcal{E}_1} |\mathcal{E}_2(\rho)\rangle\rangle = \check{S}_{\mathcal{E}_1} \check{S}_{\mathcal{E}_2} |\rho\rangle\rangle . \quad (8.13)$$

Moreover, such a representation provides a useful tool to evaluate some properties of the map. For example, the image of the map  $\mathcal{E}(\mathcal{B}(\mathsf{H}))$  corresponds to the set the operators  $A$  such that  $|A\rangle\rangle \in \text{Rng } \check{S}_{\mathcal{E}}$ , where “Rng” denotes the range (i.e. the image) of an operator. Analogously, the kernel of  $\mathcal{E}$ , i.e. the set of operators  $A$  such that  $\mathcal{E}(A) = 0$ , is exactly the set of operators  $A$  such that  $|A\rangle\rangle \in \text{Ker } \check{S}_{\mathcal{E}}$ . Finally, by definition,  $\mathcal{E}$  is invertible iff  $\check{S}_{\mathcal{E}}$  is invertible, and the two inverses are related through the identity

$$|\mathcal{E}^{-1}(\rho)\rangle\rangle = \check{S}_{\mathcal{E}}^{-1} |\rho\rangle\rangle , \quad (8.14)$$

so that

$$|\mathcal{E}^{-1} \circ \mathcal{E}(\rho)\rangle\rangle = \check{S}_{\mathcal{E}}^{-1} \check{S}_{\mathcal{E}} |\rho\rangle\rangle = |\rho\rangle\rangle .$$

Being too much geared around linearity, unfortunately the above representation of maps tells us nothing about complete positivity. In order to explore this, it is convenient to introduce another operator representation of the map  $\mathcal{E}$  in terms of the operator  $S_{\mathcal{E}} \in \mathbf{B}(\mathsf{H} \otimes \mathsf{H})$  resulting from the action of the extended map  $\mathcal{E} \otimes \mathcal{I}$  on the operator  $|I\rangle\langle I| \in \mathbf{B}(\mathsf{H} \otimes \mathsf{H})$  [12,13], namely

$$S_{\mathcal{E}} = (\mathcal{E} \otimes \mathcal{I}) [|I\rangle\langle I|] = \sum_{ijlm} \mathcal{E}_{ij}^{lm} |i\rangle\langle j| \otimes |l\rangle\langle m|. \quad (8.15)$$

The inverse relation of identity (8.15) can be easily checked to be

$$\mathcal{E}(\rho) = \text{tr}_2[(I \otimes \rho^T) S_{\mathcal{E}}]. \quad (8.16)$$

A comparison between Eq. (8.12) and Eq. (8.15) shows that  $\check{S}_{\mathcal{E}}$  and  $S_{\mathcal{E}}$  are connected by a transposition of indices: if the matrix elements of the first are  $\mathcal{E}_{ij}^{lm}$ , the ones of the second are  $\mathcal{E}_{il}^{jm}$ , or in other terms

$$\check{S}_{\mathcal{E}} = (S_{\mathcal{E}}^{T_2} E)^{T_2} = (E S_{\mathcal{E}}^{T_1})^{T_1}, \quad (8.17)$$

where  $E = \sum_{ij} |i\rangle\langle j| \otimes |j\rangle\langle i|$  is the so called *swap* operator, and  $O^{T_l}$  denotes the partial transposition of the operator  $O$  on the  $l$ -th Hilbert space.

One immediately notices that if  $\mathcal{E}$  is CP, then  $S_{\mathcal{E}}$  is a positive operator, since it results from the application of the extension  $\mathcal{E} \otimes \mathcal{I}$  of a CP map to the positive operator  $|I\rangle\langle I|$ . Actually, the converse holds too, namely any map defined through Eq. (8.16) with  $S_{\mathcal{E}} \geq 0$  is CP. In fact, given that  $S_{\mathcal{E}}$  is positive, it can be decomposed as

$$S_{\mathcal{E}} = \sum_i |A_i\rangle\langle A_i|, \quad (8.18)$$

so that by substituting the above equation into Eq. (8.16), and applying the rules of Eq. (8.10), one finds that the resulting map can be expressed in the so called *Kraus form* [19]

$$\mathcal{E}(\rho) = \sum_i A_i \rho A_i^\dagger. \quad (8.19)$$

Any map of this form is completely positive, in fact the result of the action of its extension  $\mathcal{E} \otimes \mathcal{I}$  on a positive operator  $R \in \mathbf{B}(\mathsf{H} \otimes \mathsf{K})$  is

$$R_{\mathcal{E}} = \mathcal{E} \otimes \mathcal{I}[R] = \sum_i (A_i \otimes I) R (A_i^\dagger \otimes I), \quad (8.20)$$

which is still positive since

$$\langle\langle \Psi | R_{\mathcal{E}} | \Psi \rangle\rangle = \sum_i \langle\langle A_i^\dagger \Psi | R | A_i^\dagger \Psi \rangle\rangle \geq 0, \quad \forall |\Psi\rangle\rangle. \quad (8.21)$$

Of course, any CP map  $\mathcal{E}$  admits a Kraus form that can be found by decomposing  $S_{\mathcal{E}}$  as we did in Eq. (8.18). When this decomposition is a diagonalization, i.e., when  $|A_i\rangle\langle A_i|$  are the unnormalized orthogonal eigenvectors, then the related Kraus form is said to be *canonical*, and it has the minimum required number of operators, corresponding to the eigenvectors of  $S_{\mathcal{E}}$ , i.e. the cardinality of the Kraus decomposition is  $\text{rank } S_{\mathcal{E}}$ . Any couple of Kraus decompositions  $\{A_i\}$  and  $\{B_i\}$  are connected as  $B_i = \sum_j v_{ij} A_j$ , where  $v_{ij}$  is an isometry (i.e.  $\sum_j v_{ij} v_{jk} = \delta_{ik}$ ). In terms of a Kraus decomposition  $\{A_i\}$  of the map  $\mathcal{E}$  one can also express  $\check{S}_{\mathcal{E}}$  as

$$\check{S}_{\mathcal{E}} = \sum_i A_i \otimes A_i^*, \quad (8.22)$$

as it easily follows from the definition of  $\check{S}_{\mathcal{E}}$  in Eq. (8.11) and the first rule in Eq. (8.10).

Several properties other than complete positivity can be expressed in terms of  $S_{\mathcal{E}}$  or equivalently in terms of the elements of a Kraus decomposition  $\{A_i\}$ , for example, the trace decreasing condition becomes

$$\text{tr}_1 S_{\mathcal{E}} \leq I \quad \text{or equivalently} \quad \sum_i A_i^\dagger A_i \leq I, \quad (8.23)$$

where the equality sign would imply that the map is trace preserving.

If  $\text{rank } S_{\mathcal{E}} = 1$  then the map is *pure* (i.e. it preserves purity of input states), and its Kraus decomposition has only one element. Unitary evolutions are the only pure trace preserving transformations, and they play a special role since any other deterministic map can be realized as a unitary transformation acting on the system plus an ancilla whose state is then disregarded. In fact, given a Kraus decomposition  $\{A_i\}_{i=1\dots r}$  of the map  $\mathcal{E}$  one can define an operator  $U$  on the Hilbert space  $\mathbb{H} \otimes \mathbb{C}^r$  whose action on the vectors of the basis of the form  $|m\rangle|0\rangle$  is defined as

$$U|m\rangle|0\rangle = \sum_{i=1}^r (A_i|m\rangle) |i\rangle = |m, 0\rangle', \quad (8.24)$$

Since the map is trace preserving, then  $\sum_i A_i^\dagger A_i = I$ , and this assures that the resulting vectors  $|m, 0\rangle'$  in Eq. (8.24) are orthonormal: the operator  $U$  can then be easily extended to a unitary operator using a larger orthonormal set by means of the customary Gram-Schmidt procedure. By making the ancilla prepared in the state  $|0\rangle$  interact with the system in the state  $\rho$  by means of the unitary transformation  $U$ , the final “local” state of the system only reads

$$\mathcal{E}(\rho) = \text{tr}_2 [ U (\rho \otimes |0\rangle\langle 0|) U^\dagger ]. \quad (8.25)$$

Notice that instead of disregarding the ancilla as we did in the previous equation, one could instead perform a measurement on it, for example by

measuring the orthonormal basis  $|i\rangle$ , thus obtaining the state of the system in correspondence of the outcome  $i$  in terms of the pure trace decreasing quantum operation

$$\rho_i = \frac{A_i \rho A_i^\dagger}{\text{tr} [A_i \rho A_i^\dagger]} . \quad (8.26)$$

If we do not read the result of such a measurement, we still end up with a system in the state  $\mathcal{E}(\rho) = \sum_i p(i|\rho) \rho_i$ : the emergence of a non-pure quantum operation such as  $\mathcal{E}$  can be interpreted as a “measurement without reading the outcome”, or else as an information leakage in an environment. This is another way to understand how unitary operators describe the evolution of a closed system, whereas non pure trace preserving CP maps represent the evolution of open systems in interaction with a reservoir.

The procedure used to build  $U$  actually accomplishes a *purification* of  $\mathcal{E}$  that is analogous to the purification of a mixed state, and it is a sort of purification of the operator  $S_{\mathcal{E}}$ . It also returns unitaries to their privileged role at the axiomatic level.

As we argued from Eq. (8.26), it is possible to realize a trace decreasing map by means of a suitable joint unitary evolution of the system coupled with an ancilla, followed by a final projective measurement on the ancilla. Consider for example a measurement leading to  $N$  possible results  $\{1 \dots N\}$ , and such that in relation to the outcome  $k$  the state is transformed according to a map  $\mathcal{E}^{(k)}$  whose Kraus decomposition is  $\{A_i^{(k)}\}_{i=1 \dots r_k}$ . If we do not read the outcomes of the measurement and we do not separate the reduced systems accordingly, the final ensemble will be described by the state

$$\mathcal{E}(\rho) = \sum_k p(k|\rho) \frac{\mathcal{E}^{(k)}(\rho)}{\text{tr} \mathcal{E}^{(k)}(\rho)} = \sum_k \mathcal{E}^{(k)}(\rho) . \quad (8.27)$$

The map  $\mathcal{E}$  is a non-pure deterministic map admitting  $\{A_i^{(k)}\}$  as its Kraus decomposition, hence  $\sum_{k=1}^N \sum_{i=1}^{r_k} A_i^{\dagger(k)} A_i^{(k)} = I$ . If we define  $U$  on  $\mathbb{H} \otimes \mathbb{C}^{r_{\max}} \otimes \mathbb{C}^N$  such that on the elements of the basis of the form  $|m\rangle|0\rangle|0\rangle$  it behaves as

$$U|m\rangle|0\rangle|0\rangle = \sum_{k=1}^N \sum_{i=1}^{r_k} (A_i^{(k)}|m\rangle) |i\rangle|k\rangle , \quad (8.28)$$

then  $U$  can be completed to a unitary operator on the whole space by the Gram-Schmidt procedure, since the resulting vectors in the above equation are orthonormal. The original maps can now be realized by evolving the system in the state  $\rho$  jointly with the two additional ancillas prepared in the state  $|0\rangle|0\rangle$  with the unitary  $U$ , and then performing a projective measurement  $|k\rangle\langle k|$  on the second ancilla while disregarding the first one with a partial trace, i.e.

$$\mathcal{E}^{(k)}(\rho) = {}_3\langle k| \text{tr}_2 [U (\rho \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|) U^\dagger] |k\rangle_3 . \quad (8.29)$$

Also in this case, the maps  $\mathcal{E}^{(k)}$  are non-pure because some information has leaked into the first ancilla, which has been disregarded. If we would measure also the basis of the first ancilla, instead of taking the partial trace, in correspondence with the outcome  $(i, k)$  the state of the system would be described by a pure quantum operation

$$\rho_{(i,k)} = \frac{A_i^{(k)} \rho A_i^{(k)\dagger}}{\text{tr}[A_i^{(k)} \rho A_i^{(k)\dagger}]} . \quad (8.30)$$

### 8.1.3 Positive operator valued measures (POVM)

When what matters in a quantum measurement is only the probability distribution of outcomes in relation to the state of the system, we don't need the detailed description of the measurement process given in terms of quantum operations. Following reasoning lines similar to those followed in Sec. 8.1.1, in particular from Eq. (8.3), it follows that the probability distribution of the outcomes of any quantum measurement must be linear in the state  $\rho$ , and that therefore it is described by the so called Born's rule

$$p(k|\rho) = \text{tr}[\rho P_k] , \quad (8.31)$$

where  $k$  is the outcome and the set  $\{P_k\}$  is called *positive operator valued measure (POVM)*, namely it is a set of operators that must be positive and with  $\sum_k P_k = I$ , in order to have  $p(k|\rho)$  a properly positive and normalized probability distribution [20].

In the present context we are interested in deriving the connection between QO's and POVM's using the operator representation of maps, considering a measuring process for which each outcome  $k$  is described by the CP maps  $\mathcal{E}^{(k)}$ . By means of Eq. (8.16) the probability distribution of outcomes reads

$$p(k|\rho) = \text{tr}[\mathcal{E}^{(k)}(\rho)] = \text{tr}[\rho \text{tr}_1[S_{\mathcal{E}^{(k)}}]^T] = \text{tr}[\rho P_k] , \quad (8.32)$$

and thus the measurement maps  $\mathcal{E}^{(k)}$  induce the POVM elements  $P_k$  which can also be expressed as

$$P_k = \text{tr}_1[S_{\mathcal{E}^{(k)}}]^T = \sum_i A_i^{(k)\dagger} A_i^{(k)} , \quad (8.33)$$

$\{A_i^{(k)}\}$  being a Kraus decomposition of the  $k$ -th map. On the contrary, given a POVM  $\{P_k\}$  one can always find a set of QO's  $\mathcal{E}^{(k)}$  describing a measuring process with the given POVM, for example using  $A_k = \sqrt{P_k}$  and  $\mathcal{E}^{(k)}(\rho) = A^{(k)} \rho A^{(k)\dagger}$ . By "purifying" these maps with the unitary transformation  $U$  defined in Eq. (8.28) of the previous section, we see that it is possible to realize any POVM in terms of an indirect measurement scheme in which a projective measurement is performed on an ancilla after a unitary interaction with the system.

## 8.2 Imprinting quantum operations into quantum states

Characterizing a quantum device means performing a measurement that provides information about the QO performed by the device. However, quantum measurements can only give information about the state of a system, and that's why we need to devise a way to encode the information about the QO into a quantum state. This will then allow us to use the whole theory of state-discrimination and state-tomography also for discrimination and tomography of QO's.

The way to encode the QO of a device on the state is to let the device act on some systems suitably prepared, so that their final state contains an imprinting of the device. The aim of this section is to classify the input states that support a full imprinting of the QO of the device, i. e. what we call *faithful states*. We will also contextually consider the case in which the information on the QO is carried not by a single state, but by an ensemble of them, and we will correspondingly call the ensemble *faithful*.

After briefly recalling the first proposed methods for quantum process tomography [3,4], based on the use of many different input states, we shall see how a single pure entangled state can support a full imprinting of the QO [10]. Then we will extend the analysis to mixed states, showing how entanglement is not strictly needed [17], and finally giving a complete characterization of faithful states and ensembles, along with a measure of their “faithfulness”. It will become clear that the possibility of characterizing a device with a single fixed input state is a distinctive feature of quantum mechanics with no classical analog, and it is rooted in the tensor-product nature of composite quantum systems, instead of the cartesian-product “classical” composite systems. However, the fact that entanglement is not strictly necessary for faithfulness also indicates that the classical input-output correlations are enough to represent the device itself, but using an ensemble of state, whereas the possibility of imprinting a complete description of the device into these correlations for a “single passage” of the device intimately pertains to quantum mechanics.

In what follows, we will first restrict the analysis to devices performing quantum processes (i.e. deterministic QO's), and then extend the treatment to devices performing non deterministic QO's. Finally we will also present how to encode a POVM on quantum states [21].

### 8.2.1 Ensembles of input states versus a single entangled state

The first proposed methods for quantum process tomography [3,4] exploited the linearity of the map representing the process, and since a linear operator is defined by its action on a set of vectors spanning the Hilbert space, in the same way, any is completely defined by its action on a set of operators generating the linear space of all operators  $B(H)$ . Hence, for encoding a quantum process  $\mathcal{E}$  on states, one should look for a set of states  $\rho_i$  that span  $B(H)$ , since then their respective output states  $\mathcal{E}(\rho_i)$  would completely determine  $\mathcal{E}$ , namely

the set of states would be *faithful*. In fact, the action of the map  $\mathcal{E}$  on a generic state  $\rho$  can be recovered by expanding  $\rho$  on the generators of the space,  $\rho = \sum_i c_i \rho_i$ , so that by linearity one obtains the action of the map on any state  $\rho$  as  $\mathcal{E}(\rho) = \sum_i c_i \mathcal{E}(\rho_i)$ .

As an example, consider the set of states given in Ref. [22] for quantum process tomography

$$\left\{ |m\rangle, |\phi_{mn}\rangle = \frac{|m\rangle + |n\rangle}{\sqrt{2}}, |\psi_{mn}\rangle = \frac{|m\rangle + i|n\rangle}{\sqrt{2}} \right\} \quad (8.34)$$

it is a *faithful set of states*, as it is a set of generators for  $B(H)$  because the elements of the basis  $|m\rangle\langle n|$  of  $B(H)$  can be written as

$$|m\rangle\langle n| = |\phi_{mn}\rangle\langle\phi_{mn}| + i|\psi_{mn}\rangle\langle\psi_{mn}| - \frac{1+i}{2}|m\rangle\langle m| - \frac{1+i}{2}|n\rangle\langle n|. \quad (8.35)$$

Quantum process tomography has been realized with this method in liquid nuclear magnetic resonance systems [5–7], and for qubits encoded in the polarization of a radiation mode [8, 9], all situations where the dimension of the Hilbert space of the system is small. A method using the eigenstates of the quadrature operator as inputs has also been proposed in Ref. [23], for a phase-space representation of quantum transformations.

The above method has its main drawback in the difficulty—usually impossibility!—of preparing the needed number of the order of  $\dim(H)^2$  of different inputs. As we will see in the following, the method also turns out to be quite inefficient in achieving the information on the channel with a minimal number of measurements (the point is not that the device must be used several times to imprint the information on the channel only once, since quantum tomography even of a single output state would need many measurements).

A viable alternative to the above method of “spanning states” inspired by the operator representation of a channel was presented in Refs. [10] and [11], and experimentally implemented for polarization qubits in Ref. [14–16]. By preparing a bipartite system in the initial state  $R = |A\rangle\langle A|$  and letting the first subsystem evolve under the map, as depicted in Fig. 8.1, the output state  $R_{\mathcal{E}}$  reads

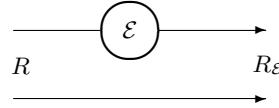
$$R_{\mathcal{E}} = (\mathcal{E} \otimes \mathcal{I}) [|A\rangle\langle A|] = (I \otimes A^T) S_{\mathcal{E}} (I \otimes A^*). \quad (8.36)$$

It is clear that whenever the operator  $A$  is invertible (i.e.  $A$  is full rank, or equivalently the bipartite system is in a maximal Schmidt’s number entangled state) it is possible to recover  $S_{\mathcal{E}}$  from  $R_{\mathcal{E}}$  by the simple inversion

$$S_{\mathcal{E}} = [I \otimes (A^T)^{-1}] R_{\mathcal{E}} [I \otimes (A^*)^{-1}], \quad (8.37)$$

and then the action of the map on a state  $\rho$  is found via Eq. (8.16), namely

$$\mathcal{E}(\rho) = \text{tr}_2[(I \otimes \rho^T) S_{\mathcal{E}}]. \quad (8.38)$$



**Fig. 8.1.** Encoding the information about a quantum device on an bipartite state. Two identical quantum systems are prepared in the state  $R$ . One of the two systems enters the device and undergoes the map  $\mathcal{E}$ , whereas the other is left untouched. The joint output state contains information on  $\mathcal{E}$ . When such information is complete the state  $R$  is called *faithful*. A pure input  $R = |A\rangle\langle A|$  is faithful iff  $\text{rank } A = \dim(\mathcal{H})$ .

Summarizing, any bipartite state with maximal Schmidt number is faithful, namely by entering a quantum device it gets imprinted the full information about its channel. This method for encoding a channel on a state exploits the quantum parallelism of entanglement, with a fixed bipartite entangled state playing the role of the several input states of the previous method. The information on the device is encoded in a “native” way, which perfectly reflects the nature of the CP map representing the device itself. Moreover, it is encoded with a single use of the device, in contrast to the many uses of the method based on the generating set of states (this feature can be exploited at best in the context of devices discrimination [18], where a single measurement is allowed). Of course, when no prior knowledge of the device is provided, in order to recover the encoded information we have to perform a full quantum tomography of the output state, whence many copies of the imprinted state are still necessary. However, the main advantage of the method based on a single entangled state resides on the fact that a generating set of states is often not available in the lab, whereas we can produce entangled states: this is the case, for example, of quantum optics (in the domain of so-called *continuous variables* in contrast single qubits encoded on polarization of single photons), where a faithful entangled state is provided by a twin-beam from parametric down-conversion of vacuum, whereas photon number states and their superpositions as in Eq. (8.34) will remain an impossible dream for many years. Another relevant advantage of the single-pure-state method versus the generating-set one is a much higher statistical efficiency, i.e. the number of measurements needed to achieve a given statistical error in the reconstruction of the map of the device. In addition, thanks to the “native way” of encoding the transformation—reflecting both complete positivity and trace preserving/decreasing property of the map—the use of the single input state allows an easy implementation of the maximum likelihood strategies for the characterization of the device.

All the above observations will be analyzed in detail later in this section, when a measure of “faithfulness” will be introduced, and also in the next section, where some practical applications of this framework for characterizing quantum devices will be exposed.

### 8.2.2 Faithful states

In the previous paragraph we showed that a pure entangled bipartite state  $|A\rangle\langle A|$  supports the imprinting of a quantum channel whenever the operator  $A$  is invertible. Here we want to extend this result to a generally non pure input state  $R$ , in order to characterize all faithful states.

So let's consider a bipartite state  $R$ , with spectral decomposition  $R = \sum_l |A_l\rangle\langle A_l|$ . By applying the relation  $|A_l\rangle\langle A_l| = (I \otimes A_l^T)|I\rangle\langle I|$ , we can rewrite the corresponding output state  $R_{\mathcal{E}} = (\mathcal{E} \otimes \mathcal{I})[R]$  as

$$\begin{aligned} R_{\mathcal{E}} = (\mathcal{E} \otimes \mathcal{I})[R] &= \sum_l (I \otimes A_l^T) (\mathcal{E} \otimes \mathcal{I})[|I\rangle\langle I|] (I \otimes A_l^*) = \\ &= \sum_l (I \otimes A_l^T) S_{\mathcal{E}} (I \otimes A_l^*) . \end{aligned} \quad (8.39)$$

If we define the completely positive map  $\mathcal{R}$  as

$$\mathcal{R}(\rho) = \sum_l A_l^T \rho A_l^* , \quad (8.40)$$

it is immediate to notice that

$$R_{\mathcal{E}} = (\mathcal{I} \otimes \mathcal{R}) [S_{\mathcal{E}}] , \quad (8.41)$$

and therefore whenever the map  $\mathcal{R}$  is invertible the output state  $R_{\mathcal{E}}$  will be in one-to-one correspondence with  $S_{\mathcal{E}}$ , and thus with the map  $\mathcal{E}$ , namely it will contain all the information about the map.

From the above considerations it follows that the input state  $R$  is faithful iff it leads to a map  $\mathcal{R}$  that is invertible. Recalling what we wrote in Sec. 8.1.2, and in particular Eq. (8.14), the invertibility of the CP map  $\mathcal{R}$  resorts to the invertibility of a customary operator. In fact, by considering the following equation involving vectors in  $\mathsf{H} \otimes \mathsf{H}$

$$|\mathcal{R}(\rho)\rangle\langle\rho| = |\sum_l A_l^T \rho A_l^*\rangle\langle\rho| = (\sum_l A_l^T \otimes A_l^\dagger) |\rho\rangle\langle\rho| \doteq \check{S}_{\mathcal{R}} |\rho\rangle\langle\rho| , \quad (8.42)$$

one realizes that the map  $\mathcal{R}$  is invertible iff the relation between vectors  $|\mathcal{R}(\rho)\rangle\langle\rho|$  is invertible, and looking at the above equation it is clear that this happens iff the operator  $\check{S}_{\mathcal{R}} \doteq \sum_l A_l^T \otimes A_l^\dagger$  on  $\mathsf{H} \otimes \mathsf{H}$  is invertible. As we already noticed in Sec. 8.1.2, the action of the inverse map  $\mathcal{R}^{-1}$  can be defined through the relation

$$|\mathcal{R}^{-1}(\rho)\rangle\langle\rho| \doteq \check{S}_{\mathcal{R}}^{-1} |\rho\rangle\langle\rho| , \quad (8.43)$$

so that  $|\mathcal{R}^{-1}(\mathcal{R}(\rho))\rangle\langle\rho| = \check{S}_{\mathcal{R}}^{-1} \check{S}_{\mathcal{R}} |\rho\rangle\langle\rho| = |\rho\rangle\langle\rho|$ . The operator  $\check{S}_{\mathcal{R}}$  can be expressed directly in terms of  $R$ , without having to evaluate its spectral decomposition, as

$$\check{S}_{\mathcal{R}} = (ER)^{T_2} E = (R^{T_2} E)^{T_1} \quad (8.44)$$

where  $E = \sum_{ij} |ij\rangle\langle ji|$  is the swap operator, and  $O^{T_l}$  denotes the partial transposition of the operator  $O$  on the  $l$ th Hilbert space.

In summary, we have found that  $R$  is faithful iff  $\check{S}_{\mathcal{R}}$  is invertible. In this case the relation between the output state  $R_{\mathcal{E}} = (\mathcal{E} \otimes \mathcal{I})[R]$  and the operator  $S_{\mathcal{E}}$  is one-to-one, with all the information about the CP map  $\mathcal{E}$  encoded in  $R_{\mathcal{E}}$ . The map  $\mathcal{E}$  can be recovered from the joint output state  $R_{\mathcal{E}}$  as follows

$$\mathcal{E}(\rho) = \text{tr}_2 \left[ (I \otimes \rho^T) (\mathcal{I} \otimes \mathcal{R}^{-1})[R_{\mathcal{E}}] \right]. \quad (8.45)$$

Later we will show some examples of faithful states, and among them there will be also separable states. On first sight this may be surprising, but it becomes obvious if one realizes that the set of faithful states is *dense*, because it is related to the set of invertible operators which is dense too.

As a further generalization, we now discuss the faithfulness of the bipartite state  $R$  of two quantum systems described by different Hilbert spaces  $H$  and  $K$ . We need now to consider vectors in either  $H \otimes K$ ,  $H^{\otimes 2}$ , or  $K^{\otimes 2}$ , and in all cases we will keep our notation  $|A\rangle\langle A|$  for the vectors, with the corresponding operator  $A$  in  $B(K, H)$ ,  $B(H)$ , or  $B(K)$  respectively.

Similarly to the previous reasoning lines, in relation to the bipartite input state  $R = \sum_l |A_l\rangle\langle A_l|$  on  $H \otimes K$ , the output reads  $R_{\mathcal{E}} = \mathcal{I} \otimes \mathcal{R}[S_{\mathcal{E}}]$ , where the map  $\mathcal{R}(\rho) = \sum_l A_l^T \rho A_l^*$  now is from  $B(H)$  to  $B(K)$ . Then, faithfulness of  $R$  is still equivalent to the invertibility of the map  $\mathcal{R}$ , but now it is more generally equivalent to its *left*-invertibility<sup>1</sup>. The operator  $\check{S}_{\mathcal{R}} = \sum_l A_l^T \otimes A_l^\dagger$  associated to  $\mathcal{R}$  now maps vectors in  $H^{\otimes 2}$  to vectors in  $K^{\otimes 2}$ , and it is still such that  $\check{S}_{\mathcal{R}}|\rho\rangle\langle\rho| = |\mathcal{R}(\rho)\rangle\langle\mathcal{R}(\rho)|$ . Again, faithfulness of  $R$  is equivalent to left-invertibility of the operator  $\check{S}_{\mathcal{R}}$  from  $H^{\otimes 2}$  to  $K^{\otimes 2}$ , that in turn is equivalent to the condition  $\text{rank } \check{S}_{\mathcal{R}} = \dim(H)^2$ . Among all the possible left-inverses of  $\check{S}_{\mathcal{R}}$  one can use the Moore-Penrose pseudo-inverse  $\check{S}_{\mathcal{R}}^\dagger$ , and thus define the left-inverse of the map  $\mathcal{R}$  as

$$|\mathcal{R}^{-1}(\rho)\rangle\langle\rho| \doteq \check{S}_{\mathcal{R}}^\dagger|\rho\rangle\langle\rho|, \quad (8.46)$$

so that one can recover  $S_{\mathcal{E}}$  from  $R_{\mathcal{E}}$  by the relation  $S_{\mathcal{E}} = (\mathcal{I} \otimes \mathcal{R}^{-1})[R_{\mathcal{E}}]$ .

---

<sup>1</sup> A generic operator  $T : H \rightarrow K$  is left-invertible iff  $\text{rank } T = \dim(H)$ . For having  $T$  left-invertible is therefore necessary that  $\dim(K) \geq \dim(H)$ , the inverse being unique whenever the equality holds, whereas non-unique in the case of a strict inequality. Among the infinitely many possible left-inverses, the Moore-Penrose pseudo-inverse  $T^\dagger$  [24] is the most used one, due to its nice properties. Starting from the *singular values decomposition* (SVD) of  $T$

$$T = \sum_i \sigma_i |v_i\rangle\langle u_i|, \quad (8.47)$$

where  $\{|v_i\rangle\}$  and  $\{|u_i\rangle\}$  are two sets of orthonormal vectors, and  $\sigma_i$  are positive real numbers (the singular values),  $T^\dagger$  is defined as

### 8.2.3 A measure of faithfulness

Even though in principle any faithful state can be used for encoding quantum processes on their outputs, the actual choice of the input will be dictated by some figure of merit depending on the particular situation. For example, consider the case in which we want to discriminate between two processes  $\mathcal{E}_1$  and  $\mathcal{E}_2$ . For input state  $R$ , their respective outputs will be

$$R_{\mathcal{E}_1} = (\mathcal{I} \otimes \mathcal{R}) [S_{\mathcal{E}_1}] \quad \text{and} \quad R_{\mathcal{E}_2} = (\mathcal{I} \otimes \mathcal{R}) [S_{\mathcal{E}_2}], \quad (8.49)$$

and thus we shall tune  $R$  in order to improve the distinguishability of these two outputs.

More generally, we see that an overall performance indicator for the faithfulness of the state  $R$  is a measure of its ability to keep outputs corresponding to different processes as far as possible in average, namely the ability of the map  $\mathcal{R}$  in Eq. (8.49) to keep its outputs as far as possible. By considering the singular value decomposition for the operator  $\check{S}_{\mathcal{R}}$

$$\check{S}_{\mathcal{R}} = \sum_i \sigma_i |V_i\rangle\langle U_i|, \quad (8.50)$$

with  $\{|V_i\rangle\}$  and  $\{|U_i\rangle\}$  sets of orthonormal vectors, and  $\sigma_i > 0$ , and by remembering that  $|\mathcal{R}(\rho)\rangle = \check{S}_{\mathcal{R}}|\rho\rangle$ , the action of  $\mathcal{R}$  on an operator  $\rho$  becomes

$$\mathcal{R}(\rho) = \sum_i \sigma_i \operatorname{tr}[U_i^\dagger \rho] V_i, \quad (8.51)$$

whence it is clear that the smaller are the singular values  $\sigma_i$ , the nearer are the outputs of  $\mathcal{R}$ , since their components on the basis  $\{|V_i\rangle\}$  will be shrunk. Therefore, in summary, the larger are the singular values of  $\check{S}_{\mathcal{R}}$  the better is the chosen input state  $R$ .

Thus, a synthetic measure of faithfulness could be for example

$$F(R) = \sum_i \sigma_i^2 = \operatorname{tr}[\check{S}_{\mathcal{R}}^\dagger \check{S}_{\mathcal{R}}]. \quad (8.52)$$

This quantity can be expressed in a more meaningful form by observing that if we use the spectral decomposition  $R = \sum_i |A_i\rangle\langle A_i|$ , with the vectors  $|A_i\rangle\langle A_i|$  being an orthogonal basis, namely  $\langle A_i | A_j \rangle = \operatorname{tr}[A_i^\dagger A_j] \propto \delta_{ij}$ , then  $\check{S}_{\mathcal{R}} = \sum_i A_i^* \otimes A_i^\dagger$ , and thus the following equations hold

---


$$T^\ddagger = \sum_i \sigma_i^{-1} |u_i\rangle\langle v_i|. \quad (8.48)$$

By definition,  $Q = T^\ddagger T$  is the orthogonal projector on  $\operatorname{Supp}(T) \equiv \operatorname{Ker}(T)^\perp$ , whence  $T^\ddagger$  inverts  $T$  on its support, which for a left-invertible operator coincides with the whole space  $\mathbb{H}$ .

$$\begin{aligned} \text{tr}[\check{S}_{\mathcal{R}}^\dagger \check{S}_{\mathcal{R}}] &= \sum_{ij} \text{tr}[A_i^T A_j^*] \text{tr}[A_i A_j^\dagger] = \sum_i \text{tr}[A_i^T A_i^*] \text{tr}[A_i A_i^\dagger] = \\ &= \sum_i (\langle\langle A_i | A_i \rangle\rangle)^2 = \text{tr}[R^\dagger R]. \end{aligned} \quad (8.53)$$

Therefore, from Eq. (8.52) one obtains

$$F(R) = \text{tr}[R^\dagger R], \quad (8.54)$$

so that faithfulness of a state turns out to be exactly its purity. This result implies that *faithful pure states* are the optimal faithful states, and that they yield outputs states encoding the maps that are the most far apart.

The definition of  $F(R)$  can be also interpreted in another way. Imagine to implement quantum process tomography using a finite number of copies of  $R$  as input states, and then reconstruct the output  $R_{\mathcal{E}}$ . The measured  $R_{\mathcal{E}}$  will be affected by experimental errors that will be mostly independent on  $R_{\mathcal{E}}$  itself, and these errors will be propagated to the experimental estimation of  $S_{\mathcal{E}}$  by the inversion map  $\mathcal{R}^{-1}$ . Since, in practice, the inversion map involves multiplications by  $\sigma_i^{-1}$ , then the smaller the singular values of  $\check{S}_{\mathcal{R}}$  are the higher the amplification of experimental errors on the measured  $S_{\mathcal{E}}$ .

For an unfaithful state  $R$ ,  $\check{S}_{\mathcal{R}}$  will have at least one null singular value, yet  $F(R)$  is different from zero. Actually, as we shall see, the state can still be used to recover the action of a device on some inputs only. Moreover, on such inputs it can achieve an even better reconstruction resolution than a faithful state, since its faithfulness is focused on a smaller subspace.

#### 8.2.4 Faithful ensembles of states

Now we will consider the case in which not a single bipartite state, but an ensemble of them  $\{R^{(n)}\}_{n=1}^N$  on  $\mathsf{H} \otimes \mathsf{K}$  is used, and we want to understand whether or not it is faithful, namely if it supports a complete imprinting of the information about a quantum process. In other words, we will discuss when the set of outputs  $\{R_{\mathcal{E}}^{(n)}\}$ , with  $R_{\mathcal{E}}^{(n)} = (\mathcal{E} \otimes \mathcal{I})[R^{(n)}]$ , is a perfect encoding of a generic channel  $\mathcal{E}$ . This analysis will bridge the scenario with the set of generating states and the one of single bipartite faithful state.

Mathematically, it is evident that the state  $R_{\text{set}}$  on  $\mathsf{H} \otimes \mathsf{K} \otimes \mathbb{C}^N$  defined as

$$R_{\text{set}} = \sum_{n=1}^N p_n R^{(n)} \otimes |n\rangle\langle n|, \quad (8.55)$$

where  $p_n$  are fixed non vanishing probabilities, is in 1-to-1 correspondence with the set of states  $\{R^{(n)}\}$ . The same correspondence holds between the output state

$$R_{\text{set}} \mathcal{E} = (\mathcal{E} \otimes \mathcal{I} \otimes \mathcal{I})[R_{\text{set}}] = \sum_{n=1}^N p_n R_{\mathcal{E}}^{(n)} \otimes |n\rangle\langle n| \quad (8.56)$$

and the set of outputs  $\{R_{\mathcal{E}}^{(n)}\}$ . Hence, if the state  $R_{\text{set}, \mathcal{E}}$  contains all the information about the map, then the same holds also for the set of outputs  $\{R_{\mathcal{E}}^{(n)}\}$ , or, equivalently, if  $R_{\text{set}}$  is faithful, then the set  $\{R^{(n)}\}$  is faithful too.

Briefly, faithfulness for the set of states  $\{R^{(n)}\}$  is translated into faithfulness for the single state  $R_{\text{set}}$ . The latter can be evaluated with the techniques exposed in the previous paragraph for bipartite states, by simply considering  $R_{\text{set}}$  as a bipartite state of  $\mathbf{H}$  and  $\mathbf{K} \otimes \mathbb{C}^N$ .

The nature of the state  $R_{\text{set}}$  can be interpreted from two subtly different points of view. On one hand, to use  $R_{\text{set}}$  is equivalent to running all the states  $\{R^{(n)}\}$  in parallel, while keeping track of each of them thanks to the tensoring with the basis  $|n\rangle\langle n|$  of  $\mathbb{C}^N$ . On the other hand,  $R_{\text{set}}$  represents the situation in which the states  $\{R^{(n)}\}$  are employed in the characterization each with a frequency equal to  $p_n$ . In fact, measuring the basis  $|n\rangle\langle n|$  on  $\mathbb{C}^N$  (either before or after the action of the device) is equivalent to preparing the input  $R^{(n)}$  with a probability  $p_n$ , where  $n$  is the outcome of the measurement.

For this reason, any quantity (e.g. the faithfulness  $F$ ) being defined for faithful states can be extended consistently to ensembles of states simply by evaluating it on the corresponding  $R_{\text{set}}$ . For example, the faithfulness of a set of generating states  $\rho_n$  (employed with the same frequency) is equivalent to the faithfulness of the bipartite state  $R_{\text{set}} = \sum_n \frac{1}{n} \rho_n \otimes |n\rangle\langle n|$ , and since the latter is a mixed state, it will lead to a non-optimal faithfulness. This shows why the encoding on an entangled state is theoretically better than the encoding on a set of generating states: while in the first case faithfulness is 1, in the second one it scales as  $\mathcal{O}[1/\dim(\mathbf{H})]$ .

### 8.2.5 Patching sets of unfaithful states

An unfaithful state  $R$  can still be useful in encoding only some quantum channels, or at least in encoding partial information about them, which can then be used to evaluate their action on some particular states. In fact, even if the map  $\mathcal{R}$  is not invertible (it maps to zero any state  $\rho$  such that  $|\rho\rangle\rangle \in \text{Ker}(\check{S}_{\mathcal{R}})$ ), one can still employ its pseudo-inverse  $\mathcal{R}^\ddagger$  defined as

$$|\mathcal{R}^\ddagger(\rho)\rangle\rangle \doteq \check{S}_{\mathcal{R}}^\ddagger |\rho\rangle\rangle. \quad (8.57)$$

This map is such that  $\mathcal{R}^\ddagger \mathcal{R} = \mathcal{Q}$ , where  $\mathcal{Q}$  is the projection map on the support of the map  $\mathcal{R}$ , and which is also defined by

$$|\mathcal{Q}(\rho)\rangle\rangle = \check{S}_{\mathcal{R}}^\ddagger \check{S}_{\mathcal{R}} |\rho\rangle\rangle = \check{S}_{\mathcal{Q}} |\rho\rangle\rangle, \quad (8.58)$$

the operator  $\check{S}_{\mathcal{Q}}$  being the projector on  $\text{Supp}(\check{S}_{\mathcal{R}}) = \text{Ker}(\check{S}_{\mathcal{R}})^\perp$ .

It is clear that such pseudo-inversion, instead of using the full operator  $S_{\mathcal{E}}$ , corresponds to its projection

$$\tilde{S}_{\mathcal{E}} = (\mathcal{I} \otimes \mathcal{R}^\ddagger)[R_{\mathcal{E}}] = (\mathcal{I} \otimes \mathcal{Q})[S_{\mathcal{E}}] \quad (8.59)$$

which represents a partial encoding of  $\mathcal{E}$ . The partially recovered map  $\tilde{\mathcal{E}}(\rho) = \text{tr}_2[(I \otimes \rho^T) \tilde{S}_{\mathcal{E}}]$  could have also been written as  $\tilde{\mathcal{E}} = \mathcal{E}\mathcal{Q}^*$ ,  $\mathcal{Q}^*$  being the map corresponding to the operator  $\tilde{S}_{\mathcal{Q}}^*$ . Clearly  $\tilde{\mathcal{E}}$  coincides with  $\mathcal{E}$  for any  $\rho$  such that  $\tilde{S}_{\mathcal{Q}}^*|\rho\rangle\rangle = |\rho\rangle\rangle$ .

For any bipartite  $R$  one can define a *number of faithfulness*  $\varphi$  as  $\varphi(R) = \text{rank}(\tilde{S}_{\mathcal{R}})$ , i.e. as the dimension of the space of input states  $R$  for which the action of the map  $\mathcal{E}$  is described faithfully. Clearly, a state is faithful iff  $\varphi(R) = \dim(\mathcal{H})^2$ . Notice that for  $\varphi(R) < \dim(\mathcal{H})^2$  one can have the situation in which  $\text{Ker}^\perp(\tilde{S}_{\mathcal{R}}) = \text{Span}\{|\rho\rangle\rangle, \rho \in \mathcal{A}\}$ , with  $\mathcal{A}$  being Abelian algebra, in which case the state  $R$  allows to reconstruct completely only “classical” channels, with the input restricted to commuting states.

The introduction of pseudo-inversion provides an alternative yet equivalent way for studying the faithfulness of a set of states  $\{R^{(n)}\}$ . Suppose they lead to the projection maps  $\{\mathcal{Q}^{(n)}\}$ , then the set will be faithful iff we can recover any operator  $\rho$  from its projections  $\mathcal{Q}^{(n)}(\rho)$ , and this is possible iff, given a basis  $\{B_i\}$  for  $\mathcal{B}(\mathcal{H})$ , one has  $\text{Span}\{\mathcal{Q}^{(n)}(B_i)\}_{i,n} = \mathcal{B}(\mathcal{H})$ . In such circumstances, any element of the basis can be expressed as a linear combination of the  $\mathcal{Q}^{(n)}(B_i)$ , i.e.  $B_i = \sum_{jn} \lambda_{ij}^n \mathcal{Q}^{(n)}(B_j)$ , and therefore it is possible to recover  $M \equiv \sum_i \text{tr}[B_i^\dagger M] B_i$  by “patching” the projections  $\mathcal{Q}^{(n)}(M)$  as

$$M = \sum_{ijn} \lambda_{ij}^n \text{tr}[B_j^\dagger \mathcal{Q}^{(n)}(M)] B_i . \quad (8.60)$$

Analogously, by patching the partial encodings  $\{\tilde{S}_{\mathcal{E}}^{(n)}\}$  (see Eq. (8.59)) we get  $S_{\mathcal{E}}$  as

$$S_{\mathcal{E}} = \sum_{ijn} \lambda_{ij}^n \text{tr}_2[(I \otimes B_j^\dagger) \tilde{S}_{\mathcal{E}}^{(n)}] \otimes B_i . \quad (8.61)$$

Of course this patching procedure can also be used with an unfaithful set of states, to obtain a more complete yet still partial encoding of the channel.

### 8.2.6 Generalization to QO's and POVM's

Suppose we have a quantum device performing the measurement described by the CP maps  $\mathcal{E}_i$ ,  $i = 1 \dots N$  being the outcomes, is it possible to encode all the maps or else their corresponding POVM? If we use a bipartite input state  $R$  and we let the device act on the first subsystem, the output state corresponding to the outcome  $i$  will be

$$R_{\mathcal{E}_i} = \frac{(\mathcal{I} \otimes \mathcal{R}) [S_{\mathcal{E}_i}]}{\text{tr}[(\mathcal{I} \otimes \mathcal{R}) [S_{\mathcal{E}_i}]]} , \quad (8.62)$$

where the denominator is also the probability of occurrence for the outcome  $i$ . In the case of  $R$  faithful, from this output it is possible to recover  $S_{\mathcal{E}_i}$  up to a normalization factor by means of the inverse map  $\mathcal{R}^{-1}$ .

After preparing an ensemble of systems described by a faithful state  $R$ , we let the measuring device act on them, and then we separate them according to the outcome  $i$ , thus obtaining  $N$  different ensembles, each labeled by the corresponding  $i$ , and described by the states  $R_{\mathcal{E}_i}$ . The denominator of Eq. (8.62) can be evaluated as the fraction of systems of the original ensemble that have been transformed into the  $i$ -th state, therefore an exact reconstruction of all the  $S_{\mathcal{E}_i}$  is possible, being equivalent to the full reconstruction of the measuring device. Notice that, in contrast to what happens for a deterministic device, in the case of a probabilistic QO a single use is not enough to imprint the whole information about it, due to of the need for evaluating the normalization factor.

In many practical situations, e.g. in a photodetector, the measuring device destroys the measured system. Here, however, with the same setup with a bipartite faithful  $R$ , the reduced state  $\rho_i$  on the unmeasured system is still available. It reads,

$$\rho_i = \text{tr}_1 R_{\mathcal{E}_i} = \frac{\mathcal{R}[\text{tr}_1 S_{\mathcal{E}_i}]}{\text{tr}[\mathcal{R}[\text{tr}_1 S_{\mathcal{E}_i}]]} = \frac{\mathcal{R}[P_i^T]}{\text{tr}[\mathcal{R}[P_i^T]]}, \quad (8.63)$$

where  $P_i$  is the POVM of the measurement relative to the outcome  $i$ . Hence, by performing a quantum tomography on the above reduced output states, one can recover the POVM of the apparatus by inverting the map  $\mathcal{R}$ , while evaluating the denominator of the previous equation as the probability of occurrence of  $i$ .

### 8.2.7 Faithfulness and separability

Since, as we have seen, faithfulness is equivalent to an invertibility condition, the set of faithful states  $R$  is *dense* within the set of all bipartite states. Therefore, there must be faithful states among mixed separable ones, which means that classical correlations in mixed bipartite states are sufficient to support the imprinting of any quantum channel. Let us see some examples of separable faithful states.

The Werner's states for dimension  $d$

$$R_f = \frac{1}{d(d^2 - 1)} [(d - f)I + (df - 1)E], \quad -1 \leq f \leq 1, \quad (8.64)$$

are separable for  $f \geq 0$ , however, they are faithful for all  $f \neq \frac{1}{d}$ . In fact, one has

$$(ER_f)^{T_2} = \frac{1}{d(d^2 - 1)} [(d - f)|I\rangle\langle I| + (df - 1)], \quad (8.65)$$

hence the singular values of  $\check{S}_{R_f}$  are  $\frac{df-1}{d(d^2-1)}$  with multiplicity  $d^2 - 1$  and  $\frac{1}{d}$  with multiplicity 1. In Ref. [15] an experiment employing these states for quantum process tomography was presented.

Similarly, the “isotropic” states

$$R_f = \frac{f}{d}|I\rangle\langle I| + \frac{1-f}{d^2-1}(I - \frac{1}{d}|I\rangle\langle I|), \quad (8.66)$$

are faithful for  $f \neq \frac{1}{d^2}$  and separable for  $f \leq \frac{1}{d}$ , the singular values of  $\check{S}_{\mathcal{R}_f}$  being  $\frac{d^2f-1}{d(d^2-1)}$  and  $\frac{f}{d}$ .

### 8.2.8 Faithfulness in infinite dimensions

For infinite dimensions (the so-called “continuous variables” in quantum optics), one needs to restrict  $B(H)$  to the Hilbert space of Hilbert-Schmidt operators on  $H$ , and this leads to the problem that the inverse map  $\mathcal{R}^{-1}$  is unbounded. The result is that we will recover the channel  $\mathcal{E}$  from the measured  $R_{\mathcal{E}}$ , however, with unbounded amplification of statistical errors, depending on the chosen complete set of operators  $B = \{B_j\}$  in  $B(H)$  used for representing the channel map. As an example, let’s consider a twin beam from parametric down-conversion of vacuum

$$|\Psi\rangle\langle\Psi| = \Psi \otimes I|I\rangle\langle I|, \quad \Psi = (1 - |\xi|^2)^{\frac{1}{2}}\xi^{a^\dagger a}, \quad |\xi| < 1 \quad (8.67)$$

for a fixed  $\xi$ ,  $a^\dagger$  and  $a$ , with  $[a, a^\dagger] = 1$ , denoting the creation and annihilation operators of the harmonic oscillator describing the field mode corresponding to the first Hilbert space in the tensor product (in the following we will denote by  $b^\dagger$  and  $b$  the creation and annihilation operators of the other field mode). The state is faithful, but the operator  $\Psi^{-1}$  is unbounded, whence the inverse map  $\mathcal{R}^{-1}$  is also unbounded. In a photon number representation  $B = \{|n\rangle\langle m|\}$ , the effect will be an amplification of errors for increasing numbers  $n, m$  of photons.

As an example, consider the quantum channel describing the *Gaussian displacement noise* [25]

$$\mathcal{N}_{\nu}(\rho) = \int_{\mathbb{C}} \frac{d\alpha}{\pi\nu} \exp[-|\alpha|^2/\nu] D(\alpha)\rho D^\dagger(\alpha), \quad (8.68)$$

where  $D(\alpha) = \exp(\alpha a^\dagger - \alpha^* a)$  denotes the usual displacement operator on the phase space. The Gaussian noise is, in a sense, analogous of the depolarizing channel for infinite dimension. The maps  $\mathcal{N}_{\nu}$  for varying  $\nu$  satisfy the multiplication rule  $\mathcal{N}_{\nu}\mathcal{N}_{\mu} = \mathcal{N}_{\nu+\mu}$ , thus the inverse map is formally given by  $\mathcal{N}_{\nu}^{-1} \equiv \mathcal{N}_{-\nu}$ . Notice that, since the map  $\mathcal{N}_{\nu}$  is compact, the inverse map  $\mathcal{N}_{\nu}^{-1}$  is necessarily unbounded. As a faithful state consider now the mixed state given by the twin-beam, with one beam spoiled by the Gaussian noise, namely

$$R = \mathcal{I} \otimes \mathcal{N}_{\nu}(|\Psi\rangle\langle\Psi|). \quad (8.69)$$

Since the (unnormalizable) vector  $|D(z)\rangle\langle D(z)| = [D(z) \otimes I]|I\rangle\langle I|$  is a eigenvector of the operator  $Z = a - b^\dagger$ , with eigenvalue  $z$ , one can easily find that

$$R = \frac{1}{\nu} (\Psi \otimes I) \exp[-(a - b^\dagger)(a^\dagger - b)/\nu] (\Psi^\dagger \otimes I), \quad (8.70)$$

thus its partial transposed on the second space reads

$$R^{T_2} = (\nu + 1)^{-1} (\Psi \otimes I) \left( \frac{\nu - 1}{\nu + 1} \right)^{\frac{1}{2}(a-b)^\dagger(a-b)} (\Psi^\dagger \otimes I), \quad (8.71)$$

where transposition is defined with respect to the basis of eigenvectors of  $a^\dagger a$  and  $b^\dagger b$ . Since our state  $R$  is Gaussian, it is separable iff its partial transposition is a positive operator [26], therefore, for  $\nu > 1$ ,  $R$  is separable (see also Ref. [27]), yet it is *formally* faithful, since the operator  $\Psi$  and the map  $N_\nu$  are both invertible. Notice that unboundedness of the inversion map can even wash out completely the information on the channel in some particular chosen representation  $B = \{B_j\}$ , e. g. when all operators  $B_j$  are out of the boundedness domain of  $\mathcal{R}^{-1}$ . This is the case, for example, of the (overcomplete) representation  $B = \{|\alpha\rangle\langle\beta|\}$ , with  $|\alpha\rangle$  and  $|\beta\rangle$  coherent states, since from the identity

$$\mathcal{N}_\nu(|\alpha\rangle\langle\alpha|) = \frac{1}{\nu + 1} D(\alpha) \left( \frac{\nu}{\nu + 1} \right)^{a^\dagger a} D^\dagger(\alpha), \quad (8.72)$$

one obtains

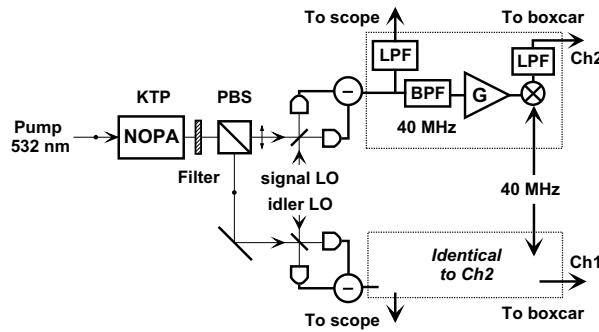
$$\mathcal{N}_\nu^{-1}(|\alpha\rangle\langle\alpha|) = \frac{1}{1 - \nu} D(\alpha) (1 - \nu^{-1})^{-a^\dagger a} D^\dagger(\alpha), \quad (8.73)$$

which has convergence radius  $\nu \leq \frac{1}{2}$ , which is the well known bound for Gaussian noise for the quantum tomographic reconstruction for coherent-state and Fock representations [28]. Therefore, we say that the state is *formally* faithful, however, we are constrained to representations that are analytical for the inverse map  $\mathcal{R}^{-1}$ .

### 8.3 Homodyne tomography of channels and POVM's

Once the information about a device is encoded into quantum states, all the techniques of quantum tomography, which are also reviewed in this set of Lecture Notes, can be applied to determine the channel or, more generally, the quantum operation describing the device. To date, several experiments of quantum process tomography have been implemented for qubits either in NMR systems [5–7] or in quantum optics [8, 9, 14, 15], the latter also deserving an entire chapter in these Lecture Notes. However, no experiments in the realm of continuous variable optical systems have been realized yet. Here, with the help of Monte Carlo simulations, we analyze the feasibility of some experiments in such context, using as a faithful state a twin-beam emerging

from parametric down-conversion of vacuum, and performing a joint homodyne tomography on both the modes of radiation at the output. The actual experimental feasibility of the technique is partly proved by the experiment of Ref. [29], in which quantum homodyne tomography of the (joint number probability distribution of) a twin-beam was achieved using the setup depicted in Fig. 8.2. After a brief introduction on homodyne tomography, we report as an



**Fig. 8.2.** A nondegenerate optical parametric amplifier (a KTP crystal) is pumped by the second harmonic of a Q-switched mode-locked Nd:YAG laser, which produces a 100-MHz train of 120-ps duration pulses at 1064 nm. The orthogonally polarized twin-beams emitted by the KTP crystal are separately detected by two balanced homodyne setups that use two independent local oscillators derived from the same laser. The output of the apparatus is a measure of the quadrature amplitudes  $X_{\phi'} \otimes X_{\phi''}$  for random phases  $\phi'$  and  $\phi''$  with respect to the local oscillators. (From Ref. [29])

example of quantum process tomography the result presented in Ref. [10] for the tomography of a displacement unitary transformation. Then we address the problem of the feasibility of the homodyne tomography of a POVM for an ON/OFF photo-detector. For the tomography of the unitary transformation the tomographic reconstruction will be performed by the method of pattern function averaging. For the tomographic of the photo-detector, on the other hand, we will also consider maximum likelihood methods, to show how they can give a huge boost to the precision of the characterization, at the sole expense of greater computational complexity.

Overall, homodyne tomography of processes and detectors will become a major diagnostic tool in quantum optics, opening new perspectives for the calibration of measuring apparatuses and the characterization of the dynamics of optical devices.

### 8.3.1 Homodyne tomography

A balanced homodyne detector in the strong oscillator limit ideally measures the field quadrature observable

$$X_\phi = \frac{a^\dagger e^{i\phi} + a e^{-i\phi}}{2}, \quad (8.74)$$

where  $a$  and  $a^\dagger$  are the annihilation and the creation operators of the mode of interest (set by the local oscillator), for a chosen value of the phase  $\phi$ . In the Fock basis  $|n\rangle$  the (unnormalizable) eigenstate  $|x\rangle_\phi$  of the quadrature  $X_\phi$  is given by

$$|x\rangle_\phi = \sum_{n=0}^{\infty} \left(\frac{2}{\pi}\right)^{\frac{1}{4}} \frac{1}{\sqrt{2^n n!}} \exp(-x^2) H_n(\sqrt{2}x) e^{inx\phi} |n\rangle, \quad (8.75)$$

$H_n(x)$  denoting Hermite polynomials. Once the phase  $\phi$  is fixed, the ideal measurement realizes the POVM  $\text{Hom}(x; \phi) = |x\rangle_\phi \langle x|$  for the “continuous variable”  $x$ , with a probability density distribution of the outcomes given byproduct

$$p(x; \phi) = \text{tr} [\rho \text{Hom}(x; \phi)], \quad (8.76)$$

$\rho$  being the state of the system. In the non-ideal situation on non-unit quantum efficiency, the POVM, and in turn the probability distribution of outcomes, becomes Gaussian convoluted with variance  $\Delta_\eta^2 = \frac{1-\eta}{4\eta}$ , the parameter  $\eta$  denoting the *quantum efficiency* of photo-detectors used in the homodyne.

Homodyne tomography is a method for estimating the state  $\rho$  from a finite sample of homodyne data, i.e. distributed according to  $p(x; \phi)$  in Eq. (8.76). The easiest strategy estimates the ensemble average of any operator  $O$  by averaging bounded *pattern function*  $\mathcal{P}_\eta[O](x, \phi)$  of homodyne data. This means that one has

$$\langle O \rangle = \text{tr}[\rho O] = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} dx p_\eta(x; \phi) \mathcal{P}_\eta[O](x, \phi), \quad (8.77)$$

and the expectation value is achieved by averaging the pattern function on the homodyne data  $\{(x_n, \phi_n)\}$  in the limit of infinitely many data

$$\frac{1}{N} \sum_{n=0}^N \mathcal{P}_\eta[O](x_n, \phi_n) \xrightarrow{N \rightarrow \infty} \langle O \rangle \quad (\text{with probability 1}). \quad (8.78)$$

By averaging the pattern functions of the form  $\mathcal{P}_\eta[j\rangle\langle i|]$ , the matrix elements  $\langle i|\rho|j\rangle$  of the state of the system are estimated. These pattern functions can be found in the first chapter of this set of Lecture Notes.

Here we are interested in the homodyne tomography of the joint state of two modes of radiation, which can be experimentally separately measured, so

that their quadratures  $X_\phi$  and  $X'_{\phi'}$  are jointly and independently measured, yielding the set of outcomes  $\{(x_n, \phi_n, x'_n, \phi'_n)\}$ . It is easy to show that the pattern function of the tensor product of two operators factorizes, namely

$$\mathcal{P}[O_1 \otimes O_2](x_n, \phi_n, x'_n, \phi'_n) = \mathcal{P}[O_1](x_n, \phi_n) \mathcal{P}[O_2](x'_n, \phi'_n), \quad (8.79)$$

whence the matrix elements of a bipartite state  $R$  can be estimated as

$$\frac{1}{N} \sum_{n=0}^N \mathcal{P}_\eta[j]\langle i](x_n, \phi_n) \mathcal{P}_\eta[m]\langle l](x'_n, \phi'_n) \rightarrow \langle i|\langle l|R|j\rangle|m\rangle. \quad (8.80)$$

Another estimation strategy for homodyne tomography is the maximum likelihood one, in which the “true” state  $\hat{\rho}$  is estimated from homodyne data  $\{(x_n, \phi_n)\}$  as the one which most likely has generated the observed data, namely the one that maximizes the likelihood functional

$$\mathcal{L}[\rho] = \sum_n \ln \text{tr}[\rho \text{Hom}_\eta(x_n; \phi_n)]. \quad (8.81)$$

Obviously, for finite samples the estimated state will differ from the true one, and an estimation of errors (statistical and systematic) is in order.

The maximum likelihood (ML) method is an effective method for solving more generally LININPOS (i.e. positive linear inverse) problems [30], and the present case of state estimation from homodyne data is just an example. Of course, the ML approach extends straightforwardly to the case of a bipartite system. A survey on the use of maximum likelihood methods in quantum mechanics is presented in this set of Lecture Notes.

### 8.3.2 Homodyne tomography of a field displacement

In this first example, the input state  $|\Psi\rangle\rangle = (1 - |\xi|^2)^{\frac{1}{2}} \sum_{n=0}^{\infty} \xi^n |n\rangle|n\rangle$  is generated by parametric downconversion of the vacuum, with  $\xi = [\bar{n}/(\bar{n} + 1)]^{\frac{1}{2}}$ ,  $\bar{n}$  being the average number of photons in each mode. A displacement unitary transformation  $D(z) = \exp(za^\dagger - z^*a)$  is then applied to one of the two beams, thus yielding the output state

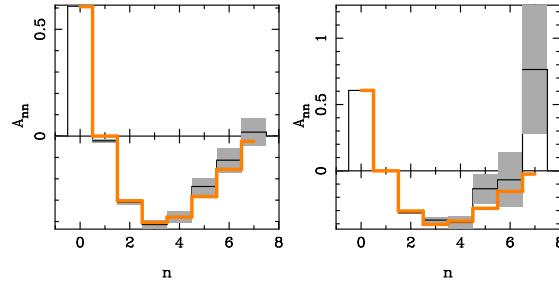
$$R_z = [D(z) \otimes I]|\Psi\rangle\rangle\langle\Psi|[D^\dagger(z) \otimes I] = (1 - |\xi|^2) |D(z) \xi^{a^\dagger a}\rangle\rangle\langle\langle D(z) \xi^{a^\dagger a}|, \quad (8.82)$$

which is then measured with two balanced homodyne setups, one for each mode.

In Fig. 8.3 some results of the Monte Carlo simulation of the proposed experiment are reported. To show how this technique is effective, the matrix elements  $\langle n|\langle n|R_z|0\rangle|0\rangle$  are estimated by pattern function averaging, and then an estimate of diagonal elements of the operator  $D(z)$  is calculated as

$$A_{nn} = \langle n|D(z)|n\rangle = (1 - |\xi|^2)^{-1/2} \xi^{-n} \frac{\langle n|\langle n|R_z|0\rangle|0\rangle}{\sqrt{\langle 0|\langle 0|R_z|0\rangle|0\rangle}}, \quad (8.83)$$

and compared with the theoretical value. As one can see, a meaningful reconstruction of the matrix elements of  $D(z)$  can be achieved in the range  $n = 0 \div 7$  with  $10^6 \div 10^7$  data, with approximately  $\bar{n} = 3$  thermal photons, and with quantum efficiency as low as  $\eta = 0.7$ . These experimental parameters correspond to those of the experiment of Ref. [29]. Improving quantum efficiency and increasing the amplifier gain (toward a maximally entangled state) have both the effect of making statistical errors smaller and more uniform versus the photon labels  $n$  and  $m$  of the matrix  $A_{nm}$ .



**Fig. 8.3.** From Ref. [10]. Homodyne tomography of the displacement of one mode of the radiation field. The estimated diagonal elements  $A_{nn}$  of the displacement operator (shown by a thin solid line on an extended abscissa range, with their respective error bars in a gray shade) are compared to the theoretical values  $\langle n | D(z) | n \rangle$  (thick solid line). Similar results are obtained for the remaining matrix elements. The reconstruction has been achieved using an entangled state  $|\Psi\rangle\langle\Psi|$  at the input corresponding to parametric downconversion of vacuum with mean thermal photon  $\bar{n}$  and quantum efficiency at homodyne detectors  $\eta$ . Left:  $z = 1$ ,  $\bar{n} = 5$ ,  $\eta = 0.9$ , and  $1.5 \times 10^6$  data have been used. Right:  $z = 1$ ,  $\bar{n} = 3$ ,  $\eta = 0.7$ , and  $6 \times 10^7$  data have been used. The last plot corresponds to the same parameters of the experiment in Ref. [29].

In the experiment of Ref. [29], the relative phases between the local oscillators of the two homodyne detectors and the pump of the twin-beam were completely random and uncontrolled, and this allowed measurement of the diagonal matrix elements  $\langle n | \langle m | R | n \rangle | m \rangle$  only of the two mode state  $R$ , since the corresponding pattern functions are the only ones not depending on the phases. This experimental limitation is difficult but not impossible to overcome.

#### *Comments on the maximum-likelihood strategy.*

The reconstruction can be made much more efficient by ML methods [31–37] also reviewed in these Lecture Notes, with a reduction of the needed number of data up to a factor 100–1000. Within our experimental scheme, the action of a generic quantum process  $\mathcal{E}$  on one mode of the twin-beam generates the

output state  $R_{\mathcal{E}} = (I \otimes \Psi^T) S_{\mathcal{E}} (I \otimes \Psi^*)$  (cfr. Eq. (8.36)),  $S_{\mathcal{E}}$  being the operator corresponding to the quantum process under analysis, which is positive and satisfies  $\text{tr}_1 S_{\mathcal{E}} = I$ . The probability distribution of the result  $(x, \phi, x', \phi')$  of a double homodyne detection on the two modes becomes

$$\begin{aligned} \Pr(x, \phi, x', \phi'; S_{\mathcal{E}}) &= \text{tr} [\text{Hom}_{\eta}(x; \phi) \otimes \text{Hom}_{\eta}(x'; \phi') R_{\mathcal{E}}] , \\ &= \text{tr} [\text{Hom}_{\eta}(x; \phi) \otimes (\Psi^* \text{Hom}_{\eta}(x'; \phi') \Psi^T) S_{\mathcal{E}}] . \end{aligned} \quad (8.84)$$

Given a set of double homodyne data  $\{(x_n, \phi_n, x'_n, \phi'_n)\}$ , the investigated quantum process can be estimated as the one whose corresponding operator  $S_{\hat{\mathcal{E}}}$  maximizes the likelihood functional

$$\mathcal{L}[S_{\mathcal{E}}] = \sum_{n=0}^N \ln [\Pr(x_n, \phi_n, x'_n, \phi'_n; S_{\mathcal{E}})] , \quad (8.85)$$

within the simplex defined by the constraints  $S_{\mathcal{E}} \geq 0$  and  $\text{tr}_1 S_{\mathcal{E}} = I$ . If some prior knowledge about the process is available (for example, one could already know that the device performs a unitary transformation) then the maximization can be further restricted to a smaller set of candidates, thus improving further the efficiency of the estimation. In contrast to what happens with pattern averaging, here, by construction, the estimated map is automatically CP and trace preserving, and can fulfill any desired additional requirement.

As regards statistical efficiency, for the ML estimator we can assert that it is in some sense the most efficient with the following reasoning. Given a generic family of probability distributions  $\Pr(x; \boldsymbol{\theta})$  depending on the independent, unconstrained parameters  $\boldsymbol{\theta} \in \mathbb{R}^d$ , one defines the Fisher information matrix as

$$F(\boldsymbol{\theta})_{mn} = \left\langle \frac{\partial \ln \Pr(x; \boldsymbol{\theta})}{\partial \theta_m} \frac{\partial \ln \Pr(x; \boldsymbol{\theta})}{\partial \theta_n} \right\rangle_x \quad (8.86)$$

and for any unbiased estimator  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$ , defined on samples of  $N$  data drawn from  $\Pr(x; \boldsymbol{\theta})$ , the covariance matrix

$$\Sigma_{mn} = \left\langle (\hat{\theta}_m - \theta_m)(\hat{\theta}_n - \theta_n) \right\rangle_{x_1 \dots x_N} . \quad (8.87)$$

The two matrices satisfy the Cramer-Rao bound

$$\Sigma F(\boldsymbol{\theta}) \geq \frac{1}{N}, \quad (8.88)$$

which puts a limit on the efficiency of the estimation that is independent of the estimator. It is possible to prove that if there exists an estimator that achieves the bound, then it coincides with the ML estimator, and that ML saturates the bound asymptotically, for increasing sample size  $N$ , when the ML estimator becomes approximately Gaussian distributed around  $\boldsymbol{\theta}$ , with a covariance matrix given by the so called CR matrix  $F^{-1}(\boldsymbol{\theta})/N$ .

When the parameters  $\boldsymbol{\theta} \in \mathbb{R}^d$  are constrained to a subset  $\Theta \subset \mathbb{R}^d$ , the problem should be reparametrized, at least in a neighborhood of the true value  $\boldsymbol{\theta}$ , and the new set of independent unconstrained parameters should be then used to calculate a new Fisher information and the related CR matrix. However, this procedure is in general inelegant and difficult to use. In Ref. [38], a much more convenient way to compute the constrained CR bound was presented, based on the distinction between regular points of  $\Theta$  (i.e. the points in the closure of the set of interior points of  $\Theta$ ) and non-regular points. As an example, for  $\Theta$  defined by the constraints  $0 \leq \theta_i \leq 1$ , all the points are regular, whereas for  $\Theta$  equal to a lower dimensional manifold embedded in  $\mathbb{R}^d$  (e.g. a surface defined by some equality constraints) all points are nonregular. The result is that if  $\boldsymbol{\theta}$  is a regular point, then the CR matrix is unaltered, whereas if  $\boldsymbol{\theta}$  is not a regular point then the CR matrix must be corrected by subtracting a positive matrix depending on  $\boldsymbol{\theta}$  that makes the CR matrix smaller and singular. The singularity of the CR matrix reflects the fact that some parameters could be actually evaluated as functions of others, and thus do not have an independent associated error. A very simple derivation of equality constrained CR bound can be found in Ref. [39], along with a proof that, for constrained problems, if the bound is achieved by an estimator, then the estimate is a stationary point for the problem of maximizing the likelihood function subject to the constraints. For the problem of  $k < d$  equality constraints  $f_j(\boldsymbol{\theta}) = 0$ , the corrected constrained CR bound becomes

$$\Sigma \geq \frac{1}{N} [F^{-1} - F^{-1}G(G^T F^{-1}G)^{-1}G^T F^{-1}] , \quad (8.89)$$

where  $G$  denotes the  $d \times k$  matrix of the gradient of the constraints  $G_{ij} = \frac{\partial f_j(\boldsymbol{\theta})}{\partial \theta_i}$ .

For the general problem of quantum process tomography, the likelihood functional  $\mathcal{L}[S_\varepsilon]$  of Eq. (8.85) is defined for a parameter  $S_\varepsilon$  living in an infinite dimensional Hilbert space. The maximum of the likelihood is not achieved over the whole space and it is more appropriate to restrict the attention to a subspace  $Q(N)$  on which the maximum exists, and to let its dimension grow with the number  $N$  of data, so to cover the whole parameter space in the limit of an infinite size sample. This method—called *sieved maximum likelihood*—has been analyzed in Ref. [40] for homodyne tomography of a quantum state, with the sieves as the span of the Fock states  $|0\rangle \dots |d(N)\rangle$ , and the function  $d(N)$  chosen in order to guarantee the consistency of the estimator, i.e. the convergence of the estimated state to the true value in the limit of infinite  $N$ .

For the particular problem at hand, because of the exponentially decreasing twin-beam components on the Fock basis, the choice of a suitable cut-off dimension will not introduce any significant bias in the estimation, and the action of the quantum channel will be reconstructed only on a finite dimensional subspace, consistently with the fact that the faithfulness of the input state rapidly vanishes for larger photon numbers.

The only downside of the ML approach is the difficulty involved in the maximization of the nonlinear functional in Eq. (8.85), which can be tackled either with standard techniques of numerical constrained maximization or with suitable modifications [41] of the iterative algorithms of the kind *expectation-maximization* (EM) for maximum likelihood [30, 42]. In practice, several technical problems may arise, as we will discuss concretely in the following example.

### 8.3.3 Homodyne tomography of an On/Off detector

In what follows we exploit the ideas of Sec. (8.2.6) for realizing the tomography of the POVM of a measuring apparatus. One of the beams in the twin-beam state  $|\Psi\rangle\langle\Psi|$  generated by parametric down-conversion of the vacuum (same setup as before) is now measured by an ON/OFF photo-detector. This is described by a two-value POVM, with elements  $\Pi^{(0)}$  and  $\Pi^{(1)} = I - \Pi^{(0)}$ . As discussed in Sec. (8.2.6), looking at Eq. (8.63), the reduced states of the remaining beam after the measurement will be

$$\rho^{(i)} = \frac{\Psi^T \Pi^{(i)T} \Psi^*}{\text{tr}[\Psi^T \Pi^{(i)T} \Psi^*]}, \quad (8.90)$$

$i$  being the measurement outcome, with the denominator of the previous expression giving its probability. Homodyne tomography is then performed on each reduced state in order to recover the POVM elements.

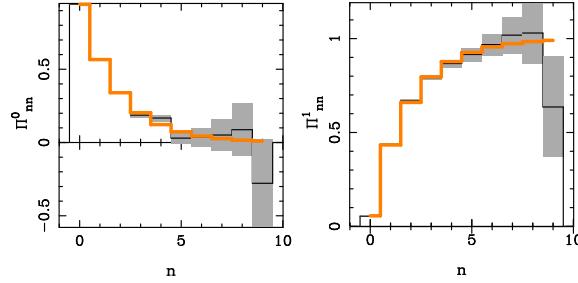
As a model of ON/OFF detector with non unit quantum efficiency and dark current, we will use an ideal ON/OFF photodetector preceeded by a beam-splitter of transmissivity  $\tau$  with one port entered by the mode of interest and with the other port fed by a thermal radiation state with mean photon number  $\mu$  [43]. The POVM element for the OFF outcome reads

$$\Pi^{(0)} = \frac{1}{\nu + 1} \sum_{n=0}^{\infty} \left(1 - \frac{\tau}{\nu + 1}\right)^n |n\rangle\langle n|, \quad (8.91)$$

$\nu = \mu(1 - \tau)$  being the resulting mean photon number of the background noise, whereas the POVM element for the ON outcome is  $\Pi^{(1)} = I - \Pi^{(0)}$ .

*Reconstruction using pattern-function averaging strategy.*

The graphs in Fig. 8.4 show that a meaningful reconstruction can be obtained within the same range of values for the parameters used in the tomography of the displacement. As usual, in order to achieve the reconstruction of the off-diagonal terms of the POVM, the phase-control for the local oscillator of the balanced homodyne detector relative to the pump of the down-converter is required. The presence of non-vanishing off-diagonal terms in the POVM would allow the detector to reveal some form of coherence in the input state,



**Fig. 8.4.** Homodyne tomography of an On/Off photo-detector having transmissivity  $\tau = 0.4$  and number of thermal noise photons  $\nu = 0.1$ . Only the diagonal matrix elements of the POVM elements  $\Pi^{(0)}$  and  $\Pi^{(1)}$  are reported (the off-diagonal ones are zero, and have similar error bars). The reconstruction is obtained by pattern-function averaging of  $1.5 \cdot 10^6$  data, for quantum efficiency  $\eta = 0.9$  and  $\bar{n} = 3$ , and presents error bars of the same magnitude as the ones for the displacement reconstruction reported in Fig. 8.3

and in our model it could be simulated by having some coherence for the thermal radiation injected in the beam-splitter. Of course, if one already knows that the detector is perfectly phase-insensitive (as for a customary photo-detector, for its intrinsic detection mechanism), one can focus the attention only on the diagonal part of the state, without the need of phase-control for the local oscillators.

It is important to notice that when the only diagonal part of the POVM of the measuring apparatus is under examination, it is not necessary to have the input state  $R$  faithful, but it is sufficient to have the matrix  $R_{mn} = \langle m| \langle n|R|m\rangle |n\rangle$  invertible, with more easily experimentally available input states. In fact, in correspondence with the measurement outcome  $i$ , the diagonal matrix elements of reduced state  $\rho^{(i)}$  of the auxiliary system are given by

$$\rho_{nn}^{(i)} = \frac{\sum_m R_{mn} \Pi_{mm}^{(i)}}{\text{tr}[\sum_{mn} R_{mn} \Pi_{mm}^{(i)}]}, \quad (8.92)$$

so that, once measured  $\rho_{nn}$ , it is possible to recover  $\Pi_{mm}^{(i)}$  given that  $R_{mn}$  is invertible. In summary, a “diagonally faithful” state and homodyne tomography (i.e. without phase control) is enough for the reconstruction of a diagonal POVM.

#### *Reconstruction using maximum-likelihood strategy.*

Now, we will analyze the data from the same experimental scheme using the maximum likelihood strategy, assuming, for simplicity, that we already know the POVM is diagonal in the Fock basis for its intrinsic detection mechanisms, such that a bipartite diagonally faithful state  $R$  and homodyne to-

mography without phase-control will suffice for the purpose of reconstructing the POVM.

Non-ideal homodyne detection can be modeled as the action of the loss map followed by ideal homodyne detection, with a suitable rescaling of outcomes, such that the POVM can be written as follows

$$\text{Hom}_\eta(x, \phi) = \sqrt{\eta} \sum_{j=0}^{\infty} V_j^\dagger e^{i\phi a^\dagger a} |\sqrt{\eta}x\rangle\langle\sqrt{\eta}x| e^{-i\phi a^\dagger a} V_j , \quad (8.93)$$

where  $V_j = (\eta^{-1} - 1)^{\frac{j}{2}} a^j \eta^{\frac{1}{2}a^\dagger a} / \sqrt{j!}$  are the elements of the Kraus decomposition of the loss map, and  $\eta$  denotes the quantum efficiency of the detectors (this scheme is equivalent to having an ideal homodyne detector preceded by a beam-splitter with transmissivity  $\eta$  and its second port fed with the vacuum state). If the phase is out of control and uniformly random, then the POVM corresponding to the measurement is the average over the phase of Eq. (8.93), which yields a diagonal POVM  $\text{Hom}_\eta(x)$  (this also makes it clear why without phase control it is impossible to reconstruct the off-diagonal matrix elements).

The probability of getting the outcome  $(i, x)$  with the photo-counter signaling outcome  $i$  and the homodyne measuring  $x$ , is given by

$$\begin{aligned} \Pr(i, x; \Pi) &= \text{tr}[R \Pi^{(i)} \otimes \text{Hom}_\eta(x)] = \sum_{mn} R_{mn} \Pi_{mm}^{(i)} \text{Hom}_{nnn}(x) = \\ &= \sum_m \Pi_{mm}^{(i)} A_m(x) , \end{aligned} \quad (8.94)$$

where  $R$  denotes twin-beam state, and the positive coefficients  $A_m(x)$  are defined as

$$A_m(x) = \sqrt{\eta} \sum_{n \geq h} R_{mn} \binom{n}{h} \eta^h (1-\eta)^{n-h} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{e^{-2\eta x^2}}{2^h h!} H_h^2(\sqrt{2\eta}x) > 0 . \quad (8.95)$$

For a given set of experimental data  $\{(i_l, x_l)\}$ , the maximum likelihood estimate  $\hat{\Pi}$  is the one maximizing the functional

$$\mathcal{L}[\Pi] = \sum_l \ln \left[ \sum_m \Pi_{mm}^{(i_l)} \cdot A_m(x_l) \right] , \quad (8.96)$$

with  $\Pi$  restricted to the simplex of diagonal POVM's.

First, one must choose the dimension of the subspace on which one is performing the maximization of the likelihood and thus the estimation of the POVM elements. In such a finite dimensional subspace, the ML estimate is well defined, being the point attaining the unique maximum of a convex functional restricted to a simplex. In principle this restriction introduces a bias in the estimation, however, in our case the exponentially decreasing

components of the twin-beam state in the Fock basis in practice makes the bias negligible, by making the components  $\langle n | \rho^{(i)} | n \rangle$  of the reduced state after the measurement rapidly vanishing for large  $n$ .

The maximization of the functional  $\mathcal{L}[\Pi]$  is a nonlinear convex programming problem, and can be faced with several different kind of algorithms as the *simplex method* (see contribution of D'Ariano *et al.* to these Lecture Notes), or the methods of *sequential quadratic programming* (SQP), or the methods of *expectation-maximization* (EM) type (see chapter by Hradil *et al.* in these Lecture Notes) that can be easily implemented in this particular example since, for fixed  $m$  and varying  $i$ , the numbers  $\Pi_m^{(i)}$  define a probability distribution. For all methods convergence is assured, since the functional to be maximized is convex and differentiable over the simplex of diagonal POVM's. However, when applying any of these methods, the convergence speed and the reliability of the result at a given iteration step are two major concerns. In fact, the derivatives of  $\mathcal{L}[\Pi]$  with respect to some of the parameters  $\theta$  defining  $\Pi$  can be very small, so that very different values of the parameters will give almost the same likelihood, thus making it hard to judge whether the point reached at a given iteration step is a good approximation of the point corresponding to the maximum: in few words, the problem becomes numerically ill conditioned, with an extremely low convergence rate.

Notice that the Fisher information matrix (Eq. (8.86)) for the probability distribution  $\Pr(i, x; \Pi)$  can be expressed in terms of the expectation value of the derivatives of the likelihood with respect to the independent parameters  $\theta_m$  defining  $\Pi$

$$F(\Pi)_{mn} = \frac{1}{N} \left\langle \frac{\partial \mathcal{L}[\Pi]}{\partial \theta_m} \frac{\partial \mathcal{L}[\Pi]}{\partial \theta_n} \right\rangle_{(i_1, x_1) \dots (i_N, x_N)}, \quad (8.97)$$

so that the derivatives of the likelihood not only affect the numerical stability of the maximization, but also limit the theoretical precision of the estimation via Cramer-Rao lower bound. This bound, in turn, can be used to check whether or not the estimation is good, depending on how much the variance of the estimator is bigger than the lower bound of Eq. (8.88). This, however, needs the calculation of the Fisher information matrix in correspondence of the unknown true value of  $\Pi$ , and this can be approximated by the Fisher information at the estimated value, which is a reasonably good approximation provided the estimated value doesn't deviate too much from the true one.

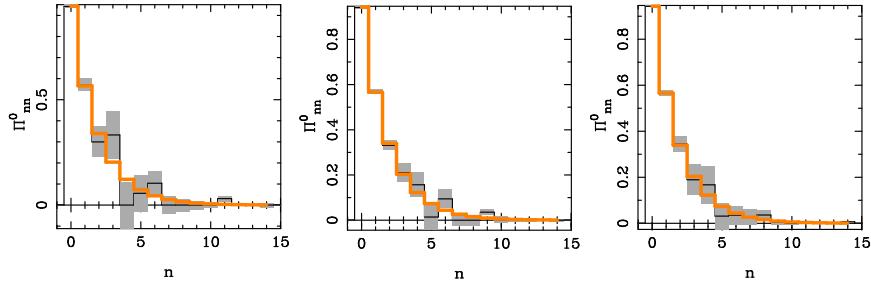
As already mentioned, in the limit of large size samples the ML estimator is Gaussian distributed around the true value with a covariance matrix equal to  $(NF)^{-1}$ . Therefore for large samples the confidence levels can be assumed to be Gaussian, with variances calculated from the Fisher information, which can be evaluated on the estimated parameter for not too large errors. However, this is an asymptotic property, so that for finite size samples sometimes there is the problem of establishing the errors and the confidence levels for the estimation. When working with Monte Carlo simulation, the virtual ho-

modyne experiment can be repeated several times, in order to evaluate the distribution of the ML estimator, and thus its confidence levels. Clearly, this approach is not satisfying for an experimentalist, who would need to collect a lot of data only to evaluate the statistical errors for a small subset of them. A valid alternative is then provided by the method of bootstrap [44], which is based on the simple idea that when some data are drawn from an unknown probability distribution, then the distribution of those data is the best approximation we have of the real probability distribution. Thus, once the experimental data are collected, we can perform the ML estimation on artificial samples repeatedly generated by random sampling of the experimental data: the distribution of such estimates approximates well the one that we would have from the real experiment, and can be used to evaluate the confidence levels for the estimator.

Back to our problem of On/Off detector tomography, we have produced a Monte Carlo simulation of the joint homodyne and on/off data distributed according to Eq. (8.94), for the POVM model presented in Eq. (8.91), with the same parameters as Fig. 8.4, and various values of the quantum efficiency  $\eta$ . The detector POVM has been estimated with the maximum likelihood method, with the only hypothesis of diagonal POVM, and putting the dimensional cut-off at the first 15 elements of the Fock basis (for a number of photons in the twin-beam equal to  $\bar{n} = 3$  this introduces almost no bias, with an actual suppression of a factor 100 between the first diagonal element of the POVM and the first excluded element). The results are reported in Fig. 8.5 for different sample sizes and quantum efficiencies, where the only “Off” element of the POVM is reported, since the “On” element is simply its complement with respect to the identity.

A direct comparison with Fig. 8.4 evidences the much higher efficiency of maximum likelihood reconstruction. The graph on the left shows how the same magnitude of errors is achieved on a larger subspace with less than one tenth of the data ( $10^5$  vs.  $10^6$ ) and with a much lower quantum efficiency (0.7 vs. 0.9). For the same quantum efficiency  $\eta = 0.9$ , here the results are much better even with as few as 1% of the data (graph in the middle), analogously, for the same amount of data ( $N = 10^6$ ), here the results are much better even for a quantum efficiency as low as 0.7 (graph on the right).

The distribution of the estimator in each bin, which is necessary for giving proper confidence levels for the result, has been evaluated by repeated Monte Carlo experiments, however, which is equivalent to the bootstrapping techniques for truly experimental data. As a result the estimator in each bin is not Gaussian distributed, a sign of the fact that the number of data used is not enough to reach the asymptotic Gaussian distribution of the ML estimator. In the plot, the only variances are reported for each bin, showing that the errors are distributed with respect to  $n$  differently than for pattern averaging.



**Fig. 8.5.** Homodyne tomography of an On/Off photodetector with transmittivity  $\tau = 0.4$  and number of thermal noise photons  $\nu = 0.1$ , with  $\bar{n} = 3$  photons in the twin-beam. The ML estimation of the diagonal of the only Off POVM element are reported for different values of sample size  $N$  and quantum efficiency  $\eta$ . Left:  $N = 10^5$ ,  $\eta = 0.7$ ; Middle:  $N = 10^4$ ,  $\eta = 0.9$ ; Right:  $N = 10^6$ ,  $\eta = 0.7$ .

The maximization has been performed numerically by the routine `donlp2` [45] that implements an SQP algorithm, and then the self-consistency of the solution has been checked by means of a few EM type iterations applied alternatively to the probability distributions corresponding to the elements  $\Pi_m^{(i)}$ , for fixed  $m$ . Of course, it would be much easier to employ the only EM algorithm, being a recursive application of an easily implementable iteration. However, this algorithm has an extremely low convergence speed, which also could make the iteration stop too early, leading to (statistically wrong!) results—which may even fit well the theoretical POVM. This paradoxical behavior is an artifact due to the smoothness of the theoretical curve of the considered model, as one can easily check by changing the model.

## Acknowledgements

This work has been sponsored by INFM through the project PRA-2002-CLON, and by EEC and MIUR through the cosponsored ATESIT project IST-2000-29681 and Cofinanziamento 2002.

## References

1. W. F. Stinespring: Proc. Am. Math. Soc. **6**, 211 (1955).
2. K. Kraus: Ann. Phys. **64**, 311 (1971).
3. I. L. Chuang, M. A. Nielsen: J. Mod. Opt. **44**, 2455 (1997).
4. J. F. Poyatos, J. I. Cirac, P. Zoller: Phys. Rev. Lett. **78**, 390 (1997).
5. M. A. Nielsen, E. Knill, R. Laflamme: Nature **396**, 52 (1998).
6. A. M. Childs, I. L. Chuang, D. W. Leung: Phys. Rev. A, 012314 (2001).

7. N. Boulant, T. F. Havel, M. A. Pravia, D. G. Cory: Phys. Rev. A **67**, 042322 (2003).
8. P. Kwiat, J. Altepeter, D. Branning, E. Jeffrey, N. Peters, T. C. Wei: Taming entanglement. In: *Proceedings of the 6th International Conference on Quantum Communications, Measurement and Computing*, ed by J. Shapiro, J. and O. Hirota (Rinton Press, Princeton, NJ, 2003).
9. M. W. Mitchell, C. W. Ellenor, S. Schneider, A. M. Steinberg: Phys. Rev. Lett. **91**, 120402 (2003).
10. G. M. D'Ariano, P. Lo Presti: Phys. Rev. Lett. **86**, 4195 (2001).
11. D. Leung: Towards robust quantum computation, PhD thesis, Stanford University (2001).
12. A. Jamiołkowski: Rep. Math. Phys. **3**, 275 (1972).
13. M.-D. Choi: Linear Algebra Appl. **10**, 285 (1975).
14. F. De Martini, A. Mazzei, M. Ricci, G. M. D'Ariano: Phys. Rev. A **67**, 062307 (2003).
15. J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, A. G. White: Phys. Rev. Lett. **90**, 193601 (2003).
16. A. G. White, A. Gilchrist, G. J. Pryde, J. L. O'Brien, M. J. Bremner, N. K. Langford, (2003): arXiv:quant-ph/0308115 (2003).
17. G. M. D'Ariano, P. Lo Presti: Phys. Rev. Lett. **91**, 047902 (2003).
18. G. M. D'Ariano, P. Lo Presti, M. G. A. Paris: Phys. Rev. Lett. **87**, 270404 (2001).
19. K. Kraus: *States, Effects, and Operations* (Springer, Berlin, 1983).
20. A. S. Holevo: *Probabilistic and statistical aspects of quantum theory*, vol 1 of *Series in Statistics and Probability* (North-Holland, Amsterdam, New York, Oxford, 1982).
21. G. M. D'Ariano, P. Lo Presti: unpublished (2003).
22. I. L. Chuang, M. A. Nielsen: *Quantum Information and Quantum Computation* (Cambridge University Press, Cambridge UK, 2000).
23. A. Luis: Phys. Rev. A **62**, 054302 (2000).
24. R. B. Bapat: *Linear algebra and linear models* (Springer, Berlin, 2000).
25. M. J. W. Hall: (1994) Phys. Rev. A **50**, 3295 (1994).
26. R. Simon: Phys. Rev. Lett. **84**, 2726 (2000).
27. M. G. A. Paris: J. Opt. B: Quantum Semiclass. Opt. **4**, 442 (2002).
28. G. M. D'Ariano, N. Sterpi: J. Mod. Opt. **44**, 2227 (1997).
29. M. Vasilyev, S.-K. Choi, P. Kumar, G. M. D'Ariano: Phys. Rev. Lett. **84**, 2354 (2000).
30. Y. Vardi, D. Lee: J. R. Statist. Soc. B **55**, 569 (1993).
31. Z. Hradil: Phys. Rev. A **55**, R1561 (1997).
32. Z. Hradil, J. Summhammer, G. Badurek, H. Rauch: Phys. Rev. A **62**, 014101 (2000).
33. J. Fiurášek, Z. Hradil: Phys. Rev. A **63**, 020101 (2001).
34. K. Banaszek, G. M. D'Ariano, M. G. A. Paris, M. F. Sacchi: Phys. Rev. A **61**, R010304 (2000).
35. M. G. A. Paris, G. M. D'Ariano, M. F. Sacchi: Maximum-likelihood method in quantum estimation. In: *Bayesian inference and maximum entropy methods in science and engineering*, vol 568 of *AIP Conf. Proc.* p 456; preprint arXiv:quant-ph/0101071 (2001).
36. M. F. Sacchi: Phys. Rev. A **64**, 022106 (2001).

37. M. F. Sacchi: Phys. Rev. A **63**, 054104 (2001).
38. J. D. Gorman, A. O. Hero: IEEE Trans. Inf. Th. **26**, 1285 (1990).
39. T. L. Marzetta: IEEE Trans. on Signal Processing **41**, 2247 (1993).
40. R. Gill, M. I. Gută: arXiv:quant-ph/0303020 (2003).
41. J. Rehacek, Z. Hradil, M. Jezek: Phys. Rev. A **63**, R040303 (2001).
42. A. P. Dempster, N. M. Laird, D. B. Rubin, J. R. Statist. Soc. B **39**, 1 (1977).
43. M. G. A. Paris: Phys. Lett. A **489**, 167 (2001).
44. B. Efron, R. Tibshirani: *An Introduction to the Bootstrap* (Chapman and Hall, New York, 1994).
45. P. Spellucci: Math. Prog. **82**, 413 (1998).

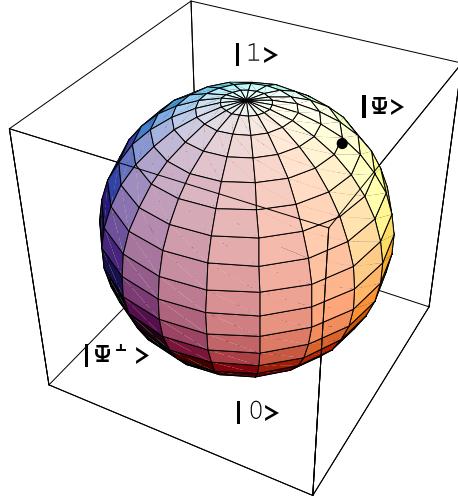
# 9 Quantum Operations on Qubits and their Characterization

Francesco De Martini, Marco Ricci and Fabio Sciarrino

Dipartimento di Fisica and Istituto Nazionale per la Fisica della Materia,  
Universita' "La Sapienza", Roma, 00185 - Italy

## 9.1 Introduction

Classical information is represented by *bits* which can be either 0 or 1. Quantum information is represented by *quantum-bits*, or *qubits*, which are two-dimensional quantum systems. A qubit, unlike a classical bit, can exist in a state  $|\Psi\rangle$  that is a superposition of two orthogonal basis states  $\{|\uparrow\rangle, |\downarrow\rangle\}$ , i.e.  $|\Psi\rangle = \tilde{\alpha}|\uparrow\rangle + \tilde{\beta}|\downarrow\rangle$ . The fact that qubits can exist in these superposition states gives to quantum information unusual properties. Specifically, information encoded in a quantum system has to obey rules of quantum physics that impose strict bounds on possible manipulations with quantum information. The common denominator of these bounds is that all quantum-mechanical transformations have to be represented by *Completely Positive (CP) maps* [1] that in turn impose a constraint on the fidelity of quantum-mechanical measurements. That is, an unknown state of a qubit cannot be precisely determined (or reconstructed) from a measurement performed on a finite ensemble of identically prepared qubits [2–4]. In particular, the mean fidelity of the best possible (optimal) state estimation strategy based on the measurement of  $N$  identically prepared qubits is  $F = (N+1)/(N+2)$ . One of the obvious consequences of this bound on the fidelity of estimation is that unknown states of quantum systems cannot be cloned, viz. copied perfectly [5], namely the perfect cloning map of the form  $|\Psi\rangle \rightarrow |\Psi\rangle|\Psi\rangle$  is not permitted by the rules of quantum mechanics. Certainly if this would be possible, then one would be able to violate the bound on the fidelity of estimation. Moreover, this possibility would trigger more dramatic changes in the present picture of the physical world, e.g. it would be possible to utilize quantum nonlocality for superluminal signaling [6–8]. Another map that cannot be performed perfectly on an *unknown* state of a qubit is the *spin-flip* or the universal-NOT, i.e. the operation  $|\Psi\rangle \rightarrow |\Psi^\perp\rangle$ , where the state  $|\Psi^\perp\rangle$  is orthogonal to the original  $|\Psi\rangle$  [3,9]. Spin-flipping is indeed an *anti-unitary* operation, i.e. a time reversal map,  $T = i\sigma_y K$ , realizing for any input qubit the *inversion* over the Bloch Sphere, as shown in Fig. 9.1. Precisely, the *phase-conjugation* operator  $K$  is responsible for anti-unitarity since  $K|\Psi\rangle = |\Psi^*\rangle$ , the complex conjugate of  $|\Psi\rangle$ . In spite of the fact that some quantum-mechanical transformations on unknown states of qubits cannot be performed perfectly one still may ask that within the given structure of quantum theory, i.e. which is *linear* and



**Fig. 9.1.** Bloch sphere, the state space of a quantum-bit (*qubit*). *Pure-states* are represented by points on the sphere while *statistical-mixtures* are points inside the sphere. The universal-NOT gate operation corresponds to the inversion of the sphere, i.e. states  $|\Psi\rangle$  and  $|\Psi^\perp\rangle$  are antipodes.

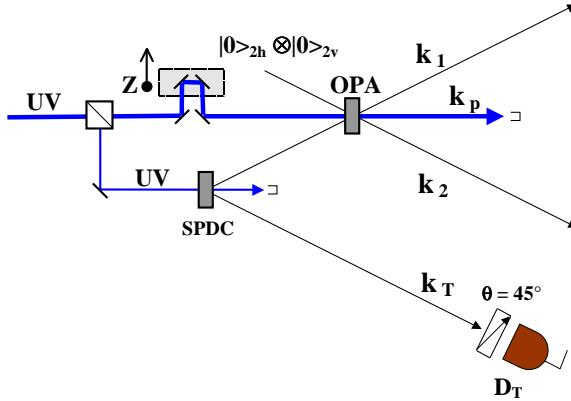
where all maps are *CP maps* [1, 10], what are the best possible approximations of these maps. To be more specific, for the sake of simplicity and with no lack of generality, these maps will be investigated, over the *bipartite entangled* Hilbert space of two-qubits, i.e.  $H_1 \otimes H_2$  with  $\dim(H_i) = 2$ ,  $i = 1, 2$ . Precisely, within this significant framework we shall investigate in a somewhat unifying manner the *optimal* universal cloning and the *optimal* Universal-NOT (U-NOT) gate. For both fundamental processes, considered here as paradigmatic ones, the *universality* condition is required to ensure that all *unknown* input states, i.e. all points on the Bloch sphere, are transformed with the same *quantum efficiency* viz. *fidelity*. Investigation of these universal optimal transformations, which are also called *universal quantum machines* [11] is extremely important since it reveals bounds on optimal manipulations with quantum information. Consequently, in recent years theoretical investigation of the Universal Optimal Quantum Cloning Machine (UOQCM), and of the U-NOT gate has been very thorough. In spite of all the successes in the theoretical analysis of the bounds on optimal manipulations with individual qubits, it is extremely difficult to realize experimentally universal quantum machines. In the domain of quantum optics this is possible by associating a cloning machine with a photon amplification process, e.g. involving inverted atoms in a laser amplifier or a nonlinear medium in a *Quantum-Injected Optical Parametric Amplifier* (QIOPA) [12]. This can be done by virtue of the existing isomorphism between any logic state of a qubit and the polarization state of the photon qubit. In the case of the mode-nondegenerate

QIOPA [12] it is generally supposed that  $N$  photons, prepared identically in an arbitrary quantum state (qubit)  $|\Psi\rangle$  of *polarization* ( $\vec{\pi}$ ) are injected into the amplifier on the input mode  $k_1$ . The amplifier then generates on the same output *cloning mode*  $M > N$  copies, or *clones* of the input qubit  $|\Psi\rangle$ . Correspondingly, the optical parametric amplifier (OPA) generates on the output *anticloning mode*,  $M - N$  states  $|\Psi^\perp\rangle$ , thus realizing a quantum NOT gate. The *nonlinear* QIOPA apparatus will be found to realize the most general cloning transformation, i.e. involving *any* value of  $N$  and  $M$ . However, it is possible to show that, in the simple case  $N = 1$  to  $M = 2$ , the cloning and the U-NOT processes are indeed related by a very simple, *linear* Quantum State Teleportation (QST) transformation [13]. This enlightening result will be investigated both theoretically and experimentally.

This chapter is organized as follows. In Section 9.2 the general theory of quantum injected amplification is reviewed with emphasis on the dynamical conditions apt to ensure, in the present context, the *universality* property of the device, i.e. implying equal quantum efficiencies for *any* arbitrary input  $\vec{\pi}$ -state. Furthermore, the multiparticle superposition state (or *Schrödinger cat* state) of the amplified field on cloning and antycloning modes will be investigated on the basis of the Wigner function of the state. Section 9.3 will show the experimental realization by a QIOPA scheme of a multiphoton, all optical *Schrödinger cat* scheme as a first, simplest application of theory elaborated in Section 9.2. In Section 9.4 the theory of the *optimal* cloning process and NOT gate will be outlined in a unitary and consistent fashion for the  $N = 1, M = 2$  case and applied to the QIOPA scheme operating in a  $\vec{\pi} - entangled$  configuration. Moreover the experimental realization of the corresponding 1 to 2 UOQCM and of the U-NOT gate will be reported. In Section 9.5 a theoretical discussion over the inner connections existing between the two basic quantum information processes – cloning and the U-NOT operation – is presented. We discuss *contextuality* of these two distinct (approximate) unitary transformations realized by the *same* overall dynamical process. These connections are not understood well yet in modern quantum information theory. In order to enlighten the preceding arguments, Section 9.6 will approach the 1 to 2 cloning and the U-NOT case by a novel, nonlocal QST protocol based on linear optics. Precisely, by means of a balanced beam splitter this protocol accomplishes the projection of the overall state onto the symmetric subspace of two qubits, which is exactly what is required for optimal cloning. Furthermore, as significant demonstration of the novel perspectives offered by quantum entanglement to modern measurement theory, Section 9.7 accounts for a new method of tomography of quantum operations, here dubbed as *Pauli tomography*. Precisely, the total parallelism of a bipartite entangled state associated with a pair of correlated photons produced by the spontaneous parametric down conversion (SPDC) is adopted to extract easily and efficiently the full information that characterizes any unknown *quantum operator*, i.e. active device.

## 9.2 Quantum-Injected Optical Parametric Amplification

Here we assume that the quantum information carriers (*qubits*) are represented by states of polarization ( $\vec{\pi}$ ), and quantum efficiency is expressed by the OPA parametric gain:  $g$ . Let us investigate theoretically the dynamics of the QIOPA apparatus making reference to Fig. 9.2 and to Ref. [12].



**Fig. 9.2.** Schematic diagram of the *quantum-injected* optical parametric amplifier in *entangled configuration*. The injection is provided by an external spontaneous parametric down conversion source of polarization ( $\vec{\pi}$ ) entangled photon states.

The active element of the device is a Type II nonlinear crystal slab operating in non-collinear configuration. In these conditions the overall amplification taking place over the coupled electromagnetic modes  $k_j$  ( $j = 1, 2$ ) is contributed to by two equal and *independent* parametric amplifiers: OPA  $A$  and OPA  $A'$  inducing uncorrelated unitary transformations respectively on two couples of (time)  $t$ -dependent field operators  $\hat{a}_1(t) \equiv \hat{a}_{1h}(t), \hat{a}_2(t) \equiv \hat{a}_{2v}(t)$  and  $\hat{a}'_1(t) \equiv \hat{a}_{1v}(t), \hat{a}'_2(t) \equiv \hat{a}_{2h}(t)$  acting on the output modes  $j = 1, 2$  along the horizontal (H) and vertical (V) directions in the  $\vec{\pi}$  plane. The interaction Hamiltonian may be expressed in the general form:

$$\hat{H}_I = i\hbar\chi[\hat{A} - e^{i\Phi'}\hat{A}'] + h.c. \quad (9.1)$$

where:  $\hat{A} \equiv \hat{a}_1(t)\hat{a}_2(t)$ ,  $\hat{A}' \equiv \hat{a}'_1(t)\hat{a}'_2(t)$ , and  $g \equiv \chi t$  is a real number expressing the amplification gain proportional to the nonlinear coupling term  $\chi$ . The dynamics of OPA  $A$  and OPA  $A'$  is expressed correspondingly by the mutually commuting unitary squeeze operators:  $\hat{U}_A(t) = \exp[-g(\hat{A}^\dagger - \hat{A})]$  and  $\hat{U}_{A'}(t) = \exp[g(e^{-i\Phi'}\hat{A}'^\dagger - e^{i\Phi'}\hat{A}')]$  implying the following Bogoliubov transformations [12]:

$$\begin{bmatrix} \hat{a}_1(t) \\ \hat{a}_2(t)^\dagger \end{bmatrix} = \begin{bmatrix} C & S \\ S & C \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2^\dagger \end{bmatrix}; \quad \begin{bmatrix} \hat{a}'_1(t) \\ \hat{a}'_2(t)^\dagger \end{bmatrix} = \begin{bmatrix} C & \tilde{S} \\ \tilde{S}^* & C \end{bmatrix} \begin{bmatrix} \hat{a}'_1 \\ \hat{a}'_2^\dagger \end{bmatrix} \quad (9.2)$$

where:  $C \equiv \cosh(g)$ ,  $S \equiv \sinh(g)$ ,  $\tilde{S} \equiv \epsilon S$ ,  $\Gamma \equiv S/C$ ,  $\tilde{\Gamma} = \epsilon \Gamma$ ,  $\epsilon \equiv -e^{-i\Phi'}$  and  $\Phi'$  is an externally adjustable *intrinsic phase* existing between the  $A$  and  $A'$  OPA devices. These transformations imply the time invariance of the interaction Hamiltonian and of the field commutators, i.e.  $\hat{H}_I(t) = \hat{H}_I(0)$ ,  $[\hat{a}_i(t), \hat{a}_j^\dagger(t)] = [\hat{a}_i, \hat{a}_j^\dagger] = \delta_{ij}$ ,  $[\hat{a}'_i(t), \hat{a}'_j^\dagger(t)] = [\hat{a}'_i, \hat{a}'_j^\dagger] = \delta_{ij}$ ,  $[\hat{a}'_i(t), \hat{a}'_j^\dagger(t)] = 0$  being  $i, j = 1, 2$  and  $\hat{a}_{i,j} \equiv \hat{a}_{i,j}(0)$ ,  $\hat{a}'_{i,j} \equiv \hat{a}'_{i,j}(0)$  the input fields at the initial time  $t = 0$ , i.e. before the OPA interaction. The evolution in the Schrödinger picture of the state acted upon by the OPA system is determined by the overall operator  $\hat{U} = \hat{U}_A \otimes \hat{U}_{A'}$  expressed in terms of the operators evaluated at  $t = 0$ :

$$\hat{U}_A = \exp[g(\hat{\sigma}_+ + \hat{\sigma}_-)]; \quad \hat{U}_{A'} = \exp[g(\hat{\sigma}'_+ + \hat{\sigma}'_-)] \quad (9.3)$$

by adopting the definitions:  $\hat{\sigma}_+ = -\hat{A}^\dagger$ ,  $\hat{\sigma}_- = \hat{A}$ ,  $\hat{\sigma}_z = (1 + \hat{a}_1^\dagger \hat{a}_1 + \hat{a}_2^\dagger \hat{a}_2) \equiv (1 + \hat{n}_1 + \hat{n}_2)$ ,  $\hat{\sigma}'_+ = -\epsilon \hat{A}'^\dagger$ ,  $\hat{\sigma}'_- = \epsilon^* \hat{A}'$ ;  $\hat{\sigma}_z = (1 + \hat{n}'_1 + \hat{n}'_2)$ . In virtue of Eq.(9.2) the following commutation properties for the sets of the  $\hat{\sigma}$  and  $\hat{\sigma}'$ pseudo-spin operators hold:  $[\hat{\sigma}_+, \hat{\sigma}_-] = \hat{\sigma}_z$ ;  $[\hat{\sigma}_z, \hat{\sigma}_\pm] = \pm 2\hat{\sigma}_\pm$  and:  $[\hat{\sigma}'_+, \hat{\sigma}'_-] = \sigma'_z$ ;  $[\hat{\sigma}'_z, \hat{\sigma}'_\pm] = \pm 2\hat{\sigma}'_\pm$ ;  $[\hat{\sigma}_\pm, \hat{\sigma}'_\mp] = [\hat{\sigma}_\pm, \hat{\sigma}'_z] = 0$ . By further adopting the definitions:  $\hat{\sigma}_x \equiv 2^{-1/2}(\hat{\sigma}_+ + i\hat{\sigma}_-)$ ,  $\hat{\sigma}_y \equiv 2^{-1/2}(\hat{\sigma}_+ - i\hat{\sigma}_-)$ , the following relevant commutators:  $[\hat{\sigma}_x, \hat{\sigma}_y] = -i\hat{\sigma}_z$ ,  $[\hat{\sigma}'_x, \hat{\sigma}'_y] = -i\hat{\sigma}'_z$  are recognized as those belonging to the symmetry group [14]. The output field may be expressed, in virtue of an appropriate operator disentangling theorem in the following form [12, 15]:

$$|\Psi\rangle = \{\exp \Gamma(\hat{\sigma}_+ + \hat{\sigma}'_+) \times \exp[-\ln C(\hat{\sigma}_z + \hat{\sigma}'_z)] \times \exp \Gamma(\hat{\sigma}_- + \hat{\sigma}'_-)\} |\Psi\rangle_{IN}. \quad (9.4)$$

Take as input state into the QIOPA system the general *qubit*:

$$|\Psi\rangle_{IN} \equiv (\tilde{\alpha} |\Psi\rangle_{IN}^\alpha + \tilde{\beta} |\Psi\rangle_{IN}^\beta); \quad |\tilde{\alpha}|^2 + |\tilde{\beta}|^2 = 1 \quad (9.5)$$

defined in the  $2 \times 2$ -dimensional Hilbert space of polarizations ( $\vec{\pi}$ ) on the 2 interacting modes  $k_1$  and  $k_2$  with basis vectors:  $|\Psi\rangle_{IN}^\alpha = |1\rangle_{1h} |0\rangle_{1v} |0\rangle_{2h} |0\rangle_{2v} \equiv |1, 0, 0, 0\rangle$ ,  $|\Psi\rangle_{IN}^\beta = |0, 1, 0, 0\rangle$ . Here the general product state  $|x\rangle_{1h} \otimes |y\rangle_{1v} \otimes |z\rangle_{2h} \otimes |t\rangle_{2v}$  has been, and shall be henceforth, expressed by the shorthand:  $|x, y, z, t\rangle$ . In virtue of the general *information preserving* property of any nonlinear (NL) transformation of parametric type, the output state is again expressed by a “massive qubit” [16]:

$$|\Psi\rangle \equiv (\tilde{\alpha} |\Psi\rangle^\alpha + \tilde{\beta} |\Psi\rangle^\beta); \quad |\tilde{\alpha}|^2 + |\tilde{\beta}|^2 = 1 \quad (9.6)$$

This (pure) state is indeed a *Schrödinger-cat* state implying the quantum superposition of the *orthonormal* multiparticle states [12]

$$|\Psi\rangle^\alpha \equiv C^{-3} \sum_{i, j=0}^{\infty} (-\Gamma)^{i+j} \epsilon^j \sqrt{i+1} |i+1, j, j, i\rangle \quad (9.7)$$

$$|\Psi\rangle^\beta \equiv C^{-3} \sum_{i, j=0}^{\infty} (-\Gamma)^{i+j} \epsilon^j \sqrt{j+1} |i, j+1, j, i\rangle \quad (9.8)$$

Consider the density operator  $\rho \equiv |\Psi\rangle\langle\Psi|$  and its reductions over the  $\vec{\pi}$  – vector spaces relative to the spatial modes  $k_1$  and  $k_2$ :  $\rho_1 = Tr_2\rho$ ;  $\rho_2 = Tr_1\rho$ . These may be expanded as weighted superpositions of  $p$  – square matrices of order  $p = (n + 2)$ , the relative weight  $\Gamma^2 = \tanh^2 g$  of each two successive matrices being determined by the parametric gain.  $\Gamma^2 = 0$  for  $g = 0$  and approaches asymptotically the unit value for large  $g$ . In turn, the  $p$  – square matrices may be expressed as sum of  $2 \times 2$  matrices as shown by the following expressions:

$$\rho_1 = C^{-6} \sum_{n=0}^{\infty} \Gamma^{2n} \times \sum_{i=0}^n \left[ \begin{array}{cc} |\tilde{\beta}|^2 (n-i+1) & \tilde{\alpha}^* \tilde{\beta} \sqrt{(i+1)(n-i+1)} \\ \tilde{\alpha} \tilde{\beta}^* \sqrt{(i+1)(n-i+1)} & |\tilde{\alpha}|^2 (i+1) \end{array} \right] \quad (9.9)$$

written in terms of the Fock basis:  $\{|i\rangle_{1h}|n-i+1\rangle_{1v}; |i+1\rangle_{1h}|n-i\rangle_{1v}\}$ . Correspondingly:

$$\rho_2 = C^{-6} \sum_{n=0}^{\infty} \Gamma^{2n} \times \sum_{i=0}^{n+1} \left[ \begin{array}{cc} |\tilde{\beta}|^2 (n-i+1) & \epsilon^* \tilde{\alpha}^* \tilde{\beta} \sqrt{(n-i+1)i} \\ \epsilon \tilde{\alpha} \tilde{\beta}^* \sqrt{(n-i+1)i} & |\tilde{\alpha}|^2 i \end{array} \right] \quad (9.10)$$

in terms of the Fock basis:  $\{|n-i\rangle_{2h}|i\rangle_{2v}; |n-i+1\rangle_{2h}|i-1\rangle_{2v}\}$ . Interestingly enough, the value  $n$  of the sum indices appearing in Eqs.(9.9) and (9.10) coincides with the *number* of photon pairs generated by the QIOPA amplification. Note that all the  $2 \times 2$  matrices in Eqs.(9.9), (9.10) and then the  $p$  – square matrices resulting from their sums over the  $i$  – index, are non-diagonal as implied by the quantum superposition property of any *Schrödinger-cat* state. Note also that the OPA *intrinsic phase*  $\Phi'$  only affects the *antcloning channel*, i.e. the mode  $k_2$ .

**Universality.** A necessary common property of the QIOPA system in the present context is its *universality*, i.e. implying the *same* quantum efficiency of the amplifying apparatus for any input unknown qubit. That is, for a qubit spanning the entire Bloch sphere. In the experiment here described the qubits are assumed to represent states of polarization ( $\vec{\pi}$ ), as said, and quantum efficiency is expressed by the QIOPA parametric *gain*:  $g$ . We shall find that universality implies an important symmetry property, namely the invariance of the coupling Hamiltonian  $H_{int}$  under simultaneous general  $SU(2)$  transformations of the polarization  $\vec{\pi}$  of the spatial modes  $k_j$  ( $j = 1, 2$ ) [12, 17]. Assume that under a simultaneous general rotation  $R$  of  $\vec{\pi}$  on both modes  $k_j$ , the field set  $\{\hat{a}_j, \hat{a}'_j\}$  is changed into the set  $\{\hat{a}_{Rj}, \hat{a}'_{Rj}\}$  ( $j = 1, 2$ ). A general  $R(\vartheta, \xi)$  transformation, expressed in terms of complex parameters for which:  $|\vartheta|^2 + |\zeta|^2 = 1$ , relates the two field sets as follows:

$$\begin{bmatrix} \hat{a}_{R1} \\ \hat{a}'_{R1} \end{bmatrix} = R^\dagger \begin{bmatrix} \hat{a}_1 \\ \hat{a}'_1 \end{bmatrix} R = \begin{bmatrix} \vartheta & \zeta \\ -\zeta^* & \vartheta^* \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}'_1 \end{bmatrix}; \quad (9.11)$$

$$\begin{bmatrix} \hat{a}'_{R2} \\ \hat{a}_{R2} \end{bmatrix} = R^\dagger \begin{bmatrix} \hat{a}'_2 \\ \hat{a}_2 \end{bmatrix} R = \begin{bmatrix} \vartheta & \zeta \\ -\zeta^* & \vartheta^* \end{bmatrix} \begin{bmatrix} \hat{a}'_2 \\ \hat{a}_2 \end{bmatrix} \quad (9.12)$$

It can be easily checked that Eq.(9.1) can be re-expressed in terms of the new field set into the invariant form:  $\hat{H}_{RI} = i\hbar\chi[\hat{A}_R - \hat{A}'_R] + h.c.$ , where  $\hat{A}_R \equiv \hat{a}_{R1}\hat{a}_{R2}$ ,  $\hat{A}'_R \equiv \hat{a}'_{R1}\hat{a}'_{R2}$ , *only* by setting the OPA *intrinsic phase*  $\Phi' = 0$ , i.e.  $\epsilon = -1$ . Interestingly, note that the same dynamical condition  $\Phi' = 0$  implies the well known  $SU(2)$  invariance of the  $\vec{\pi}$ -entangled "singlet states" generated by SPDC, that is by the OPA when it is not *quantum-injected* (or, when it is only "injected" by the *vacuum field* on both input modes  $k_j$ ).

Since, in general, the input  $N \geq 1$  qubits injected into the amplifier are quantum superpositions of  $\vec{\pi}$ -states, the dynamical condition  $\Phi' = 0$  finally implies the *universality* of the overall cloning and U-NOT transformations. For the sake of a clearness in the future discussions, we find it convenient to re-cast the invariant Hamiltonian with  $\Phi' = 0$ , in the following form:

$$\hat{H}_{int} = i\hbar\chi \left( \hat{a}_\pi \hat{b}_{\pi\perp} - \hat{a}_{\pi\perp} \hat{b}_\pi \right) + h.c. \quad (9.13)$$

where  $\hat{a}$  and  $\hat{b}$  are the overall field operators acting respectively on the output modes  $k_1$  and  $k_2$ . For reasons that will become clear in the following sections these modes are referred to as the *cloning* and the *anticloning* modes, respectively. Furthermore, since  $g = \chi t$  is independent of any *unknown* polarization state of the injected field, we have denoted the creation  $\hat{a}^\dagger$ ,  $\hat{b}^\dagger$  and annihilation  $\hat{a}$ ,  $\hat{b}$  operators of a single photon in modes  $k_1$ ,  $k_2$  with subscripts  $\pi$  or  $\pi^\perp$  to indicate the invariance of the process with respect to the polarization states of the input particles. Of course, the  $SU(2)$  transformation for the fields  $\hat{a}$  is again expressed as follows:

$$\begin{bmatrix} \hat{a}_{R\pi} \\ \hat{a}_{R\pi\perp} \end{bmatrix} = R^\dagger \begin{bmatrix} \hat{a}_\pi \\ \hat{a}_{\pi\perp} \end{bmatrix} R = \begin{bmatrix} \vartheta & \zeta \\ -\zeta^* & \vartheta^* \end{bmatrix} \begin{bmatrix} \hat{a}_\pi \\ \hat{a}_{\pi\perp} \end{bmatrix} \quad (9.14)$$

and:  $|\vartheta|^2 + |\zeta|^2 = 1$ . The same R-transformation is valid for the fields  $\hat{b}$ . The polarization conditions  $\pi$  and  $\pi^\perp$  will be expressed respectively by the field state vectors  $|\Psi\rangle$  and  $|\Psi^\perp\rangle$  on the cloning and antycloning output modes of the apparatus.

**Wigner Function.** In order to inspect at a deeper lever the *Schrödinger-cat* condition, consider the Wigner function of the output field for the QIOPA apparatus shown in Fig. 9.2. Evaluate first the symmetrically-ordered characteristic function of the set of complex variables  $(\eta_j, \eta_j^*, \xi_j, \xi_j^*) \equiv \{\eta, \xi\}$ , ( $j = 1, 2$ ):  $\chi_s \{\eta, \xi\} \equiv \langle \Psi_0 | D[\eta_1(t)]D[\eta_2(t)]D[\xi_1(t)]D[\xi_2(t)] | \Psi_0 \rangle$  expressed in terms of the *displacement* operators:  $D[\eta_j(t)] \equiv \exp[\eta_j(t)\hat{a}_j^\dagger - \eta_j^*(t)\hat{a}_j]$ ,  $D[\xi_j(t)] \equiv \exp[\xi_j(t)\hat{a}_j^\dagger - \xi_j^*(t)\hat{a}_j]$  where:  $\eta_1(t) \equiv (\eta_1 C - \eta_2^* S)$ ;  $\eta_2(t) \equiv (\eta_2 C - \eta_1^* S)$ ;  $\xi_1(t) \equiv (\xi_1 C - \xi_2^* S)$ ;  $\xi_2(t) \equiv (\xi_2 C - \xi_1^* S)$ . The Wigner function, expressed in terms of the corresponding complex phase-space variables  $(\alpha_j, \alpha_j^*, \beta_j, \beta_j^*) \equiv \{\alpha, \beta\}$  is the eight-dimensional Fourier transform of  $\chi_s \{\eta, \xi\}$ , namely:

$$W\{\alpha, \beta\} = \frac{1}{\pi^8} \iiint d^2\eta_1 d^2\eta_2 d^2\xi_1 d^2\xi_2 \chi_s\{\eta, \xi\} e^{\sum_j [(\eta_j^* \alpha_j + \xi_j^* \beta_j) - c.c.]} \quad (9.15)$$

where  $d^2\eta_j \equiv d\eta_j d\eta_j^*$ , etc. By a lengthy application of operator algebra and integral calculus, reported in [18], we could evaluate analytically in closed form both  $\chi_s\{\eta, \xi\}$  and  $W\{\alpha, \beta\}$ . The *exact* expression of the Wigner function is found:

$$W\{\alpha, \beta\} = -\overline{W_A}\{\alpha\} \overline{W_{A'}}\{\beta\} \left[ 1 - |e^{i\Phi} \Delta_A\{\alpha\} + \Delta_{A'}\{\beta\}|^2 \right] \quad (9.16)$$

where  $\Delta_A\{\alpha\} \equiv 2^{-\frac{1}{2}}(\gamma_{A+} - i\gamma_{A-})$ ,  $\Delta_{A'}\{\beta\} \equiv 2^{-\frac{1}{2}}(\gamma_{A'+} - i\gamma_{A'-})$  are expressed in terms of the squeezed variables:  $\gamma_{A+} \equiv (\alpha_1 + \alpha_2^*)e^{-g}$ ;  $\gamma_{A-} \equiv i(\alpha_1 - \alpha_2^*)e^{+g}$ ;  $\gamma_{A'+} \equiv (\beta_1 + \beta_2^*)e^{-g}$ ;  $\gamma_{A'-} \equiv i(\beta_1 - \beta_2^*)e^{+g}$ . The Wigner functions

$$\begin{aligned} \overline{W_A}\{\alpha\} &\equiv 4\pi^{-2} \exp\left(-[|\gamma_{A+}|^2 + |\gamma_{A-}|^2]\right) \\ \overline{W_{A'}}\{\beta\} &\equiv 4\pi^{-2} \exp\left(-[|\gamma_{A'+}|^2 + |\gamma_{A'-}|^2]\right) \end{aligned}$$

definite positive over the 4-dimensional spaces  $\{\alpha\}$  and  $\{\beta\}$  represent the effect of squeezed-vacuum, i.e. emitted respectively by  $OPA_A$  and  $OPA_{A'}$  in absence of any injection. Inspection of Eq.(9.16) shows that precisely the quantum superposition character of the injected state  $|\Psi\rangle_{IN}$  determines the dynamical quantum superposition of the devices  $OPA_A$  and  $OPA_{A'}$ , the ones that otherwise act as *uncoupled* and *independent* objects. In other words, the quasi-probability functions  $\overline{W_A}\{\alpha\}$ ,  $\overline{W_{A'}}\{\beta\}$  corresponding to the two macrostates  $|\Psi\rangle^\alpha$  and  $|\Psi\rangle^\beta$  in absence of quantum superposition are defined in two totally separated and independent spaces  $\{\alpha\}$  and  $\{\beta\}$ . The link between these spaces is provided by the superposition term in Eq.(9.16):  $2Re[e^{i\Phi} \Delta_A\{\alpha\} \Delta_{A'}^*\{\beta\}]$ . This term provides precisely the first-order quantum interference of the macrostates  $|\Psi\rangle^\alpha$  and  $|\Psi\rangle^\beta$ . In addition, and most importantly, Eq.(9.16) shows the non definite positivity of  $W\{\alpha, \beta\}$  over its definition space. This assures the overall quantum character of our multiparticle injected amplification scheme [12, 19].

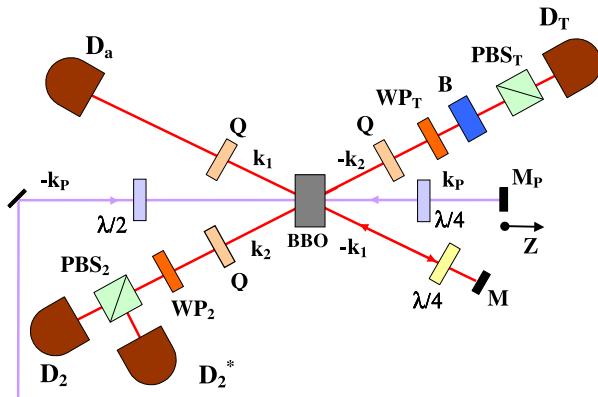
### 9.3 Realization of an “all optical” Schrödinger-Cat

Since the golden years of quantum mechanics the interference of classically distinguishable quantum states, first introduced by the famous *Schrödinger-cat* analogue [20] has been the object of extensive theoretical studies and recognized as a major conceptual paradigm of physics [21, 22]. In modern times the sciences of quantum information and quantum computation deal precisely with collective processes involving a multiplicity of interfering states, generally mutually entangled and rapidly de-phased by decoherence [23]. For many

respects the experimental implementation of this intriguing classical-quantum condition represents today an open problem in spite of recent successful studies carried out mostly with atoms, photon-atom, superconducting and Bose Einstein condensate systems [24–26]. In this section we report the first experimental demonstration of an all-optical Schrödinger cat, namely made of a multi-photon assembly in a quantum superposition state. In addition, at variance with the previous results with atoms, the output state is directly accessible to measurement or to further exploitation, e.g. by a quantum information *gate*. The scheme is based on the process of the *quantum-injected* optical parametric amplification of a single photon in a superposition state of polarization ( $\pi$ ), i.e. a  $\pi$ -encoded *qubit* [12, 27].

Conceptually, the method consists of transferring the well accessible and easily achievable condition of quantum superposition characterizing a single input qubit to a *macroscopic*, i.e. multi-particle, amplified state. In virtue of the very general *information preserving* property of the optical parametric amplification (OPA) the generated macroscopic state is found to keep the *same* superposition character and the interfering capabilities of the injected input qubit, thus realizing a genuine multi-photon Schrödinger-cat state [16]. Since the amplification scheme realized here basically consists of a *universal quantum cloning* process, i.e. acting identically on *any* input qubit, it will be necessarily affected by *squeezed vacuum* (SV) noise, i.e. arising from the amplification of the input vacuum-state [9, 12, 16, 27].

Let's refer to the experimental apparatus represented in Fig. 9.3. The



**Fig. 9.3.** Experimental realization of the Multi-particle Entangled Schrödinger cat. A single photon,  $N = 1$ , was injected with a definite  $\pi$ -state into the nonlinear crystal of OPA along the mode  $k_1$ . The output state was detected on mode  $k_2$ .

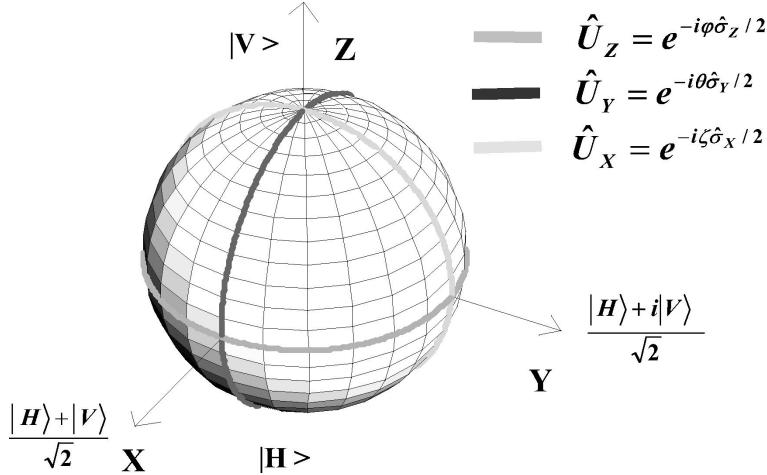
OPA active element was a nonlinear crystal slab (BBO:  $\beta$ -barium borate), 1.5 mm thick cut for Type II phase-matching, able to generate by SPDC

$\vec{\pi}$  – entangled pairs of photons. Precisely, the OPA *intrinsic phase* was set as to generate by SPDC *singlet* entangled states on the output optical modes. As it is well known, this condition ensured the *universality* of the quantum cloning transformation in the QIOPA scheme [9,27]. The photons of each pair were emitted with equal wave lengths  $\lambda = 795\text{nm}$  over two spatial modes  $-k_1$  and  $-k_2$  owing to a SPDC process excited by a coherent pump UV field provided by a Ti:Sa Coherent MIRA mode-locked pulsed laser coupled to a second harmonic generator and associated with a mode with wave vector  $\mathbf{k}_p$  and wave length  $\lambda_p = 397.5\text{nm}$ . The average UV power was  $0.25W$ , the pulse repetition rate  $76\text{ MHz}$  and the time duration of each UV pulse as well as of the generated single photon pulses were  $\tau = 140\text{ fs}$ . The SPDC process implied a 3-wave nonlinear parametric interaction taking place towards the right hand side (r.h.s.) of Fig. 9.3. The UV pump was back-reflected over the mode  $-\mathbf{k}_p$  onto the nonlinear crystal by a spherical mirror  $\mathbf{M}_p$ , with micro-metrically adjustable position  $\mathbf{Z}$ , thus exciting the main OPA amplification, or quantum cloning process towards the left hand side (l.h.s.) of Fig. 9.3. By the combined effect of two adjustable optical UV wave plates ( $\lambda/2 + \lambda/4$ ) acting on the projections of the linear polarization  $\pi_p$  of the UV field on the fixed optical axis of the BBO crystal for the  $\mathbf{k}_p$  and  $-\mathbf{k}_p$  counter-propagating excitation processes, the SPDC excitation was always kept at a very low level while the main OPA amplification could reach *any large intensity*, as we shall see shortly. Precisely, by a smart balancing of the orientation angles  $\vartheta_{\frac{1}{2}}$  and  $\vartheta_{\frac{1}{4}}$  of the UV wave plates, the probabilities for the SPDC emission, i.e. towards the r.h.s. of Fig. 9.3, of two simultaneous correlated photon pairs and of a single pair was always kept below  $10^{-2}$  in any high gain condition. One of the photons of the SPDC emitted pair, back-reflected by a fixed mirror  $\mathbf{M}$ , was re-injected onto the amplifying nonlinear crystal by the input mode  $\mathbf{k}_1$ , while the other photon emitted over mode  $(-\mathbf{k}_2)$  excited the detector  $D_T$ , the *trigger* of the overall *conditional* experiment. A proper setting of  $\mathbf{Z}$  secured the space-time overlapping into the nonlinear crystal of the interacting re-injected pulses with wl's  $\lambda_p$  and  $\lambda$ , and then determined the optimal QIOPA condition. In all experiments the time optical walk-off effects due to the crystal birefringence were compensated by inserting in the modes  $\mathbf{k}_1$  and  $\mathbf{k}_2$  three fixed X-cut quartz plates  $Q$  and one  $\lambda/4$  wave plate. Before re-injection into the nonlinear crystal, the input qubit on mode  $\mathbf{k}_1$ ,

$$|\Psi\rangle_{in} = (\tilde{\alpha}|\Psi\rangle_{in}^{\alpha} + \tilde{\beta}|\Psi\rangle_{in}^{\beta}), |\tilde{\alpha}|^2 + |\tilde{\beta}|^2 = 1 \quad (9.17)$$

represented by the Bloch sphere shown in Fig. 9.4, underwent unitary  $SO(3)$  rotation  $\Phi$  transformations:  $\widehat{U}_j \equiv \exp(-i\sigma_j\Phi/2)$  around the three cartesian axes  $j = x, y, z$  by the combined action of the  $\lambda/2$  wave plate  $WP_T$ , of the micro-metrically adjustable *Babinet compensator*  $B$  and of the polarizing beam-splitter  $PBS_T$  acting on mode  $(-\mathbf{k}_2)$  in virtue of the nonlocality correlating modes  $\mathbf{k}_1$  and  $\mathbf{k}_2$ . These  $SO(3)$  transformations are represented in Fig. 9.4 by the circles drawn on the surface of the Bloch-sphere related to the input

qubit. The main OPA process, i.e. acting towards the l.h.s. of Fig. 9.3 on the



**Fig. 9.4.** Transformations on the Bloch sphere of the injected qubit.  $|\Psi\rangle_{IN}$  was moved along three different trajectories corresponding to  $\sigma_X$  (white curve),  $\sigma_Y$  (black curve) and  $\sigma_Z$  (gray curve) rotations.

injected qubit,  $|\Psi\rangle_{in}$  was characterized by two different excitation regimes, establishing two corresponding “sizes” of the generated Schrödinger cat: a) *low gain* regime, characterized by a low excitation UV energy ( $3.5nJ$ ) per pulse, leading to a small value of the nonlinear parametric *gain*  $g = 0.07$ ; b) *high gain* regime, characterized by a larger value of the *gain* about  $g = 1.15$ . This last condition was attained by the adoption of a Ti-Sa regenerative amplifier REGA operating at 250 kHz with a pulse duration equal to  $180fs$ .

Assume that the input *pure state* qubit, Eq.(9.17) represents the superposition of the following two polarization ( $\pi$ )-states [27]:  $|\Psi\rangle_{in}^\alpha = |1, 0, 0, 0\rangle$  and  $|\Psi\rangle_{in}^\beta = |0, 1, 0, 0\rangle$ , accounting respectively for 1 photon with horizontal ( $h$ ) polarization on input mode  $\mathbf{k}_1$ , vacuum state on the input mode  $\mathbf{k}_2$ , and 1 photon with vertical ( $v$ ) polarization on mode  $\mathbf{k}_1$ , vacuum state on the mode  $\mathbf{k}_2$ . As seen in the previous section, the *exact* expression of the output *pure state* is found to be expressed by a “massive qubit”:  $|\Psi\rangle \equiv (\tilde{\alpha} |\Psi\rangle^\alpha + \tilde{\beta} |\Psi\rangle^\beta)$ , where the states  $|\Psi\rangle^\alpha$  and  $|\Psi\rangle^\beta$  are expressed in Eqs.(9.7) and (9.8). We consider the reduced density operators of the pure state  $|\Psi\rangle$  over the modes  $\mathbf{k}_1$  and  $\mathbf{k}_2$  given in Eqs.(9.9) and (9.10). Note also that all the involved matrices are *non-diagonal* as implied by the quantum superposition property of the overall *Schrödinger-cat* state. Furthermore this one possesses the nonlocal property of all entangled, non separable quantum systems: it is indeed a bipartite *entangled Schrödinger cat*. It is worth noting that the Von Neumann

entropies  $S(\rho_j) = Tr(\rho_j \log_2 \rho_j)$ ,  $j = 1, 2$  are equal, thus implying the same *degree of mixedness* on the 2 output channels:  $S(\rho_1) = S(\rho_2)$  [27, 28].

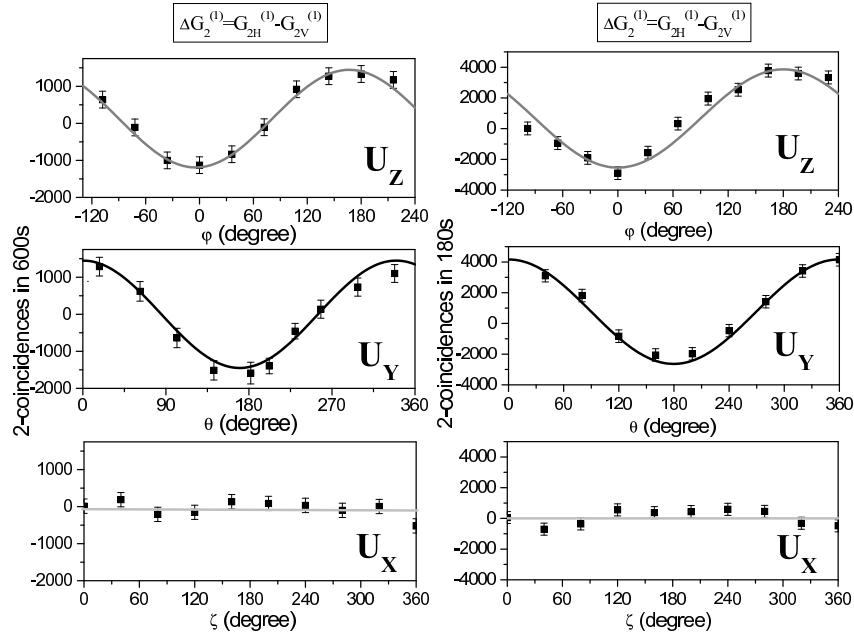
The Schrödinger-cat detection at the OPA output could be undertaken either (a) on the injection, i.e. cloning mode  $\mathbf{k}_1$ , or (b) on the anticoloning mode  $\mathbf{k}_2$ , or (c) on both output modes. We selected the option (b) since *any* interference effect registered on mode  $\mathbf{k}_2$  was by itself an unambiguous demonstration that the Schrödinger-cat condition was attained. Indeed, we tested carefully that in absence of any QIOPA action only (non interfering) *vacuum-* or *squeezed-vacuum fields* could be detected on the output mode  $\mathbf{k}_2$ .

By the two detectors  $D_2$  and  $D_{2*}$  coupled to the two mutually orthogonal fields  $\hat{c}_j(t)$ ,  $j = 1, 2$  emerging from the polarization beam-splitter  $PBS_2$  inserted in the output mode  $\mathbf{k}_2$ , the first-order correlation functions  $G_{2j}^{(1)} \equiv \langle \Psi | \hat{N}_j(t) | \Psi \rangle$  could be measured. They are ensemble averages of the number operators  $\hat{N}_j(t) \equiv \hat{c}_j^\dagger(t) \hat{c}_j(t)$  written in terms of the detected fields:  $\hat{c}_1(t) \equiv 2^{-\frac{1}{2}}[\hat{b}_h(t) + \hat{b}_v(t)]$ ,  $\hat{c}_2(t) \equiv 2^{-\frac{1}{2}}[\hat{b}_h(t) - \hat{b}_v(t)]$  where  $\hat{b}_i(t)$ , defined on the mode  $\mathbf{k}_2$ , underwent a  $45^0\pi$ -rotation by the  $\lambda/2$  wave plate ( $WP_2$ ) before injection into  $PBS_2$  (see Fig. 9.3). By expressing Eq.(9.17) as:  $|\Psi\rangle_{IN} \equiv e^{i\gamma} (\alpha |\Psi\rangle_{IN}^\alpha + \beta e^{i\varphi} |\Psi\rangle_{IN}^\beta)$  with  $\alpha$  and  $\beta$  real numbers, the  $G_{2j}^{(1)}$  could be given in the form,

$$G_{21}^{(1)} = \bar{n} + \frac{1}{2}\bar{n}[1 - 2\alpha\beta \cos \varphi]; G_{22}^{(1)} = \bar{n} + \frac{1}{2}\bar{n}[1 + 2\alpha\beta \cos \varphi] \quad (9.18)$$

showing the expected superposition character of the output field with respect to  $\alpha, \beta$  and  $\varphi$ . By comparison with the corresponding averages over the input *vacuum state*:  $G_{21,vac}^{(1)} = G_{22,vac}^{(1)} = \bar{n}$ , which are independent of  $\varphi$  and  $\alpha, \beta$ , as said, the signal-to-noise-ratio related to the Schrödinger-cat detection was found  $S/N = 2$ , for  $\varphi = 0$  and  $\alpha = \beta = 2^{-\frac{1}{2}}$ .

The above result immediately suggests a *1<sup>st</sup>-order*  $\pi$ -interferometric method for Schrödinger-cat detection on a *single*  $\mathbf{k}_j$  beam, with *visibility*:  $V = (G_{2\max}^{(1)} - G_{2\min}^{(1)})/(G_{2\max}^{(1)} + G_{2\min}^{(1)}) \geq \frac{1}{3}$ . Since the correlation-function  $G^{(1)}$  exhibits unambiguously the *1<sup>st</sup>-order* interference property of any quantum object [19], the fringe patterns shown in Fig. 9.5 are indeed clear evidence of the Schrödinger-cat condition. Precisely, these patterns express, in correspondence with the two different dynamical regimes of *low gain* and *high gain*, the quantity  $\Delta G_2^{(1)} \equiv (G_{21}^{(1)} - G_{22}^{(1)})$  i.e., the *difference* between the correlation functions measured at the two outputs of  $PBS_2$ .  $\Delta G_2^{(1)}$  is free from noise as the two components  $G_{21}^{(1)}$  and  $G_{22}^{(1)}$  are affected by the *same* amount of *SV noise*, i.e. the one implied by any approximate *quantum cloning* process [9, 27]. Fig. 9.5 (left panel) refers to the *low gain* excitation condition:  $g = 0.07$ . At this level, approximately only one single photon pair, i.e.  $n = 1$  in Eqs.(9.7), (9.8), was generated by the QIOPA system and then only an overall average number of three photons were making the “Schrödinger kitten.” A relatively *High gain* condition ( $g = 1.15$ ) is expressed by the plots given by Fig. 9.5



**Fig. 9.5.** Experimental result of  $\Delta G_2^{(1)}$  in case of low (left) and high (right) gain regime amplification. The three patterns correspond to the trajectories represented in Fig. 9.4. As expected no oscillation is present in the case of  $U_X$  transformation. The two regime conditions exhibit the same coherence behavior.

(right panel) showing the realization of a much “fatter” *Schrödinger cat*, i.e. made of about  $n = 4$  quantum-interfering photon pairs. The value of  $n$  has been estimated from the number of detected photons and the overall quantum efficiency of the detectors. Theoretically, by moving the injected state over the Bloch sphere, a visibility  $V_{th} = (G_{2\max}^{(1)} - G_{2\min}^{(1)}) / (G_{2\max}^{(1)} + G_{2\min}^{(1)}) = 33\%$  of the fringe-pattern was expected. Due to experimental imperfections thus far we have measured  $V \approx 4\%$  with a 2-coincidence detection measurement scheme and  $V \approx 7\%$  with a 4-coincidence scheme. In spite of that, the plots of Fig. 9.5 show, with no ambiguities, the quantum interference properties expected by the Schrödinger cat [12, 19, 20]. Of course, this highly nontrivial result is expected to be *linearly scaled* by adoption of a more efficient nonlinear crystal, e.g. made of two mutually stuck, orthogonally oriented Type-I crystals [29] and of a more powerful UV source.

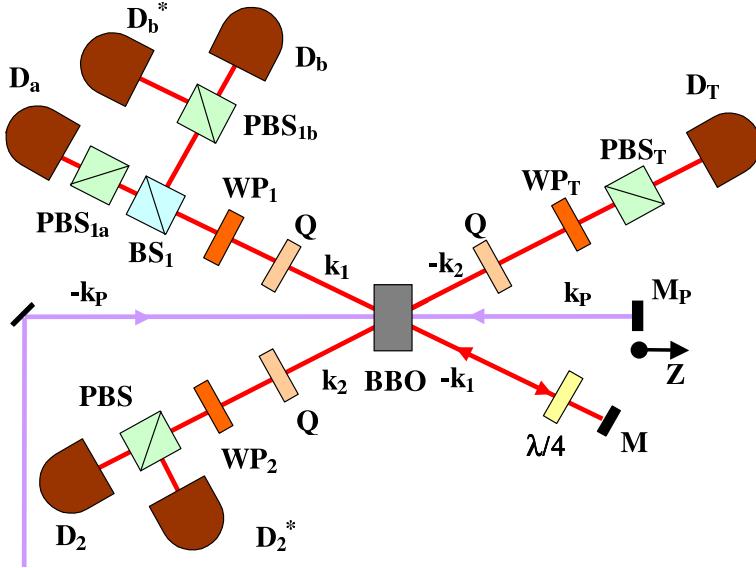
In conclusion, we have reported the successful implementation of a multi-particle bipartite *Entangled Schrödinger cat*, i.e. possessing all the appealing nonlocal properties of the nonseparable quantum systems. Since the latter additional feature was absent in the 1935 proposal by Schrödinger [20], indeed the very first proponent and supporter of quantum entanglement, we believe that the conceptual relevance of the present realization goes even beyond that

already intriguing quantum riddle. Furthermore, we believe that the present realization will open a new trend of studies on the persistence of the validity of several crucial laws of quantum mechanics for entangled mixed-state systems of increasing complexity [30]. Another example would be the study of the violation of Bell-type inequalities in the multi-particle regime. Indeed, an outstanding new perspective in quantum nonlocality has been investigated in our Laboratory [31]. In the framework of quantum information technology the generation of *massive-qubits* carrying the information of the corresponding *single-particle* input qubits will make possible the practical realization of the universal multiple-qubit logic gates using photons as information carrying particles. For instance a multi-particle *massive-qubit* acting as *control qubit* could be the only solution towards the actual implementation of the problematic *quantum-nondemolition* function implied by the XOR or Fredkin gates.

#### 9.4 Universal Optimal Quantum Cloning Machine and U-NOT gate

In the present section the adopted experimental apparatus, the quantum-injected optical parametric amplifier is transformed simultaneously into a universal optimal quantum cloning machine and into a Universal NOT quantum-information gate. The two processes, indeed *forbidden* in their *exact* form for fundamental quantum limitations, will be found to be *universal* and *optimal*, i.e. the measured *fidelity* of both processes  $F < 1$  will be found close to the limit values evaluated by quantum theory. Consider first the general case of an input  $\vec{\pi}$ -encoded qubit  $|\Psi\rangle_{in}$  associated with a single photon with wave length  $\lambda$ , injected on the input mode  $k_1$  of the QIOPA, the other input mode  $k_2$  being in the vacuum state [9, 12, 27]. We adopted the same OPA configuration used in the experiment reported in the previous section. Particularly the nonlinear crystal orientation was set as to realize the universality of the OPA dynamical interaction, i.e. the insensitivity of the amplification quantum efficiency to *any* input qubit state  $|\Psi\rangle$ . As said before, the peculiar interaction Hamiltonian Eq.(9.13) ensures the necessary *universality* property of the cloning machine and of the U-NOT gate under investigation.

The QIOPA apparatus was arranged in the folded *self-injected* configuration shown in Fig. 9.6. The UV pump beam, back reflected by a spherical mirror  $M_p$  with 100% reflectivity and micrometrically adjustable position  $\mathbf{Z}$ , excited the nonlinear crystal OPA in both directions  $-k_p$  and  $k_p$ , i.e. correspondingly oriented towards the right and the left sides of Fig. 9.6. A SPDC process excited by the  $-k_p$  pump mode created single pairs of photons in entangled singlet-states of linear polarization,  $\vec{\pi}$ . One photon of each pair, emitted over  $-k_1$  was reflected by a spherical mirror  $M$  into the nonlinear crystal where it provided the  $N = 1$  *quantum injection* into the OPA excited by the UV-pump beam associated with the back reflected mode  $k_p$ . Because



**Fig. 9.6.** Schematic diagram of the *universal optimal quantum cloning machine* realized on the cloning channel (mode  $k_1$ ) of a *self-injected* OPA and of the universal-NOT gate realized on the antycloning channel,  $k_2$ . Micrometric adjustments of the coordinate  $Z$  of the UV mirror  $M_p$  ensured the time superposition in the active nonlinear crystal of the UV 140 femtosecond pump pulses and of the single photon pulse injected via back reflection by the fixed mirror  $M$ .

of the low pump intensity, the probability of the unwanted  $N = 2$  photons injection has been evaluated to be  $10^{-2}$  smaller than the one for  $N = 1$ . The twin SPDC generated photon emitted over  $-k_2$  was selected by the devices (wave plate and polarizing beam splitter:  $WP_T + PBS_T$ ), detected by  $D_T$  and provided the trigger of the overall conditional experiment. Because of the Einstein-Podolsky-Rosen non-locality implied by the SPDC emitted singlet state, the selection on mode  $-k_2$  provided the realization on  $k_1$  of the polarization states  $|\Psi\rangle_{in}$  of the injected photon. By adopting  $\lambda/2$  or  $\lambda/4$  wave plates with different orientations of the optical axis, the following states were injected into the OPA:  $|H\rangle$ ,  $2^{-1/2}(|H\rangle + |V\rangle)$ , and  $2^{-1/2}(|H\rangle + i|V\rangle)$ .

Let us consider the injected photon in the mode  $k_1$  to have *any* polarization  $\vec{\pi}=\Psi$ , corresponding to the (*unknown*) input qubit  $|\Psi\rangle_{in}$ . We will describe this polarization state as  $\hat{a}_{\Psi}^{\dagger}|0,0\rangle_{k_1} = |1,0\rangle_{k_1}$  where  $|m,n\rangle_{k_1}$  represents a state with  $m$  photons of the mode  $k_1$  having the polarization  $\Psi$ , while  $n$  photons have the polarization  $\Psi^\perp$ . Assume that the mode  $k_2$  is in the vacuum state. The initial polarization state of modes  $k_i$  ( $i = 1, 2$ ) reads  $|\Psi\rangle_{in} = |1,0\rangle_{k_1} \otimes |0,0\rangle_{k_2}$  and it evolves according to the unitary operator  $\hat{U} \equiv \exp(-i\hat{H}_{int}t)$ :

$$\widehat{\mathbf{U}} |\Psi\rangle_{in} \simeq |1,0\rangle_{k_1} \otimes |0,0\rangle_{k_2} + g \left( \sqrt{2} |2,0\rangle_{k_1} \otimes |0,1\rangle_{k_2} - |1,1\rangle_{k_1} \otimes |1,0\rangle_{k_2} \right) \quad (9.19)$$

This represents the first approximation for the general output state vector  $|\Psi\rangle_{out}$ , a *pure* state describing the two modes at times  $t > 0$  [9]. Be  $\varrho \equiv (|\Psi\rangle\langle\Psi|)_{out}$  the corresponding density operator. It has been shown that  $|\Psi\rangle_{out} = \widehat{\mathbf{U}} |\Psi\rangle_{in}$  is equal to the output state of a general UOQCM, i.e.  $|\Psi\rangle_{out} = |\Psi_N\rangle_M$  where  $N$  and  $M > N$  are respectively the number of input qubits to be cloned and the number of clones, as said [4]. The linearization procedure implying the above approximation, i.e. the restriction to the simplest case  $N = 1$ ,  $M = 2$ , is justified in the present experiment by the very small value of the amplification gain:  $g \approx 0.1$ . The zero order term in Eq.(9.19) corresponds to the process when the input photon in the mode  $k_1$  did not interact in the nonlinear medium, while the second term describes the first order amplification process in the OPA. Here the state  $|2,0\rangle_{k_1}$  describing two photons of the mode  $k_1$  in the polarization state  $\Psi$  corresponds to the state  $|\Psi\Psi\rangle$ . This state-vector describes the cloning of the original qubit  $N = 1$  [33]. The vector  $|0,1\rangle_{k_2}$  describes the state of the mode  $k_2$  with a single photon in with the polarization  $\Psi^\perp$ . That is, this state vector represents the flipped version of the input qubit on mode  $k_1$  and then the QIOPA acts on the output mode  $k_2$  as a universal NOT-gate [9].

To see that the stimulated emission is indeed responsible for creation of the flipped qubit, let us compare the state Eq.(9.19) with the output of the OPA when the vacuum is injected into the nonlinear crystal on *both* input modes  $k_i$  ( $i = 1, 2$ ). In this SPDC case is  $|\Psi\rangle_{in}^0 = |0,0\rangle_{k_1} \otimes |0,0\rangle_{k_2}$ , and we obtain to the same order of approximation as above:

$$\widehat{\mathbf{U}} |\Psi\rangle_{in}^0 \simeq |0,0\rangle_{k_1} \otimes |0,0\rangle_{k_2} + g \left( |1,0\rangle_{k_1} \otimes |0,1\rangle_{k_2} - |0,1\rangle_{k_1} \otimes |1,0\rangle_{k_2} \right) \quad (9.20)$$

We see that the flipped qubit described by the state vector  $|0,1\rangle_{k_2}$  in the right-hand sides of Eqs.(9.19) and (9.20) do appear with different amplitudes corresponding to the ratio of probabilities to be equal to  $R^* = 2 : 1$ . Sometimes  $R^*$  is referred to as signal-to-noise ratio:  $S/N$ . Note in Eqs.(9.19), (9.20) that, by calling by  $R$  the ratio of the probabilities of detecting 2 and 1 photons on mode  $k_1$  only, we obtain:  $R = R^*$ . In other words the *same value* of  $S/N$  affects both cloning and U-NOT processes realized simultaneously on the two different output modes:  $k_1$  and  $k_2$ . We do believe that this circumstance is highly significant. The ratios  $R$  and  $R^*$  are indeed the quantities measured in the present experiments for both the UOQCM and U-NOT-gate processes, respectively. These ratios will be used to determine, for both processes, the corresponding values of the *fidelity*, which can be defined as follows [4, 9, 33]:

$$\text{UOQCM fidelity} : F = \text{Tr}(\rho_1 \hat{n}_{1\pi}) / \text{Tr}(\rho_1 \hat{n}_1) = \frac{2R+1}{2R+2} = 5/6 \quad (9.21)$$

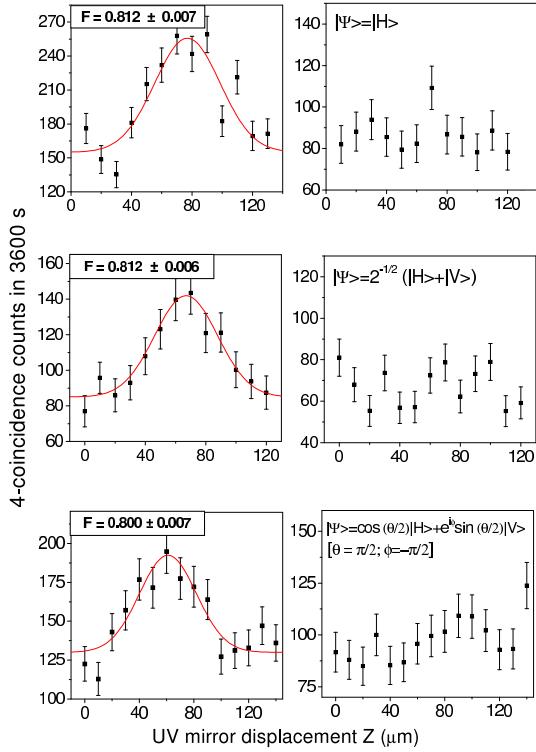
$$\text{U-NOT fidelity} : F^* = \text{Tr}(\rho_2 \hat{n}_{2\perp}) / \text{Tr}(\rho_2 \hat{n}_2) = \frac{R^*}{R^* + 1} = 2/3$$

where  $\hat{n}_i = \hat{n}_{i\Psi} + \hat{n}_{i\perp}$ , being  $\hat{n}_{1\Psi} \equiv \hat{a}_{\Psi}^{\dagger} \hat{a}_{\Psi}$ ,  $\hat{n}_{1\perp} \equiv \hat{a}_{\Psi\perp}^{\dagger} \hat{a}_{\Psi\perp}$ ,  $\hat{n}_{2\Psi} \equiv \hat{b}_{\Psi}^{\dagger} \hat{b}_{\Psi}$ ,  $\hat{n}_{2\perp} \equiv \hat{b}_{\Psi\perp}^{\dagger} \hat{b}_{\Psi\perp}$  and the reduced density operators of the output state on the modes  $k_i$  ( $i = 1, 2$ ) are:  $\varrho_1 = Tr_2\rho$  and  $\varrho_2 = Tr_1\rho$ . The values  $F = 5/6$  and  $F^* = 2/3$  are the optimal *fidelity* values allowed by quantum mechanics respectively for cloning  $N = 1$  into  $M = 2$  qubits and for realizing a U-NOT gate by single qubit flipping.

As said, the goal of the *cloning* experiment was to measure, under injection of the state  $|\Psi\rangle$ , the OPA amplification  $R$  on the output mode  $k_1$ , here referred to as the cloning mode carrying the *same*  $\vec{\pi}$  condition corresponding of the input state. Contextually, no amplification should affect the output state corresponding to the polarization  $\vec{\pi}^\perp$  orthogonal to  $\vec{\pi}$ . In order to perform this task, the  $PBS_2$  was removed on the mode  $k_2$  and the photons on the same mode detected by a single detector:  $D_2$ . The two-cloning photons associated with the cloning mode,  $k_1$  were separated by means of a 50 : 50 conventional beam splitter ( $BS_1$ ) and their polarization states were analyzed by the combinations of  $WP_1$  and of  $PBS_{1a}$  and  $PBS_{1b}$ . For each injected state of polarizations  $|\Psi\rangle$ ,  $WP_1$  was set in order to detect  $|\Psi\rangle$  by detectors  $D_a$  and  $D_b$  and to detect the state orthogonal to  $|\Psi\rangle$ , viz.  $|\Psi^\perp\rangle$  by  $D_b^*$ . Hence any coincidence event detected by  $D_a$  and  $D_b$  corresponded to the realization of the state  $|\Psi\Psi\rangle_1$  over the cloning mode, while a coincidence detected by  $D_a$  and  $D_b^*$  corresponded to the state  $|\Psi\Psi^\perp\rangle_1$ .

The measurement of  $R$  was carried out by 4-coincidence measurements involving simultaneously the detectors of the set  $[D_2, D_T, D_a, D_b]$  and the one of the set:  $[D_2, D_T, D_a, D_b^*]$ . The experimental data reported on the left hand side (l.h.s.) of Fig. 9.7 corresponds to the 4-coincidence measurement by  $[D_2, D_T, D_a, D_b]$ , that is, to the emission over the C-mode of the state  $|\Psi\Psi\rangle_1$  under injection of the input state  $|\Psi\rangle_{IN}$ . The resonance peaks found by this last measurement identified the position  $Z$  of the UV retro-reflecting mirror for which the overlapping of the pump and of the injected pulses was realized, i.e. the actual realization of the *quantum injected* amplification. The right hand side (r.h.s.) column of Fig. 9.7 reports the corresponding 4-coincidence data obtained by the set  $[D_2, D_T, D_a, D_b^*]$ , i.e. implying the realization on the cloning mode of the state  $|\Psi\Psi^\perp\rangle_1$ . As expected from theory no amplification peaks are present in this case, here referred to as a noise condition. The values of the signal-to-noise ratio  $R$  were determined as the ratio of the peak values of the plots on the l.h.s. of Fig. 9.7 and of the values of the corresponding plots on the r.h.s. measured at the same value of  $Z$ . The results are:  $F_H = 0.812 \pm 0.007$ ;  $F_{H+V} = 0.812 \pm 0.006$ ;  $F_{left} = 0.800 \pm 0.007$ , to be compared with the optimal value given above in Eq.(9.21):  $F_{th} = 5/6 \approx 0.833$ .

The U-NOT gate operation has been demonstrated by restoring the  $PBS_2$  on the AC-mode,  $k_2$  coupled to the detectors  $D_2$ ,  $D_{2*}$ , via the  $WP_2$ , as shown in Fig. 9.6. The  $\vec{\pi}$ -analyzer consisting of ( $PBS_2 + WP_2$ ) is set as to reproduce the *same* filtering action of the analyzer ( $PBS_T + WP_T$ )



**Fig. 9.7.** Demonstration of the UOQCM by single photon qubit self-injection and 4-detector coincidences. The plots on the l.h.s. and on the r.h.s. of the figure mutually correspond. The values of the *fidelity* of the process  $F$  evaluated by each test are expressed in the upper side of the l.h.s. plots while the corresponding test qubits are expressed in r.h.s. plots.

for the trigger signal. On the cloning channel,  $k_1$ , the devices  $PBS_{1a}$  and  $PBS_{1b}$  were removed and the field was simply coupled by the *normal* Beam Splitter ( $BS_1$ ) to the detectors  $D_a$  and  $D_b$ . A coincidence event involving these ones was the *signature* for a *cloning* event. The values of the signal-to-noise ratio  $R^*$  evaluated on the basis of the data of the 4-coincidence experiments involving the sets  $[D_2, D_T, D_a, D_b]$  and  $[D_{2*}, D_T, D_a, D_b]$  and reported in Fig. 9.8, were adopted to determine the corresponding values of the U-NOT *fidelity*  $F^*$  formally expressed by Eq.(9.21). The results are:  $F_H^* = 0.630 \pm 0.008$ ;  $F_{H+V}^* = 0.625 \pm 0.016$ ;  $F_{left}^* = 0.618 \pm 0.013$  to be compared with the optimal value:  $F^* = 2/3 \approx 0.666$ . The 4-coincidence method was adopted in all experiments because it allowed a better spatial mode filtering of the system, leading to a larger experimental value of  $R$ . Note also that the height of the amplification peak in Figs. 9.7 and 9.8 does not decrease towards zero for large values of  $Z$ , as expected. This effect is

attributable entirely to the limited time resolution of the 4-coincidence apparatus. The effect would disappear if the resolution could be pushed into the sub-picosecond range, precisely of the order of the time duration of the interacting pump and injection pulses:  $\tau' \simeq 140$  fs. Such a resolution is hardly obtainable with the present technology. Note that all results reported in the left columns of Figs. 9.7 and 9.8 show an amplification efficiency that is almost identical for the adopted different  $\vec{\pi}$  conditions corresponding to the input state:  $|\Psi\rangle_{in}$ . This significant result represents the first demonstration of the universality of the QIOPA system carried out by *quantum injection* of a *single* qubit.

## 9.5 More on Contextuality and Universality

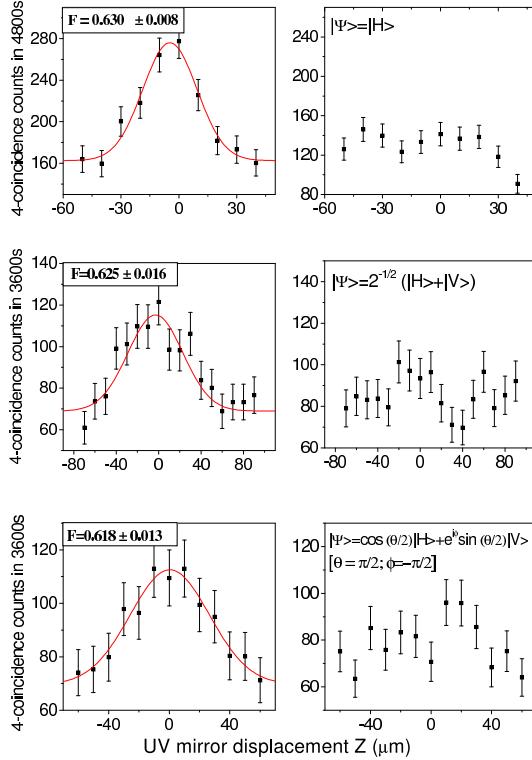
A remarkable, somewhat intriguing aspect of the previous experiment is that both processes of quantum cloning and the U-NOT gate are realized contextually by the *same* physical apparatus, by the same unitary transformation and correspondingly by the same quantum logic network. To the best of our knowledge it is not well understood yet why these “forbidden” processes can be so closely related. We may try to illuminate at least one formal aspect of this correlation.

Recall that the two processes are detected in our experiment over the two  $\vec{\pi}$ -*entangled* output cloning and antycloning channels, physically represented by the two corresponding wave vectors  $\vec{k}_j$  ( $j = 1, 2$ ). In addition, the overall output vector state  $|\Psi\rangle$  given by Eq.(9.6) is a *pure state* since the *unitary* evolution operator  $\hat{U} = \hat{U}_A \otimes \hat{U}_{A'}$  (Eq.(9.3)) acts on a *pure* input state. As a consequence, a result already considered in section 9.2, the reduced density matrices  $\rho_1$  and  $\rho_2$  have the same eigenvalues and the entanglement of the *bi-partite* state  $|\Psi\rangle$  can be conveniently measured by the *entropy of entanglement*:

$$E(\Psi) = S(\rho_1) = S(\rho_2) \quad (9.22)$$

where  $S(\rho_j) = -Tr\rho_j \log_2 \rho_j$  is the Von Neumann entropy of either the cloning or antycloning subsystem:  $j$  ( $j = 1, 2$ ) [28]. Let’s consider the approximate *cloning* process performed by our *universal optimal quantum cloning machine* acting on the cloning channel,  $k_1$ . What has been actually realized in the experiment is a procedure of “linearization” of the cloning map that is *nonlinear* and, as such cannot be realized exactly by nature [1, 5, 6]. By this procedure a *mixed-state* condition of the output state  $\rho_1$  was achieved corresponding to the *optimal* limit value of the entropy  $S(\rho_1) > 0$ . In virtue of Eq.(9.22) the *same mixing* condition, the *same* entropy  $S(\rho_2)$  is shared by the state  $\rho_2$  on the antycloning channel leading to the *noisy* measurement process for the *flipped-qubit*  $|\Psi^\perp\rangle$ .

Note also that the present argument relies *only* on the *map-linearity* without accounting explicitly for the *non complete-positivity* of the time reversal map



**Fig. 9.8.** Demonstration of the Universal Optimal NOT-gate by single photon qubit self-injection and 4-detector coincidences. The plots on the l.h.s. and on the r.h.s. of the figure mutually correspond. The values of the fidelity of the process,  $F^*$  and the expression of the corresponding test qubits are shown as for Fig. 9.5.

$T$ , the one which is actually responsible for the flipping dynamics [3]. The presence of the quoted *unavoidable* quantum noise is shown by the experimental plots in the r.h.s. of Fig. 9.8. Furthermore, in Section 9.4 it is shown that the measurement process implied by the investigation on qubit-flipping on the AC-channel parallels almost identically the one realized to investigate cloning on the C-channel.

Very interestingly the above argument could be *reversed*, as the mixed-state condition  $S(\rho_2) > 0$  imposed by the  $T$  map on the antycloning channel is transferred to the cloning channel, in virtue of Eq.(9.22), and accounts for the unavoidable quantum noise affecting the cloning results shown in Fig. 9.7. In summary, according to Eq.(9.22), the *same* degree of *mixedness* affects *also* the output state realized on the antycloning channel, as implied by the same limit values of the signal-to-noise ratios,  $R = R^*$  affecting the experimental plots shown in Figs. 9.7 and 9.8. Since on the antycloning channel an (approximate) CP map is realized that is generally distinct from any process realized

on the cloning channel, e.g. here the *cloning* process, Eq.(9.22) appears to establish a significant *symmetry condition* in the context of axiomatic quantum theory [1,27]. In this same context it has been noted that the transformation connecting the cloning and U-NOT processes also realizes contextually the *optimal entangling* process and the *universal probabilistic quantum processor* [34].

Note in this connection that the adoption of the *classical* OPA apparatus within an investigation pushed at the extreme quantum measurement limits enlightened the conceptual significance of the quantum noise, originating in this case from the vacuum field. We recall here that the failure of an old proposal for superluminal signaling by an *amplified* Einstein-Podolsky-Rosen scheme was attributed to the impossibility of realizing a single photon *ideal* amplifier, i.e., one for which the signal-to-noise ratio would be:  $R > 2$  [6,27].

For the sake of further clarification on the *universality* concept, let's analyze the simple case of a *mode-degenerate* QIOPA amplifier in which *only* two electromagnetic modes,  $k_{\parallel}$  and  $k_{\perp}$  with the same wave length  $\lambda = 2\lambda_p$ , interact in the nonlinear crystal. In work [35] this case has been realized by a Type II BBO crystal cut for *collinear* emission over a *single* wave vector  $\vec{k}$  and the modes  $k_{\parallel}$  and  $k_{\perp}$  express two orthogonal states of polarization, e.g. of linear ( $\pi'$ ). This condition is still represented by the device shown in Fig. 9.2 in which the modes  $k_1$  and  $k_2$  are made to collapse into a single  $k$ . In general the *quantum injection* is still provided there by a qubit encoded in a single photon:  $|\Psi\rangle_{IN} \equiv (\tilde{\alpha}|1,0\rangle + \tilde{\beta}|0,1\rangle)$  where again  $|\tilde{\alpha}|^2 + |\tilde{\beta}|^2 = 1$  and:  $|x,y\rangle \equiv |x\rangle_{\parallel} \otimes |y\rangle_{\perp}$ . For the present purpose the problem may be simplified, with no loss of generality by assuming:  $\tilde{\alpha} = 1$ ,  $\tilde{\beta} = 0$ . Applying the theory given in Section 9.2, we get the output state to the first order of approximation:

$$|\Psi\rangle \approx |1,0\rangle - \sqrt{2}\Gamma|2,1\rangle \quad (9.23)$$

to be compared with Eq.(9.19). Likewise, the no-injection condition leads to the output state  $|\Psi\rangle_0 \approx |0,0\rangle - \Gamma|1,1\rangle$ , to be compared with Eq.(9.20). The "no-interaction" first term is easily discarded by a coincidence measurement. The second term consists of a *pure* state realizing both *exact* cloning and *qubit-flipping* on the output modes. This means that any test aimed at the detection of the two processes *is not affected* by quantum noise. However, the device works *just* for *one particular* input qubit, i.e. for one particular choice of the parameters  $\tilde{\alpha}$  and  $\tilde{\beta}$ . Any other choice would lead to a different amplification quantum efficiency. In other words, the device is not a *universal machine*. Referring ourselves to the theoretical results of Section 9.2, this happens because the interaction Hamiltonian of the present OPA system  $\hat{H}_I = i\hbar\chi\hat{A} + h.c.$ ,  $\hat{A} \equiv \hat{a}_{\parallel}\hat{a}_{\perp}$  cannot be made formally invariant under displacements of the injected qubit over the entire Bloch sphere (see Fig. 9.1). Indeed, at least four nonlinearly interacting electromagnetic modes are needed to attain such invariance of the optical amplifier/squeezing

Hamiltonian. All these results are fully consistent with the general theory of the optimal quantum machines [4].

In summary, the work presented in this section, and in the preceding one, shows experimentally in a unifying manner, and perhaps for the first time, that the concepts of no-signaling, i.e. causality, linearity and complete-positivity have yet little explored inter-connections within the inner structure of quantum mechanics [6, 9].

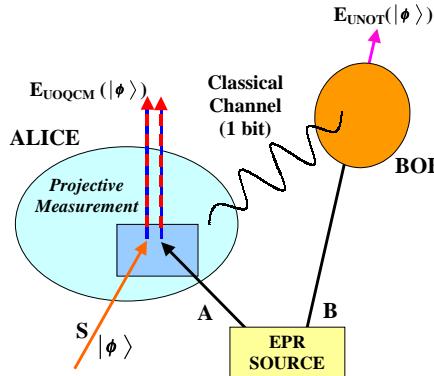
## 9.6 Teleportation scheme implementing the Universal Optimal Quantum Cloning Machine and the Universal NOT gate

In the previous sections the universal optimal quantum cloning machine and the U-NOT gate have been implemented in a somewhat intriguing “contextual” fashion by exploiting the field amplification property of the *nonlinear* quantum injected optical parametric amplifier. We may reasonably ask at this point whether there is something special in the QIOPA device or in any other nonlinear *active*, i.e. amplifying device. In the present section we show that indeed this is not the case. In fact, we shall show that the UOQCM and the U-NOT gate acting on a single qubit can be implemented very simply and generally by slightly modifying the QST protocol within a *linear* optical scheme, i.e. by using only linear optical devices [36]. Precisely, we shall consider again polarization-encoded qubits and deal with the  $1 \rightarrow 2$  cloning process.

We may recall first that the QST protocol implies that an unknown input qubit  $|\phi\rangle_S = \alpha|0\rangle_S + \beta|1\rangle_S$  is destroyed at a sending place (Alice:  $\mathcal{A}$ ) while its perfect replica appears at a remote place (Bob:  $\mathcal{B}$ ) via dual quantum and classical channels. To demonstrate that two qualitatively different resources are needed: a quantum resource that cannot be cloned, and a classical resource that cannot travel faster than light [13, 37]. Let us assume that Alice and Bob share the entangled singlet state  $|\Psi^-\rangle_{AB} = 2^{-\frac{1}{2}}(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B)$ , and that we want to teleport the generic qubit  $|\phi\rangle_S$ . The singlet state is adopted here and hereafter because its well known invariance under SU(2) transformations will ensure the universality of the cloning and U-NOT processes as we have seen in the previous sections. The overall state of the system is then  $|\Omega\rangle_{SAB} = |\phi\rangle_S|\Psi^-\rangle_{AB}$ . Alice performs a Bell measurement by projecting the qubits  $S$  and  $A$  into the four Bell states  $\{|\Psi^-\rangle_{SA}, |\Psi^+\rangle_{SA}, |\Phi^-\rangle_{SA}, |\Phi^+\rangle_{SA}\}$  and then sends the result to Bob by means of 2 bits of classical information. In order to obtain  $|\phi\rangle_B$ , Bob applies to the received state the appropriate unitary transformation  $U_B$  according to the following protocol:  $|\Psi^-\rangle_{SA} \rightarrow U_B = \mathbb{I}$ ,  $|\Psi^+\rangle_{SA} \rightarrow U_B = \sigma_Z$ ,  $|\Phi^-\rangle_{SA} \rightarrow U_B = \sigma_X$ ,  $|\Phi^+\rangle_{SA} \rightarrow U_B = \sigma_Y$  where the kets express the received corresponding information,  $I, \sigma_Z, \sigma_X$  are respectively the identity, phase flip operators and  $\sigma_Y = -i\sigma_Z\sigma_X$ . The

teleportation channel, after the unitary operation  $U_B$ , acts as the identity operator:  $E_{QST}(\rho_S) = \rho_B$  where  $\rho_S$  is the density matrix representing the state of the qubit  $S$ . In absence of the classical information channel, Bob cannot apply the appropriate  $U_B$ , and the apparatus realized the map  $E_B(\rho_S) = \frac{\mathbb{I}_B}{2}$ , corresponding to the depolarizing channel  $E_{DEP}$ . This is the worst possible information transfer because any information about the initial state  $|\phi\rangle$  is lost.

In order to implement the UOQCM and U-NOT at Alice's and Bob's sites, we modify the QST protocol by performing a different measurement on the system  $S + A$ . This leads to a different content of information to be transferred by the *classical* channel from  $\mathcal{A}$  to  $\mathcal{B}$ . Precisely, the *Bell measurement*, able to discriminate between the 4 Bell states, is replaced here by a dichotomic *projective measurement* able to identify  $|\Psi^-\rangle_{SA}$ , i.e. the *anti-symmetric* subspace of  $H \equiv H_A \otimes H_S$ , and its complementary *symmetric* subspace. Let us analyze the outcomes of such a strategy, schematically represented by Fig. 9.9.



**Fig. 9.9.** General scheme for the simultaneous realization of the Tele-UNOT Gate and of the *probabilistic* universal quantum cloning machine.

With a probability  $p = \frac{1}{4}$  the  $|\Psi^-\rangle_{SA}$  is detected by  $\mathcal{A}$ . In this case the correct QST channel  $E_{QST}$  is realized. However, if this is not the case, with probability  $p = \frac{3}{4}$ , Bob cannot apply any unitary transformation to the set of the non identified Bell states  $\{|\Psi^+\rangle_{SA}, |\Phi^-\rangle_{SA}, |\Phi^+\rangle_{SA}\}$ , and then the QST channel implements the statistical map  $E(\rho) = \frac{1}{3} [\sigma_Z \rho \sigma_Z + \sigma_X \rho \sigma_X + \sigma_Y \rho \sigma_Y]$ . As we shall see, this map coincides with the map  $E_{\text{U-NOT}}(\rho_S)$  that realizes the optimal universal NOT gate, i.e. the one that approximates *optimally* the flipping of one qubit  $|\phi\rangle$  into the orthogonal qubit  $|\phi^\perp\rangle$ , i.e.  $\rho_S$  into  $\rho_S^\perp \equiv |\phi^\perp\rangle\langle\phi^\perp|$ . Bob identifies the two different maps realized at his site by reading the information (1 bit) received by Alice on the classical channel. For example, such

*bit* can assume the value 0 if Alice identifies the Bell state  $|\Psi^-\rangle_{SA}$  and 1 if she does not. We name this process Tele-UNOT since it consists in the teleportation of an *optimal* antiunitary map acting on any input qubit, the U-NOT gate [9, 38]. In Sections 9.4 and 9.5 we have enlightened the contextuality in a bipartite entangled system of the *optimal* U-NOT gate together with an *optimal* quantum cloning process. Therefore it is worth analyzing what happens when the overall state  $|\Omega\rangle_{SAB}$  is projected onto the subspace orthogonal to  $|\Psi^-\rangle_{SA} \langle \Psi^-|_{SA} \otimes H_B$  by the projector:

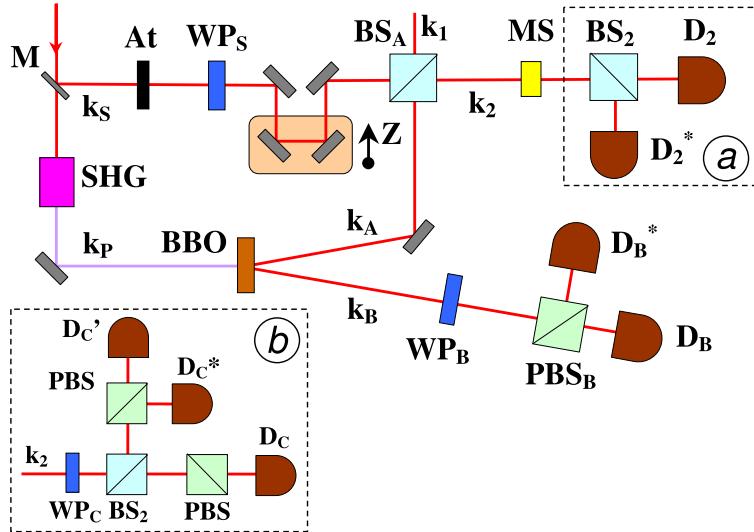
$$P_{SAB} = (\mathbb{I}_{SA} - |\Psi^-\rangle_{SA} \langle \Psi^-|_{SA}) \otimes \mathbb{I}_B. \quad (9.24)$$

This procedure generates the normalized state  $|\tilde{\Omega}\rangle \equiv P_{SAB} |\Omega\rangle_{SAB} = \sqrt{\frac{2}{3}}[|\xi_1\rangle_{SA} \otimes |1\rangle_B - |\xi_0\rangle_{SA} \otimes |0\rangle_B]$  where:  $|\xi_1\rangle_{SA} = \alpha |0\rangle_S |0\rangle_A + \frac{1}{2}\beta(|1\rangle_S |0\rangle_A + |0\rangle_S |1\rangle_A)$  and  $|\xi_0\rangle_{SA} = \beta |1\rangle_S |1\rangle_A + \frac{1}{2}\alpha(|1\rangle_S |0\rangle_A + |0\rangle_S |1\rangle_A)$ . By tracing this state over the *SA* and *B* manifolds we get:  $\rho_{SA} \equiv \text{Tr}_B |\tilde{\Omega}\rangle \langle \tilde{\Omega}| = \frac{2}{3} |\phi\rangle \langle \phi|_{SA} \langle \phi|_{SA} + \frac{1}{3} |\{\phi, \phi^\perp\}\rangle_{SA} \langle \{\phi, \phi^\perp\}|_{SA}$  and  $\rho_B^{out} \equiv \text{Tr}_{SA} |\tilde{\Omega}\rangle \langle \tilde{\Omega}| = \frac{1}{3}(2\rho_S^\perp + \rho_S)$ , where:  $|\{\phi, \phi^\perp\}\rangle_{SA} = 2^{-\frac{1}{2}}(|\phi^\perp\rangle_S |\phi\rangle_A + |\phi\rangle_S |\phi^\perp\rangle_A)$ . We further project on the spaces *A* and *B* and obtain the output states, which are found mutually equal:  $\rho_S^{out} \equiv \text{Tr}_A \rho_{SA} = \frac{1}{6}(5\rho_S + \rho_S^\perp) = \rho_A^{out} \equiv \text{Tr}_S \rho_{SA}$ . At last, by these results the expected *optimal* values for the fidelities of the two “forbidden” processes are obtained:  $F_{CLON} = \text{Tr}[\rho_S^{out} \rho_S] = \text{Tr}[\rho_A^{out} \rho_S] = \frac{5}{6}$  and  $F_{UNOT} = \text{Tr}[\rho_B^{out} \rho_S^\perp] = \frac{2}{3}$  [9, 27].

We want also to emphasize that it is possible to obtain the same results by performing a projective measurement on different Bell states. It's easily verified that by projection along  $(\mathbb{I}_{SA} - |\Phi^+\rangle_{SA} \langle \Phi^+|_{SA}) \otimes \mathbb{I}_B$ , the teleportation channel realizes the optimal universal approximation of the transpose map. This map is related to the U-NOT map by a unitary transformation:  $E_{\text{TRANSPPOSE}} = E_{\sigma_Y}(E_{\text{U-NOT}}(\rho))$ , where  $E_{\sigma_Y}(\rho) = \sigma_Y \rho \sigma_Y$ . It is straightforward to verify that the cloning machine, in this case, generates  $\rho'_S = \rho'_A = \sigma_Y \rho_S \sigma_Y$ , that is, the clones of the input qubit on which acts the unitary transformation  $\sigma_Y$ . Similarly, by choosing another Bell state in (9.24), analogue transformations are achieved. Of course, in this case the universality of the processes is retrieved by applying an adequate unitary transformation to the output states of qubits labelled by *SA* and *B*.

In the past years many experimental realizations of quantum state teleportation have been achieved [39]. Recently the first “active”, i.e. *complete* version of the QST protocol was realized by our laboratory by physically implementing the Bob's unitary operation [40]. In the experiment described here the input qubit was codified as the *polarization* state of a single photon belonging to the input mode  $k_S$ :  $|\phi\rangle_S = \alpha |H\rangle_S + \beta |V\rangle_S$ ,  $|\alpha|^2 + |\beta|^2 = 1$  (see Fig. 9.10).

Here  $|H\rangle$  and  $|V\rangle$  correspond to the horizontal and vertical polarizations, respectively. In addition, an entangled pair of photons, *A* and *B* were gen-



**Fig. 9.10.** Setup for the optical implementation of the *Tele-UNOT gate* and the *probabilistic UOQCM*. The measurement setup used for the verification of the cloning experiment is reported in the inset (b).

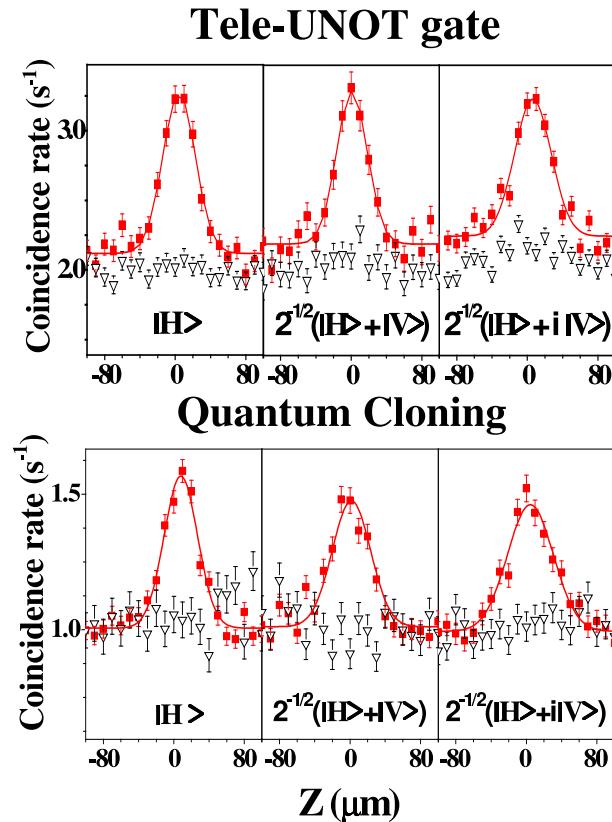
erated on the modes  $k_A$  and  $k_B$  by SPDC in the *singlet* state:  $|\Psi^-\rangle_{AB} = 2^{-1/2} (|H\rangle_A |V\rangle_B - |V\rangle_A |H\rangle_B)$ . The *projective measurement* in the space  $H = H_A \otimes H_S$  was realized by linear superposition of the modes  $k_S$  and  $k_A$  on a 50 : 50 *beam-splitter*,  $BS_A$ . Consider the overall output state realized on the two output modes  $k_1$  and  $k_2$  of  $BS_A$  and expressed by a linear superposition of the states:  $\{|\Psi^-\rangle_{SA}, |\Psi^+\rangle_{SA}, |\Phi^-\rangle_{SA}, |\Phi^+\rangle_{SA}\}$ . As it is well known, the realization of the singlet  $|\Psi^-_{SA}\rangle$  is identified by the emission of one photon on each output mode of  $BS_A$  while the realization of the set of the other three Bell states implies the emission of 2 photons either on mode  $k_1$  or on mode  $k_2$ . The realization of the last process, sometimes dubbed as “Bose Mode Coalescence” (BMC) or “mode-occupation enhancement” [41] was experimentally identified by the simultaneous clicking of the detectors  $D_2$  and  $D_2^*$  coupled to the output mode  $k_2$  by the 50 : 50 beam-splitter  $BS_2$ . The identical effect expected on mode  $k_1$  was not exploited, for simplicity. As shown by the theoretical analysis above, the last condition implied the simultaneous realization in our experiment of the U-NOT and UOQCM processes, here detected by a *post-selection* technique. Interestingly enough, the *symmetry* of the projected subspace of  $H$  identified by BMC is implied by the intrinsic *Bose symmetry* of the 2 photon Fock state realized at the output of  $BS_A$ .

The source of the SPDC process was a Ti:Sa mode-locked pulsed laser with wave length  $\lambda = 795\text{nm}$  and repetition rate  $76\text{MHz}$ . A weak beam, deflected from the laser beam by mirror  $M$ , strongly attenuated by grey

filters ( $At$ ), delayed by  $Z = 2c\Delta t$  via a micrometrically adjustable optical “trombone” was the source of the *quasi* single-photon state injected into  $BS_A$  over the mode  $k_s$ . The average number of injected photons was  $\bar{n} \simeq 0.1$  and the probability of a spurious 2 photon injection was evaluated to be a factor 0.05 lower than for single photon injection. Different qubit states  $|\phi\rangle_S$  were prepared via the optical wave plate  $WP_S$ , either a  $\lambda/2$  or  $\lambda/4$  wave plate. The main UV laser beam generated by second harmonic generation, focused into a nonlinear crystal, excited the SPDC source of the *singlet*  $|\Psi^-\rangle_{AB}$ . The photons  $A$  and  $B$  of each entangled pair were emitted over the modes  $k_A$  and  $k_B$  with equal wls  $\lambda = 795nm$ . All adopted photodetectors ( $D$ ) were equal SPCM-AQR14 single photon counters. One interference filter with bandwidth  $\Delta\lambda = 3nm$  was placed in front of each  $D$  and determined the coherence time of the optical pulses:  $\tau_{coh} \simeq 350fs$ .

In order to realize the *Tele – UNOT* protocol, the BMC process on the output mode  $k_2$  of  $BS_A$  was detected by a coincidence technique involving  $D_2$  and  $D_2^*$  [Fig. 9.10, inset: (a)]. In order to enhance the *visibility* of the Ou-Mandel interference at the output of  $BS_A$ , a  $\pi$ -preserving mode-selector ( $MS$ ) was inserted on mode  $k_2$ . On the Bob’s site, the polarization  $\pi$ –*state* on the mode  $k_B$  was analyzed by the combination of the wave plate  $WP_B$  and of the *polarization beam splitter*  $PBS_B$ . For each input  $\pi$ –*state*  $|\phi\rangle_B$ ,  $WP_B$  was set in order to make the  $PBS_B$  to transmit  $|\phi\rangle_B$  and to reflect  $|\phi^\perp\rangle_B$ , by then exciting  $D_B$  and  $D_B^*$  correspondingly. First consider QST turned off, by setting the optical delay  $|Z| \gg c\tau_{coh}$  i.e. by spoiling the interference of photons  $S$  and  $A$  in  $BS_A$ . In this case, since the states  $|\phi\rangle_B$  and  $|\phi^\perp\rangle_B$  were realized with the same probability on mode  $k_B$ , the rate of coincidences detected by the  $D$ -sets  $[D_B, D_2, D_2^*]$  and  $[D_B^*, D_2, D_2^*]$  were expected to be equal. By turning on the QST, i.e. by setting  $|Z| \ll c\tau_{coh}$ , the output state  $\rho_B^{out} = (2\rho_S^\perp + \rho_S)/3$  was realized implying a factor  $R = 2$  *enhancement* of the counting rate  $[D_B^*, D_2, D_2^*]$  and *no enhancement* of  $[D_B, D_2, D_2^*]$ . The actual measurement of  $R$  was carried out, and the *universality* of the *Tele – UNOT* process demonstrated, by the experimental results shown in Fig. 9.11. These 3-coincidence results, involving the sets  $[D_B, D_2, D_2^*]$  and  $[D_B^*, D_2, D_2^*]$  correspond to the injection of three different input states:  $|\phi\rangle_S = |H\rangle$ ,  $|\phi\rangle_S = 2^{-1/2}(|H\rangle + |V\rangle)$ ,  $|\phi\rangle_S = 2^{-1/2}(|H\rangle + i|V\rangle)$ . In Fig. 9.11 the square and triangular markers refer respectively to the  $[D_B^*, D_2, D_2^*]$  and  $[D_B, D_2, D_2^*]$  coincidences versus the delay  $Z = 2c\Delta t$ . We may check that the *Tele – UNOT* process only affects the  $|\phi^\perp\rangle_B$  component, as expected. The *signal-to-noise* ( $S/N$ ) ratio  $R$  was determined as the ratio between the peak values, i.e. for  $Z \simeq 0$ , and the no BMC enhancement values, i.e. for  $|Z| \gg c\tau_{coh}$ . The experimental values of the *U – NOT fidelity*  $F = R(R+1)^{-1}$  are:  $F_H = 0.641 \pm 0.005$ ;  $F_{H+V} = 0.632 \pm 0.006$ ;  $F_{H+iV} = 0.619 \pm 0.006$ , for the three injection states  $|\phi\rangle_S$ . These results, to be compared with the *optimal* value  $F_{th} = 2/3 \approx 0.666$  corresponding to the *optimal*  $R = 2$ , have been evaluated by taking into account the reduction, by a factor  $\xi = 0.7$ , of the coincidence

rate due to the spurious simultaneous injection of two photons on the mode  $k_S$  and to the simultaneous emission of two SPDC pairs. The factor  $\xi$  was carefully evaluated by a side experiment involving the detectors  $D_2$  and  $D_2^*$ . Note that the experimental Tele-UNOT peaks shown in Fig. 9.11 indeed demonstrate the simultaneous, *contextual* realization of the quantum cloning and U-NOT processes on Alice's and Bob's sites, respectively.



**Fig. 9.11.** Experimental results of the *Tele-UNOT gate* and the UOQCM for three input qubits. *Filled squares*: plots corresponding to the “correct” polarization; *Open triangles*: plots corresponding to the “wrong” polarization. The solid line represents the best gaussian fit expressing the *correct* polarization.

For the sake of completeness, we wanted to gain insight into the linear *probabilistic* UOQCM process by investigating whether at the output of  $BS_A$  the photons  $S$  and  $A$  were indeed left in the state  $\rho_S^{out} = (5\rho_S + \rho_S^\perp)/6 = \rho_A^{out}$  after the state projection. The cloning analysis was realized on the  $BS_A$  output mode  $k_2$  by replacing the measurement set (a) with the (b) one, shown in the inset of Fig. 9.10. The polarization state on mode  $k_2$  was

analyzed by the combination of the wave plate  $WP_C$  and of the polarizer beam splitters  $PBS$ . For each input  $\pi$ -state  $|\phi\rangle_S$ ,  $WP_C$  was set in order to make  $PBS$ 's to transmit  $|\phi\rangle_S$  and to reflect  $|\phi^\perp\rangle_S$ . The cloned state  $|\phi\phi\rangle_S$  was detected on mode  $k_2$  by a coincidence between the detectors  $D_C$ ,  $D'_C$ . The generation of an entangled pair was assured by detecting one photon on the mode  $k_B$ ; in this case  $PBS_B$  was removed and the field of mode  $k_B$  was coupled directly to  $D_B$ . Any coincidence detected by the sets  $[D_C, D'_C, D_B]$  and  $[D_C, D'_C, D_B]$  implied the realization of the states  $|\phi\phi\rangle_S$  and  $|\phi\phi^\perp\rangle_S$ , respectively. In analogy with the previous experiment, when  $|Z| \gg c\tau_{coh}$  the rate of coincidences detected by  $[D_C, D'_C, D_B]$  and  $2 \times [D'_C, D'_C, D_B]$  were expected to be equal. By turning on the cloning machine,  $|Z| \ll c\tau_{coh}$ , an *enhancement* by a factor  $R = 2$  of the counting rate by  $[D_C, D'_C, D_B]$  and *no enhancement* by  $[D_C, D'_C, D_B]$  were expected. The experimental results of the  $S/N$  ratio  $R$ , carried out by coincidence measurements involving  $[D_C, D'_C, D_B]$  and  $[D_C, D'_C, D_B]$  are reported in the lower plots of Fig. 9.11, again for the three different input states:  $|\phi\rangle_S = |H\rangle$ ,  $|\phi\rangle_S = 2^{-1/2}(|H\rangle + |V\rangle)$ ,  $|\phi\rangle_S = 2^{-1/2}(|H\rangle + i|V\rangle)$ . The square and triangular markers there refer respectively to the  $[D_C, D'_C, D_B]$  and  $[D_C, D'_C, D_B]$  coincidence plots Vs the delay  $Z$ . The following values of the *cloning fidelity*  $F = (2R+1)(2R+2)^{-1}$  were found:  $F_H = 0.821 \pm 0.003$ ;  $F_{H+V} = 0.813 \pm 0.003$ ;  $F_{H+iV} = 0.812 \pm 0.003$ , to be compared with the *optimal*  $F_{th} = 5/6 \approx 0.833$  corresponding to the limit  $S/N$  value:  $R = 2$ . As for the *Tele-UNOT* gate experiment, the factor  $\xi = 0.7$  has been corrected during the evaluation of the fidelities. Finally note that, in the present context the entangled *singlet* state  $|\Psi^-\rangle_{AB}$  was not strictly necessary for the sole implementation of quantum cloning as we could model the *local* effect of the *singlet* on the input mode  $k_A$  by a *fully mixed* state  $\rho_A = \frac{1}{2}\mathbb{I}_A$  spanning a two dimensional space. This has been realized successfully in another experiment [36].

In summary, two relevant quantum information processes, forbidden by quantum mechanics in their exact form are found to be connected *contextually* by a modified quantum state teleportation scheme and can be optimally realized. At variance with previous experiments, the complete implementation of the new protocol has been successfully performed by a fully *linear* optical setup. The results are found in full agreement with theory.

## 9.7 Quantum Entanglement and Tomographic analysis of Quantum Operations

The reported work on quantum cloning and on the related *optimal* and *universal* transformations involving any quantum state estimation can be considered the most significant modern contribution to the science of Quantum Measurement. We have seen in the previous sections the fundamental unifying role of quantum entanglement – the basis of the quantum parallelism of computers of the future – as establishing an unexpected, and as yet not

well understood, *contextuality network* among the optimally realized non-unitary operations. Since the apparent significance of all that relies on the as yet unexplored character of quantum entanglement, it appears fit to present here a new method of tomography of quantum operations, dubbed a “Pauli tomography” in a two dimensional Hilbert space. As we shall see, by cleverly exploiting the total parallelism of the bipartite entanglement associated with a pair of SPDC correlated photons, the full information characterizing an unknown quantum operator, i.e. an active device, can be retrieved easily and efficiently. In the framework of the new information technology [42] the relevance of the new method is indeed motivated by the need of characterizing any quantum devices making a complete identification and characterization of their functioning. We present here the first full experimental quantum characterization of a single-qubit device.

How do we usually characterize the operation of a device? Actually, we are interested just in linear devices, since quantum dynamics is intrinsically linear. Any linear device, either quantum or classical (examples are: an optical lens or a good amplifier), can be completely described by a *transfer matrix* which gives the output vector by matrix-multiplying the input vector. In quantum mechanics the inputs are density operators  $\rho_{in}$  and the role of the transfer matrix is played by the so called *quantum operation* [1] of the device, that here we will denote by  $E$ . Thus the output state  $\rho_{out}$  is given by the quantum operation applied to the input state as follows:

$$\rho_{out} = \frac{E(\rho_{in})}{Tr[E(\rho_{in})]} \quad (9.25)$$

and the normalization constant  $Tr[E(\rho_{in})]$  is also the probability of occurrence of the transformation  $E$ , when there are other possible alternatives, such as when we consider the state transformation due to a measuring device for a given outcome.

Now the problem is: how to reconstruct the form of  $E$  experimentally? One would be tempted to adopt the conventional method [42] of running a *basis* of all possible inputs, and measuring the corresponding outputs by *quantum tomography* [43]. However, since the states  $\rho$  are actually operators, not vectors, in order to get all possible matrix elements we would need to run a complete orthogonal basis of quantum states  $|n\rangle$  along with their linear combination  $2^{\frac{1}{2}}(|n'\rangle + i^k|n''\rangle)$ , with  $k = 0, 1, 2, 3$  and  $i$  denoting the imaginary unit. This is a simple consequence of the polarization identity. However, the availability of such a set of states in the laboratory is, by itself, a very hard technological problem since states with a precise varying number of photons and, even worst, their superposition, are still a dream for experimentalists.

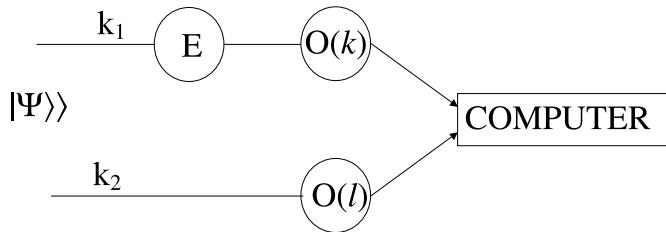
The quantum parallelism intrinsic of entanglement now comes to help us, running all possible input states in parallel by using only a single entangled state as the input! This was first shown in [44]. Hence, we don't need to prepare a complete set of states, but just one entangled state, a state commonly available in modern quantum optical laboratories!

### 9.7.1 Pauli tomography

Assume for simplicity, and with no loss of generality, that the entangled state spans a 4-dimensional Hilbert space  $H_1 \otimes H_2$ , each space  $H_i$  of dimension  $d = 2$ . The input entangled state,

$$|\Psi\rangle\rangle = \sum_{nm} \Psi_{nm} |nm\rangle \quad (9.26)$$

expressed in terms of the basis vectors  $|n\rangle \otimes |m\rangle = |nm\rangle$  of the two spaces  $H_i$  ( $i = 1, 2$ ) is used as shown in Fig. 9.12 where the entangled system consists of two single-mode optical beams.



**Fig. 9.12.** General experimental scheme of the method for the tomographic estimation of the quantum operation of a single qubit device. Two identical quantum systems, e.g. two optical beams as in the present experiment, are prepared in an entangled state  $|\Psi\rangle\rangle$ . One of the systems undergoes the quantum operation  $E$ , whereas the other is left untouched. At the output one makes a quantum tomographic estimation, by measuring jointly two observables from a *quorum*  $\{O(l)\}$ . In the present experiment the quorum is represented by the set of Pauli operators.

The key feature of the method implies that only one of the two systems, say system  $i = 1$ , is input into the *unknown* device  $E$ , whereas the other is left untouched. This setup leads to the output state  $R_{out}$ , which in tensor notation writes as follows

$$R_{out} = E \otimes I |\Psi\rangle\rangle \langle\langle \Psi| \quad (9.27)$$

where  $I$  denotes the identical operation. It is a result of linear algebra that  $R_{out}$  is in one-to-one correspondence with the quantum operation  $E$ , as long as the state  $|\Psi\rangle\rangle$  is full-rank, i.e. it has non-vanishing components on the whole state-space of each system, such as, for instance, a so called maximally entangled state. Full-rank entangled states can be easily generated by spontaneous parametric down conversion of the vacuum state, as in the experiment reported here. Note that by this method the problem of availability of all possible input states is solved: we just need a single entangled state  $|\Psi\rangle\rangle$ , which works as all possible inputs in a sort of quantum parallelism!

Now, how to characterize the entangled state  $R_{out}$  at the output? We obviously need to perform many measurements on an ensemble of equally prepared quantum systems, since, due to the *no-cloning theorem* [5] we cannot determine the state of a single system [45]. For this purpose a technique for the full determination of the quantum state has been introduced and developed since 1994. The method named *quantum tomography* [43] has been initially introduced for the state of a single-mode of radiation, the so called *homodyne tomography*, and thereafter it has been generalized to any quantum system. The basis of the method is just performing measurements of a suitably complete set of observables called *quorum*. For our needs, we just have to measure jointly a quorum of observables, here the spin observables  $\sigma_i$  ( $i = 0, 1, 2, 3$ ), on the two entangled systems at the output, in order to determine the output state  $R_{out}$ , and hence the quantum operation  $E$ .

In this section we present the first complete experimental characterization of a quantum device, which in our case will be a single-qubit device. We shall refer to this protocol as “Pauli tomography”. The qubit is encoded on polarization of single photons in the following way

$$|0\rangle = |1\rangle_h |0\rangle_v, |1\rangle = |0\rangle_h |1\rangle_v$$

namely with the “logical zero” state corresponding to a single horizontally polarized photon and the “logical one” state corresponding to a single vertically polarized photon. In the following we will denote by  $h$  and  $v$  the annihilation operators of the horizontally and vertically polarized modes of radiation associated to a fixed wave vector,  $\mathbf{k}$ . Using single photon states we encode a qubit on the polarization. In the polarization representation, the Pauli matrices write as follows:

$$\sigma_1 = h^\dagger v + v^\dagger h, \quad \sigma_2 = i(h^\dagger v - v^\dagger h), \quad \sigma_3 = h^\dagger h - v^\dagger v$$

The ring of Pauli matrices is completed by including the identity  $\sigma_0 = h^\dagger h + v^\dagger v$ . In the following we will denote by  $\boldsymbol{\sigma}$  the column three-vector of operators  $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ , and by  $\boldsymbol{\sigma}$  the column tetra-vector  $\boldsymbol{\sigma} = (\sigma_0, \sigma_1, \sigma_2, \sigma_3)$ , and use Greek indices for three-vectors components  $\alpha = 1, 2, 3$  (or  $\alpha = x, y, z$ ), and Latin indices for tetra-vector components:  $i = 0, 1, 2, 3$ .

A wave plate changes the two radiation modes according to the matrix transformation:

$$\begin{pmatrix} h \\ v \end{pmatrix} \longrightarrow w(\phi, \theta)^\dagger \begin{pmatrix} h \\ v \end{pmatrix} w(\phi, \theta) \equiv W(\phi, \theta) \begin{pmatrix} h \\ v \end{pmatrix} \quad (9.28)$$

where the matrix  $W(\phi, \theta)$  is given by:

$$\begin{aligned} W(\phi, \theta) &= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \\ &= \begin{pmatrix} z_+ + cz_- & sz_- \\ sz_- & z_+ - cz_- \end{pmatrix} \end{aligned} \quad (9.29)$$

where  $s = \sin 2\theta$ ,  $c = \cos 2\theta$ ,  $\theta$  is the wave plate orientation angle around the wave vector  $\mathbf{k}$ ,  $z_{\pm} = \frac{1}{2}(1 \pm e^{i\phi})$ ,  $\phi = \frac{2\pi\delta}{\lambda}$ ,  $\lambda$  is the wave length and  $\delta$  is the optical path through the plate. Special cases are the  $\frac{\lambda}{4}$  plate that can be used with  $\theta = \frac{\pi}{4}$  to give the right and left circularly polarized *modes*

$$\begin{pmatrix} r \\ l \end{pmatrix} = W\left(\frac{\pi}{2}, \frac{\pi}{4}\right) \begin{pmatrix} h \\ v \end{pmatrix} = 2^{-\frac{1}{2}} e^{i\pi/4} \begin{pmatrix} h + iv \\ -ih + v \end{pmatrix} \quad (9.30)$$

and the  $\frac{\lambda}{2}$  plate that can be used to give the diagonal linearly polarized *modes*

$$\begin{pmatrix} a \\ b \end{pmatrix} = W\left(\pi, \frac{\pi}{8}\right) \begin{pmatrix} h \\ v \end{pmatrix} = 2^{-\frac{1}{2}} \begin{pmatrix} h + v \\ h - v \end{pmatrix}. \quad (9.31)$$

The Heisenberg picture evolution of the Pauli matrices corresponding to the unitary  $U$  on the qubit Hilbert space will be given by a rotation, which we will denote as follows

$$U^\dagger \boldsymbol{\sigma} U = R(U) \boldsymbol{\sigma}. \quad (9.32)$$

For a  $\phi$ -wave plate we have the rotation matrix

$$R(W(\phi, \theta)) = \begin{pmatrix} s^2 + c^2 \cos \phi & -c \cos \phi & sc(1 - \cos \phi) \\ c \sin \phi & \cos \phi & -s \sin \phi \\ sc(1 - \cos \phi) & s \sin \phi & c^2 + s^2 \cos \phi \end{pmatrix} \quad (9.33)$$

In particular, for a  $\frac{\lambda}{2}$  wave plate we have

$$W(\pi, \theta) = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$$

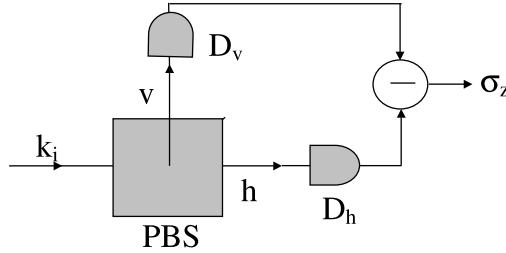
$$R(\pi, \theta) = \begin{pmatrix} -\cos 4\theta & 0 & \sin 4\theta \\ 0 & -1 & 0 \\ \sin 4\theta & 0 & \cos 4\theta \end{pmatrix} \quad (9.34)$$

The  $\sigma_z$ -photodetector is achieved as in Fig. 9.13. From Eq.(9.33) we can see that a  $\sigma_x$ -detector can be obtained by preceding the  $\sigma_z$ -detector with a  $\lambda/2$  wave plate oriented at  $\theta = \pi/8$ , whereas a  $\sigma_y$ -detector is obtained by preceding the  $\sigma_z$ -detector with a  $\lambda/4$  wave plate oriented at  $\theta = \pi/4$ . When collecting data at a  $\sigma_\alpha$ -detector, we will denote by  $\sigma_\alpha = \pm 1$  the random outcome, with  $\sigma_\alpha = \pm 1$  corresponding to the h-detector flashing, and  $\sigma_\alpha = -1$  corresponding to the v-detector flashing instead.

The experimental averages for the complete setup must coincide with the following theoretical expectation values

$$\overline{s_i^{(1)} s_j^{(2)}} = \langle \langle \Psi | (U^\dagger \otimes I)(\sigma_i^{(1)} \otimes \sigma_j^{(2)})(U \otimes I) | \Psi \rangle \rangle \quad (9.35)$$

and, in particular,  $\overline{s_i^{(1)}} \equiv \overline{s_i^{(1)} s_0^{(2)}}$  and  $\overline{s_i^{(2)}} \equiv \overline{s_0^{(1)} s_2^{(2)}}$  now  $s_i^{(n)}$  denoting the random outcome of the detector of the  $n$ th beam ( $n = 1, 2$ ) in the entangled



**Fig. 9.13.** Pauli-matrix measurement apparatus for photon polarization qubits inserted at the end of each test optical beam.

state. For maximally entangled states we have also  $\overline{s_\alpha^{(1)}} = \overline{s_\alpha^{(2)}} = 0$  for all  $\alpha = x, y, z$ . The theoretical expectations for the setup without the device to be characterized are given by:

$$\langle\langle\Psi|\sigma_i^{(1)} \otimes \sigma_j^{(2)}|\Psi\rangle\rangle = Tr [\Psi^+ \sigma_i \Psi \sigma_j^*] \quad (9.36)$$

Where  $\Psi$  denotes the matrix of the state on the customary basis of eigenvectors of  $\sigma_z$ . In particular, for the four Bell states

$$2^{-\frac{1}{2}} |\sigma_j\rangle\rangle = 2^{-\frac{1}{2}} \sigma_j \otimes I |I\rangle\rangle \quad (9.37)$$

we have

$$2^{-\frac{1}{2}} \Delta_{ij}(\sigma_k) = \delta_{ij} H_{kj}, \quad H = \begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix} \quad (9.38)$$

Using Eq.(9.32) and (9.36) we obtain

$$\overline{s_\alpha^1 s_\beta^2} = \sum_\lambda R_{\alpha\lambda}(U) \Delta_{\lambda\beta}(\Psi) \quad (9.39)$$

In particular, in the lab we use the multiplet state corresponding to  $\Psi = 2^{-\frac{1}{2}} \sigma_x$ , whence we have

$$\overline{s_\alpha^1 s_\beta^2} = R_{\alpha\beta}(U) \Delta_{1\beta}(\Psi)$$

### 9.7.2 Method, Apparatus and Results

The SPDC was adopted to generate on the two modes  $\mathbf{k}_i$   $i = 1, 2$ , single photon pairs in a polarization entangled triplet state, viz.  $|\Psi\rangle\rangle = 2^{-1/2} |\sigma_x\rangle\rangle$ , according to Eq.(9.37). The measurement apparatus consisted of two equal polarizing beam splitters  $PBS_i$  with output modes coupled to four equal Si-avalanche photo-detectors SPCM-AQR14. The beams exciting the detectors

were filtered by equal interference filters within a bandwidth  $\Delta\lambda = 3nm$ . The detector output signals were finally analyzed by a computer. We now want to determine experimentally by this apparatus the matrix elements of the state  $|\Psi\rangle\rangle$  expressed by Eq.(9.26). This can be achieved as follows: from the trivial identity

$$\langle nm| \Psi\rangle\rangle = \Psi_{nm} \quad (9.40)$$

we obtain the matrix  $\Psi_{nm}$  for the *input states* in terms of the following ensemble averages

$$\Psi_{nm} = e^{i\varphi} \frac{\langle\langle \Psi | 01 \rangle \langle nm | \Psi \rangle\rangle}{\sqrt{\langle\langle \Psi | 01 \rangle \langle 01 | \Psi \rangle\rangle}} \quad (9.41)$$

where the unmeasurable phase factor is given by:  $\exp(i\varphi) = \Psi_{01}/|\Psi_{01}|$ . The choice of the vector  $|01\rangle$  is arbitrary as it is needed only for the sake of normalization: e.g. we could have used  $|10\rangle$  or  $|11\rangle$ , instead. Using the *tomographic expansion* over the four Pauli matrices [43], [44] we see that, in virtue of Eq.(9.41), the matrix element of the input state is obtained from the following experimental averages:

$$\Psi_{nm} = \frac{1}{4\sqrt{p}} \sum_{ij} Q_{ij}(nm) \overline{s_i^{(1)} s_j^{(2)}} \quad (9.42)$$

where

$$p = \langle\langle \Psi | 01 \rangle \langle 01 | \Psi \rangle\rangle = \frac{1}{4} \left(1 + s_3^{(1)}\right) \left(1 - s_3^{(2)}\right) \quad (9.43)$$

is the fraction of coincidences with both  $\sigma_z$ -detectors firing on h, and the matrix  $Q(nm)$  is given by

$$Q_{ij}(nm) = \langle n | \sigma_i | 0 \rangle \langle m | \sigma_j | 1 \rangle \quad (9.44)$$

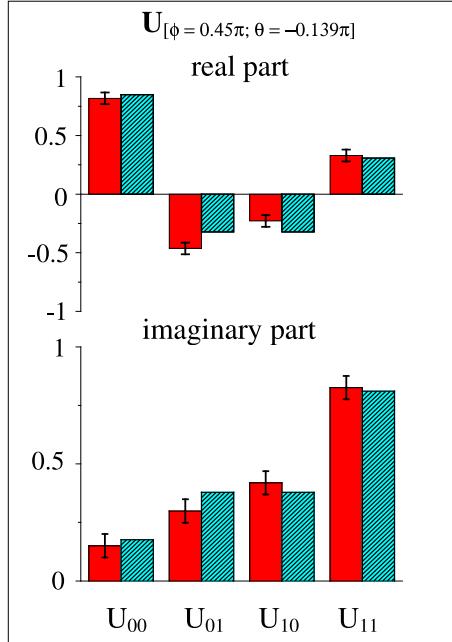
the unitary matrix  $U_{nm}$  of the device is obtained with the same averaging above but for the state at the output of the device:  $|U\Psi\rangle\rangle = (U \otimes I)|\Psi\rangle\rangle$ . Therefore, we now have:

$$(U\Psi)_{nm} = e^{i\varphi} \frac{\langle\langle U\Psi | 01 \rangle \langle nm | \Psi U \rangle\rangle}{\sqrt{\langle\langle U\Psi | 01 \rangle \langle 01 | \Psi U \rangle\rangle}} \quad (9.45)$$

where we use again Eqs.(9.43), (9.44), but the average expressed by (9.41) is carried out over the output state  $|U\Psi\rangle\rangle$ . The (complex) parameters  $U_{nm}$  are obtained from Eq.(9.45) by matrix inversion. This is of course possible since the matrix  $\Psi$  is invertible, in virtue of the entangled character of  $|\Psi\rangle\rangle$ .

The experimental demonstration of the tomographic process is given in Figs. 9.14 and 9.15 where both *real* and *imaginary parts* of the four components of the matrix  $U$  are reported for two different unknown devices inserted in the mode  $\mathbf{k}_1$  of the tomographic apparatus.

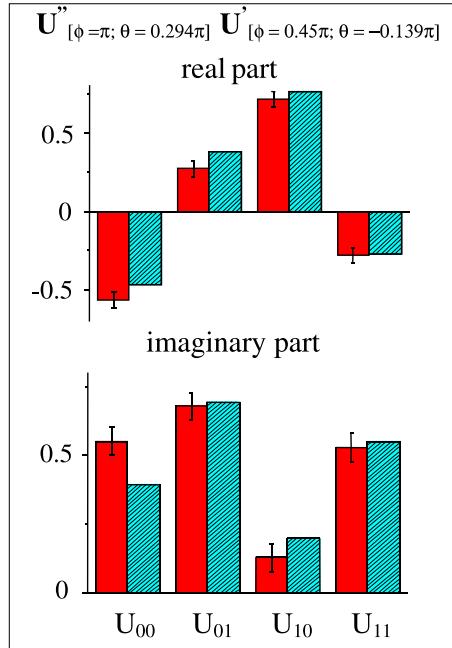
The experimental results are shown together with the corresponding data evaluated theoretically. Furthermore the experimental variance of the data



**Fig. 9.14.** Experimental characterization by Pauli tomography of a single optical  $\lambda/2$  wave plate  $[\varphi = \pi; \theta = +0.249\pi]$  inserted on channel  $k_1$  with the following optical properties: retardation phase  $\varphi = (\pi)$ ; orientation angle of the optical axis respect to the laboratory horizontal direction “h”:  $\theta = (0.249\pi)$ . The experimental real and imaginary parts of the four matrix elements  $U_{ij}$  of the wave plate are shown together with the related measured statistical variances. The corresponding theoretical values are shown for comparison.

are also reported. Each unknown device is represented in one of the figures as follows. In Fig. 9.14 a single  $\lambda/2$  wave plate  $[\varphi = \pi; \theta = +0.249\pi]$  i.e., with retardation phase  $\varphi = (\pi)$  and orientation angle in respect to the horizontal direction “h” :  $\theta = (0.249\pi)$ . In Fig. 9.15 a combination of two wave plates: i.e. a wave plate  $[\varphi = 0.45\pi; \theta = -0.138\pi]$  followed by a  $\lambda/2$  wave plate  $[\varphi = \pi; \theta = +0.29\pi]$ . As we may see, the experimental results are found in good agreement with theory [46]. In summary, we have given the first demonstration, in a simple single-qubit context, of a novel tomographic method that is able to fully characterize the properties of any device acting on a quantum system by exploiting for the first time the complete intrinsic parallelism of the quantum entanglement. This method establishes a new fundamental framework of utterly paradigmatic relevance in the domains of modern quantum measurement theory and quantum information.

Our method is expected to be of general and far reaching relevance. In fact, it can be adopted within more general and complex multi-qubit systems. For instance by this method a full characterization of a two-qubits device,



**Fig. 9.15.** Experimental characterization by Pauli tomography of a combination of two optical wave plates: the wave plate  $[\varphi = 0.45\pi; \theta = -0.138\pi]$  (cfr. Fig. 9.14) followed by a  $\lambda/2$  wave plate  $[\varphi = \pi; \theta = +0.29\pi]$ . The experimental real and imaginary parts of the four matrix elements  $U_{ij}$  of the combination are shown together with the related measured statistical variances. The corresponding theoretical values are shown for comparison

such as a controlled-NOT, can be achieved. In this case we just need to double the input and the measurement setup, by providing two input entangled states and four detectors coupled at the outputs of the two pairs of the device channels. The full quantum characterization of the device is finally obtained by a joint tomographic reconstruction on both channels of the device.

## Acknowledgments

We are indebted with Vladimir Buzek, Sandu Popescu, Christoph Simon, Andrea Mazzei, Carlo Sias, Paolo Lo Presti and Mauro D'Ariano for enlightening discussions and suggestions. This work has been supported by the FET European Network on Quantum Information and Communication (Contract IST-2000-29681: ATESIT), by PRA-INFM “CLON” and by PAIS-INFM 2002 (QEUPCO).

## References

1. K. Kraus: *States, Effects and Operations* (Springer-Verlag, Berlin, 1983); J. Preskill: Lecture Notes on Quantum Computation, <http://www.theory.caltech.edu/people/ph229/#lecture>. A *completely-positive map* (CP-map)  $\Lambda(\rho)$  preserves *positivity* for: (a) any *local* state in the Hilbert space  $H$ , (b) when *tensor-multiplied* with the identical map  $\mathbf{I}$  acting on *any* Hilbert space  $K$ , the extended map  $\Lambda(\rho) \otimes \mathbf{I}$  is positive for any state in the entangled space  $H \otimes K$ , for *any* extension of  $K$ . A *positive map* (P-map) only satisfies property (a).
2. A. S. Holevo: *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982)
3. S. Massar and S. Popescu: *Phys. Rev. Lett.* **74**, 1259 (1995); N. Gisin and S. Popescu: *Phys. Rev. Lett.* **83**, 432 (1999); R.G. Sachs: *The Physics of Time Reversal* (U. of Chicago Press, 1987).
4. R. Derka, V. Buzek and A. Ekert: *Phys. Rev. Lett.* **80**, 1571 (1998); V. Buzek, M. Hillery, R.F. Werner: *Phys. Rev. A* **60**, R2626 (1999); R.F. Werner: *Phys. Rev. A* **58**, 1827 (1998).
5. W.K. Wootters and W.K. Zurek: *Nature* (London) **299**, 802 (1982).
6. N. Herbert: *Found. Phys.* **12**, 1171 (1982). The first precise formulation of the no-cloning theorem based on the linear structure of quantum theory was indeed expressed in 1981 by G. Ghirardi in a Referee Report for Foundations of Physics to the paper by N. Herbert (G. Ghirardi: private communication).
7. N. Gisin: *Phys. Lett. A* **242**, 1 (1998); D. Bruss, G.M. D'Ariano, C. Macchiavello, and M.F. Sacchi: *Phys. Rev. A* **62**, 062302 (2000).
8. C. Simon, V. Buzek, and N. Gisin: *Phys. Rev. Lett.* **87**, 170405 (2001).
9. F. De Martini, V. Buzek, F. Sciarrino, and C. Sias: *Nature* (London) **419**, 815 (2002).
10. A. Peres: *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1993).
11. A. Alber *et al*: *Quantum Information: An Introduction to Basic Theoretical Concepts and Experiments* (Springer-Verlag, Berlin, 2001).
12. F. De Martini: *Phys. Rev. Lett.* **81**, 2842 (1998); F. De Martini, V. Mussi, and F. Bovino: *Opt. Comm.* **179**, 581 (2000).
13. C. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. Wootters: *Phys. Rev. Lett.* **70**, 1895 (1993).
14. A. Perelomov: *Generalized Coherent States and Their Applications* (Springer-Verlag, Berlin, 1985).
15. M.J. Collett: *Phys. Rev. A* **38**, 2233 (1988).
16. F. De Martini, G. Di Giuseppe and S. Padua: *Phys. Rev. Lett.* **87**, 150401 (2001).
17. C. Simon, G. Weihs and A. Zeilinger: *Phys. Rev. Lett.* **84**, 2993 (2000).
18. F.A. Bovino, F. De Martini and V. Mussi: arXiv:quant-ph/9905048 (1999).
19. D.F. Walls and G.I. Milburn: *Quantum Optics* (Springer Verlag, Berlin, 1994).
20. E. Schrödinger: *Naturwissenschaften* **23**, 807 (1935).
21. A. Caldeira, A. Leggett: *Physica A* **121**, 587 (1983).
22. A. Leggett, A. Garg: *Phys. Rev. Lett.* **54**, 857 (1985).
23. W. Zurek: *Physics Today*, October 1991, p 36.
24. C. Monroe, D. Meekhof, B. King, D. Wineland: *Science* **272**, 1131 (1996).

25. M. Brune, E. Hagley, J. Dreyer, X. Maistre, A. Maali, C. Wunderlich, J.M. Raimond, S. Haroche: Phys. Rev. Lett. **77**, 4887 (1996); L. Davidovich, M. Brune, J.M. Raimond and S. Haroche: Phys. Rev. A **53**, 1295 (1996); J.R. Friedman, V. Patel, W. Chen, S.K. Tolpygo, and J.E. Lukens: Nature (London) **406**, 43 (2000); M. Grenier, O. Mandel, T. Hänsch, I. Bloch: Nature (London) **419**, 51 (2002).
26. M.W. Noel, C.R. Stroud: Phys. Rev. Lett. **77**, 1913 (1996).
27. D. Pelliccia, V. Schettini, F. Sciarrino, C. Sias and F. De Martini: Phys. Rev. A **68**, 042306 (2003).
28. C.H. Bennett, D.P. Di Vincenzo, J.A. Smolin, and W.K. Wootters: Phys. Rev. A **54**, 3824 (1996).
29. P.G. Kwiat, E. Waks, A.G. White, I. Appelbaum and P.H. Eberhard: Phys. Rev. A **60**, R773 (1999).
30. M. Horodecki, P. Horodecki and R. Horodecki. In: *Quantum Information: An Introduction to Basic Theoretical Concepts and Experiments* (Springer, Berlin, 2001).
31. M.D. Reid, W.L. Munro, and F. De Martini: Phys. Rev. A **66**, 033801 (2002).
32. M. Redhead: *Incompleteness, Nonlocality and Realism* (Clarendon Press, Oxford, 1987).
33. A. Lamas-Linares, C. Simon, J. C. Howell and D. Bouwmeester: Science **296**, 712 (2002) first reported a  $N = 1, M = 2$  cloning experiment with a semi-classical strongly attenuated *coherent* injection state reaching a UOQCM fidelity  $F \approx 0.81$ .
34. V. Buzek: private communication to FDM; V. Buzek and M. Hillery: Phys. Rev. A **62**, 022303 (2000); M. Hillery, V. Buzek, and M. Ziman: Phys. Rev. A **65**, 022301 (2002).
35. F. De Martini: Phys. Lett. A **250**, 15 (1998). The collinear QIOPA is found to be an efficient generator of multiphoton Schrödinger Cat  $\vec{\pi}$ -states. The no-interaction term in Eq.(9.23) is easily discarded by a simple 2-detector coincidence measurement on the mode  $k$ .
36. M. Ricci, F. Sciarrino, C. Sias, and F. De Martini: arXiv:quant-ph/0304070 (2003); arXiv:quant-ph/0310021 (2003).
37. C.H. Bennett, and D. Di Vincenzo: Nature (London) **404**, 247 (2000).
38. D. Gottesman, and I.L. Chuang: Nature (London) **402**, 390 (1999).
39. D. Boschi, S. Branca, F. De Martini, L. Hardy, S. Popescu: arXiv:quant-ph/9710013 (1997); Phys. Rev. Lett. **80**, 1121 (1998); D. Bouwmeester *et al.*: Nature (London) **390**, 575 (1997); E. Lombardi, F. Sciarrino, S. Popescu, and F. De Martini: Phys. Rev. Lett. **88**, 070402 (2002).
40. S. Giacomin, F. Sciarrino, E. Lombardi, and F. De Martini: Phys. Rev. A **66**, 030302(R) (2002).
41. C.K. Hong, Z.Y. Ou and L. Mandel: Phys. Rev. Lett. **59**, 2044 (1987); J.C. Howell, I.A. Khan, D. Bouwmeester, and N.P. Bigelow: arXiv:quant-ph/0301170 (2003).
42. I.L. Chuang, M.A. Nielsen: *Quantum Information and Quantum Computation* (Cambridge Univ. Press, Cambridge, 2000).
43. G.M. D'Ariano. In: *Experimental Quantum Computation and Information*, Proceedings of the Intl.School “E. Fermi,” Varenna, ed by F. De Martini and C. Monroe (Editrice Compositori, Bologna, 2001).
44. G.M. D'Ariano and P. Lo Presti: Phys. Rev. Lett. **86**, 4195 (2001).

45. G.M. D'Ariano and H.P. Yuen: *Phys. Rev. Lett.* **76**, 2832 (1996).
46. F. De Martini, A. Mazzei, M. Ricci, and G.M. D'Ariano: *Phys. Rev. A* **67**, 062307 (2003).



# 10 Maximum-Likelihood Estimation in Experimental Quantum Physics

Gerald Badurek<sup>1</sup>, Zdeněk Hradil<sup>2</sup>, Alexander Lvovsky<sup>3</sup>, Gabriel Molina-Teriza<sup>4</sup>, Helmut Rauch<sup>1</sup>, Jaroslav Řeháček<sup>2</sup>, Alipasha Vaziri<sup>4</sup>, and Michael Zawisky<sup>1</sup>

<sup>1</sup> Atominstitut der Österreichischen Universitäten, Stadionallee 2, A-1020 Wien, Austria [rauch@ati.ac.at](mailto:rauch@ati.ac.at)

<sup>2</sup> Department of Optics, Palacky University, 17. listopadu 50, 772 00 Olomouc, Czech Republic [hradil@aix.upol.cz](mailto:hradil@aix.upol.cz)

<sup>3</sup> Universität Konstanz, Fakultät für Physik, Fach 696, D-78457 Konstanz, Germany, URL <http://www.uni-konstanz.de/quantum-optics/quantech/>

<sup>4</sup> Institut für Experimentalphysik, Universität Wien, Boltzmanngasse 5, A-1090 Wien, Austria

## 10.1 Introduction

Over the past hundred years, quantum theory proved to be an extremely useful tool for describing and understanding nature on a microscopic level. Its usual applications consist in predicting the outcomes of future experimental results given a quantum state describing the physical system.

However, in experimental physics one often faces a different question: “Given the outcomes of a particular set of measurements, what quantum state do they imply?” Such *inverse problems* may arise for instance in the stage of setting up and calibrating laboratory sources of quantum states, in the analysis of decoherence and other deteriorating effects of the environment, or in some special tasks in quantum information processing such as eavesdropping on a quantum channel in quantum cryptography.

Quantum state reconstruction is a highly nontrivial problem. The quantum state of a system, however simple, cannot be determined by a single measurement. Repeated identical measurements performed on multiple identical copies of a quantum state generally will not yield the complete information about the ensemble in question. Such a set of measurements will, however, provide the probability distribution of the ensemble measured over the eigenstates of the measurement apparatus.

To fully characterize an ensemble one needs to perform a set of multiple, *different* measurements on an even larger set of the identically prepared systems. By modifying the configuration of the apparatus, one acquires the quantum state statistics associated with various measurement bases. A set of these distributions contains complete information about the system, which can be then extracted in the form of a density matrix. In a sense, many different and mutually incompatible observations are compressed to a single mathematical object.

Various theoretical approaches to this problem are reviewed in other chapters in this volume. The purpose of this chapter is to provide the reader with some examples of quantum tomography taken from the modern experimental praxis. Maximum likelihood estimation will be applied to physical systems of increasing complexity starting with a simple one-dimensional problem of quantum phase, continuing through the absorption and phase neutron tomographies, further discussing quantum tomography of discrete quantum systems, and closing with the homodyne tomography of an infinite dimensional system—a mode of light.

## 10.2 Maximum-likelihood phase estimation

Phase measurements do not belong to the category of conventional measurements since a Hermitian phase operator does not exist in the canonical sense [1–5]. However, this does not mean that phase cannot be described in quantum theory or even measured. Let us first briefly review some basic facts related to experimental reality.

From the point of view of estimation theory [6] phase is a c-number parameter appearing in the transformation describing the action of a phase shifter on the input state  $|\Psi\rangle$

$$|\Psi(\bar{\theta})\rangle = e^{-i\bar{\theta}\hat{N}}|\Psi\rangle, \quad (10.1)$$

$\hat{N} = \hat{a}^\dagger\hat{a}$  being the photon-number operator. Any phase measurement is completely described by the statistics  $p(\theta|\bar{\theta})$  of its outcomes  $\theta$  conditioned on the true value  $\bar{\theta}$  of the phase shift. Basically, two strategies may arise. One can either look for the *ideal measurement* [7–11], i.e. the measurement optimal from some point of view, or, provided that a measurement cannot be chosen at our will, one should choose the statistical data analysis extracting as much information about the parameter of interest as possible.

The statistics of the ideal phase measurement are just the statistics of the Susskind-Glogower phase operator [12],

$$p(\theta|\bar{\theta}) = \frac{1}{2\pi}|\langle\theta|\Psi(\bar{\theta})\rangle|^2, \quad \hat{a}\hat{N}^{-1/2}|\theta\rangle = e^{i\theta}|\theta\rangle. \quad (10.2)$$

Although the eigenstates of this operator are not orthogonal, they are overcomplete, and thus generate probability operator valued measure (POVM)  $\Pi(\theta) = |\theta\rangle\langle\theta|$ , that defines the ideal phase measurement in the sense of generalized measurements.

Though there are ways to simulate measurement (10.2) by means of post selection [13], this does not seem to be a practical solution. What is usually measured in practice is energy, and phase sensitive devices called interferometers are used to transform phase shifts into variations of output energies. Due to the statistical nature of quantum theory, the resulting relationship

between the measured quantities and the parameters of interest is not deterministic. Such an indirect inference is usually called quantum estimation, and its scheme is the following:

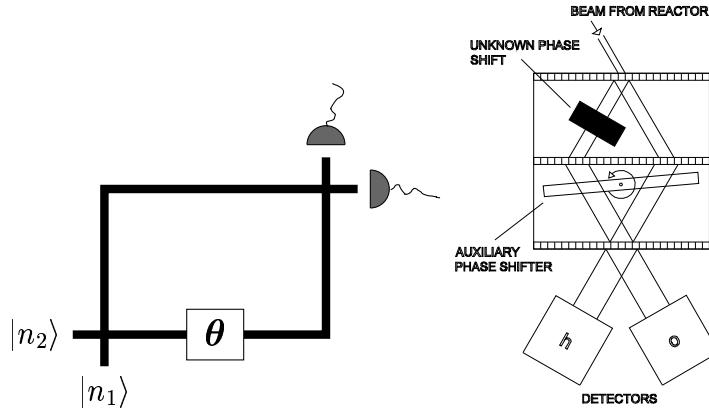
$$\rho \rightarrow \boxed{\begin{array}{l} \text{true phase} \\ \text{shift } \bar{\theta} \end{array}} \rightarrow \rho(\bar{\theta}) \rightarrow \boxed{\text{detection}} \rightarrow \mathbf{n}(\bar{\theta}) \rightarrow \boxed{\text{estimation}} \rightarrow \theta(\mathbf{n}).$$

An interferometer provides the input-output transformation of the known initial state  $\rho$ . The subsequent measurement yields a phase sensitive data  $\mathbf{n}$  that are processed to get a phase estimate  $\theta$ . The true phase shift  $\bar{\theta}$  inside the interferometer, which is a non-fluctuating parameter controlled by the experimentalist, should carefully be distinguished from the phase estimate  $\theta$ , which is generally a random quantity.

The performance of the estimation of course depends on the choice of the estimator. The point estimators of phase corresponding to the ML estimation will be used here [8,14]. In accordance with the ML approach [15], the sought-after phase shift is given by the value, which maximizes the likelihood function

$$\mathcal{L} \equiv p(\theta|\mathbf{n}) \propto p(\mathbf{n}|\theta). \quad (10.3)$$

A schematic representation of a Mach-Zehnder (MZ) interferometer is shown in Fig. 10.1. Its input ports are fed by  $n_1$  and  $n_2$  particles. Formally,



**Fig. 10.1.** Left side: Mach-Zehnder optical interferometer; right side: silicon perfect crystal interferometer used in neutron optics.

this device can be described by Lie algebra SU(2), the correspondence being provided by the Schwinger representation of the angular momentum operators.

The phase resolution of the MZ interferometer strongly depends on the properties of the input signal. Provided that the interferometer is operated

in the usual (classical) manner with the light entering one input port only,  $|in\rangle = |N, 0\rangle$ , the phase error is proportional to  $1/\sqrt{N}$ . This regime is usually referred to as the standard limit of phase measurements. When both the input ports of the interferometer are fed by the signal with an equal number of particles,  $|in\rangle = |N/2, N/2\rangle$ , the theory predicts the ultimate phase resolution of  $\Delta\theta \approx 1/N$ , which is the quantum limit of phase measurements [9, 10].

While the classical limit can be achieved using a coherent source, or by accumulating data from repeated single particle interference experiments, the quantum limit requires a bright source of highly nonclassical exotic states such as Fock states or highly squeezed states [7, 9], which are still not readily available in laboratories. But even in the absence of such resources one can still benefit from the proper data manipulation as will be shown on the example of a semi-classical and fully quantum phase reconstruction.

The quantum phase reconstruction can also be regarded as the simplest case of a quantum-state reconstruction. A single particle being in a superposition of two paths is equivalent to a two-level system and can therefore be visualized by a point inside the Bloch sphere. The action of the interferometer is related to the equatorial cut through this sphere: The phase shift rotates the Bloch vector along the equator, while the decoherence affecting the visibility contracts this vector. So the natural parameter space of the phase reconstruction is a unit circle.

### 10.2.1 Quantum phase estimation

First, let us show that the operational phase concepts can naturally be embedded in the general scheme of quantum estimation theory [6, 16] as was done in Refs. [17–20].

Let us consider a MZ interferometer, where the measurement is performed with zero and  $\pi/2$  auxiliary phase shifters. The auxiliary shifter is needed to get a unique phase estimate. Such a device is equivalent to the 8-port homodyne detection scheme [21] with the four output channels numbered by indices 3, 4, 5, and 6, where the numbers of particles are registered in each run. Assume that the these quantities fluctuate in accordance with some statistics. The mean intensities are modulated by a phase parameter  $\bar{\theta}$

$$\bar{n}_{3,4} = \frac{N}{2}(1 \pm V \cos \bar{\theta}), \quad \bar{n}_{5,6} = \frac{N}{2}(1 \pm V \sin \bar{\theta}), \quad (10.4)$$

where  $N$  is the total intensity and  $V$  is the visibility of the interference fringes. Provided that a particular combination of outputs  $\{n_3, n_4, n_5, n_6\}$  has been registered, the phase shift can be inferred.

Now, for a while, let us assume that the phase sensitive device operates with a Gaussian signal with phase insensitive noise. This is only an approximation to the real situation since realistic signals are discrete. Under such approximation, the likelihood function corresponding to the detection of given data reads

$$\mathcal{L} \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=3}^6 [n_i - \bar{n}_i]^2 \right\}. \quad (10.5)$$

Here the variance  $\sigma^2$  represents the phase insensitive noise of each channel. The sampling of intensities may serve for the estimation of the phase shift and visibility simultaneously. Likelihood function (10.5) is maximized on the physically allowed space of parameters  $V \leq 1$  by the following phase and visibility,

$$e^{i\theta} = \frac{n_3 - n_4 + i(n_5 - n_6)}{\sqrt{(n_3 - n_4)^2 + (n_5 - n_6)^2}}, \quad (10.6)$$

$$V = \min \left( 2 \frac{\sqrt{(n_3 - n_4)^2 + (n_5 - n_6)^2}}{\sum_{i=3}^6 n_i}, 1 \right). \quad (10.7)$$

Notice that the prediction of this semi-classical theory (10.6) coincides with the operational quantum phase introduced by Noh, Fougères and Mandel (NFM) [22]. This means that operational phase concepts can be thought of as special cases of ML phase estimation – ML estimation for Gaussian signals. Such predictions are optimal only for signals represented by a continuous Gaussian signal with phase independent and symmetrical noises. Let us note that a generalization of this concept to a larger number of auxiliary phase shifts detected [23] is known as the phase of the discrete Fourier transformation.

Since realistic signals are discrete, the theory can be refined by considering the actual statistics of the experiment. This can be demonstrated on the case of Poissonian signals. These are frequently encountered in laboratories as ideal lasers, or thermal sources of particles, such as neutron beams. ML estimation based on the Poissonian likelihood function,

$$\mathcal{L} \propto \prod_{i=3}^6 \frac{\bar{n}_i^{n_i}}{n_i!} e^{-\bar{n}_i}, \quad (10.8)$$

gives optimum values for the phase shift and visibility

$$e^{i\theta} = \frac{1}{V} \left[ \frac{n_4 - n_3}{n_4 + n_3} + i \frac{n_6 - n_5}{n_6 + n_5} \right], \quad (10.9)$$

$$V = \sqrt{\left( \frac{n_4 - n_3}{n_4 + n_3} \right)^2 + \left( \frac{n_6 - n_5}{n_6 + n_5} \right)^2} \leq 1, \quad (10.10)$$

provided that the estimated visibility (10.10) is smaller than unity. If not, it is necessary to maximize the likelihood function (10.8) numerically on the boundary ( $V = 1$ ) of the physically allowed region of parameters.

Relations (10.9-10.10) provide a correction of the semiclassical Gaussian theory with respect to the discrete Poissonian signals. The principle of inference together with the two different assumptions about the nature of the

signal have given rise to two different phase estimates (10.6) and (10.9). One may wonder, whether the improvement of phase inference gained by taking the correct statistics of the experiment into account is worth giving up the simple NFM formalism and resorting to numerical methods. Could the optimization of the information yield from the measured data lead to a significant increase of the accuracy of the phase fitting?

The difference between recipes (10.6) and (10.9) can be tested in a controlled phase measurement. The phase difference is adjusted to a certain value and estimated independently using both the methods (10.6) and (10.9) in repeated experiments. The efficiency of both methods can then be compared.

Of course, some measure of the estimation error is needed for this. Dispersion defined by the relation

$$\sigma^2 = 1 - |\langle e^{i\theta} \rangle|^2 \quad (10.11)$$

is one such reasonable measure. Here, the average is taken over the posterior phase distribution of the corresponding phase estimator.

The evaluation of the average quadratic cost (10.11) is not the only way to compare the efficiencies of different estimation procedures. Another possibility is to use the rectangular cost function

$$C(\theta - \bar{\theta}) = \begin{cases} -1 & |\theta - \bar{\theta}| \leq \Delta\theta \\ 0 & |\theta - \bar{\theta}| > \Delta\theta \end{cases} \quad (10.12)$$

The averaged rectangular cost  $\langle C(\theta - \bar{\theta}) \rangle$  measures how many times the estimate  $\theta$  falls within the chosen window  $\Delta\theta$  spanning around the true phase  $\bar{\theta}$ . The difference

$$\Delta E = \langle C(\theta - \bar{\theta}) \rangle_{\text{Gauss}} - \langle C(\theta - \bar{\theta}) \rangle_{\text{Poiss}} \quad (10.13)$$

then measures how much the Poissonian prediction is better than the Gaussian one. If this quantity is found to be positive, the ML estimation is better than its NFM counterpart.

Although the dispersion (10.11), and  $\Delta E$  cannot be calculated explicitly for an arbitrary input intensity  $N$ , it is possible to analyze the limit cases [20]. Obviously, both the predictions (10.6) and (10.9) will coincide provided that there is almost no information available in the low intensity limit,  $N \rightarrow 0$ . Not so obvious is the fact that both predictions will also coincide in the high intensity limit,  $N \gg 1$ , provided the visibility is low  $V \rightarrow 0$ . To see this, let us compare the asymptotic dispersion of the NFM estimator,

$$\sigma_G^2 \approx \frac{1}{V^2} N^{-1} + O\left(\frac{1}{N^2}\right), \quad (10.14)$$

with the asymptotic expression for the Cramér-Rao lower bound (CRLB) on the dispersion of any estimator,

$$\sigma_{CRLB}^2 = \frac{V^2 - 1 - \frac{1}{4}V^4 \sin^2 2\bar{\theta}}{V^2 - 1 - \frac{1}{2}V^2 \sin^2 2\bar{\theta}} V^{-2} N^{-1} + O\left(\frac{1}{N^2}\right). \quad (10.15)$$

If the visibility is low, both expressions will become identical. Therefore the semiclassical theory is optimal in this limit case [23].

Asymptotic expressions for various phase estimators in the opposite limit of high visibility  $V \approx 1$  are given in Table 10.1. Estimator  $ML'$  is the phase

**Table 10.1.** Asymptotic dispersions and overall quadratic costs of various phase estimators.  $ML'$  – unconstrained ML estimation;  $ML_1$  – single-parameter ML estimation. For comparison, CRLB is also shown.

Estimator	$\sigma^2$	$\bar{C} \equiv \int \sigma^2 d\bar{\theta}$
NFM	$1/N$	$2\pi/N$
$ML'$	$(1 + \cos^2 2\bar{\theta})/2N$	$\frac{3}{2}\pi/N$
$ML$	$\approx (1 + 0.5 \cos^2 2\bar{\theta})/2N$	$\approx \frac{5}{4}\pi/N$
$ML_1$	$1/2N$	$\pi/N$
CRLB	$1/2N$	$\pi/N$

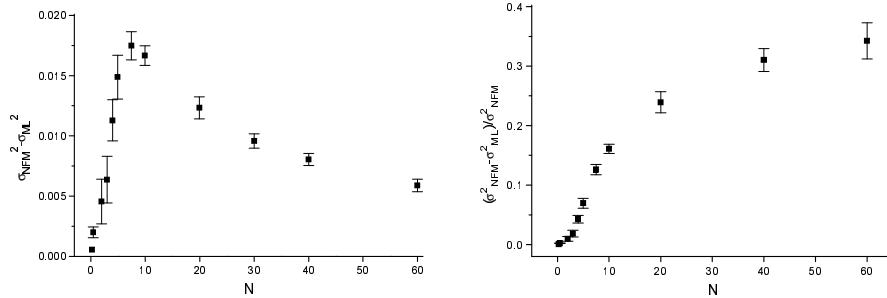
prediction given by Eq. (10.9), that is, one does not care about the possible unphysical inferred visibilities  $V > 1$ , and  $ML_1$  is the single-parameter estimation obtained by setting  $V = 1$  in the likelihood function (10.8) and maximizing it only with respect to phase.

Notice that the uncertainties of all estimators scale as  $1/\sqrt{N}$ . This is to be expected in accordance with the standard quantum limit. However the constant of proportionality depends on the estimator used. By taking physical constraints into account the accuracy of phase fitting is improved.

The single-parameter  $ML_1$  estimator is seen to provide best phase predictions. It attains the CRLB and hence is optimum. It yields a phase prediction whose uncertainty is reduced by the factor of  $\sqrt{2}$ , that is by about 30%, compared with the semi-classical theory. However, estimating phase alone implicitly presumes good a-priori knowledge of the visibility. If the actual value of the visibility is not known or fluctuates during the experiment, the single-parameter estimator may lead to biased phase predictions. For large intensities the bias might completely spoil the estimation [20].

### 10.2.2 Experiments

The performance of the semiclassical NFM and ML phase estimators have been determined in a series of experiments utilizing two principal sources of particles – beams of thermal neutrons [19] and laser light [20]. The main goal of the experiments was to compare the optimum phase prediction with the semi-classical theory in the regime of only a few input particles. As a side result, the theoretical asymptotic uncertainties given in Table 10.1 were tested experimentally.



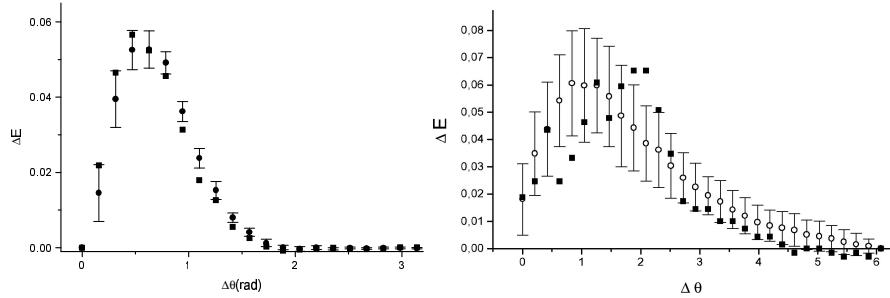
**Fig. 10.2.** The experimentally observed absolute (left panel) and relative (right panel) difference between the dispersions of the NFM and ML estimators as a function of the input mean number of photons  $N$  for fixed true phase  $\bar{\theta} = \pi/3$ . Error bars corresponding to 68% confidence intervals are also shown.

The dispersions (10.11) of the NFM (or equivalently Gaussian) and ML phase estimators found in experiments with light are shown in Fig. 10.2. The true phase was fixed at  $\bar{\theta} = \pi/3$ . The number of detected quadruples  $\{n_3, n_4, n_5, n_6\}$  used for the calculation of the dispersions varied from 1000 samples for the input mean number of photons  $N = 60$  to more than 100,000 samples for  $N = 0.1$ . The error bars corresponding to these finite numbers of samples are the result of numerical simulation. The visibility during the experiments was better than 99.6%.

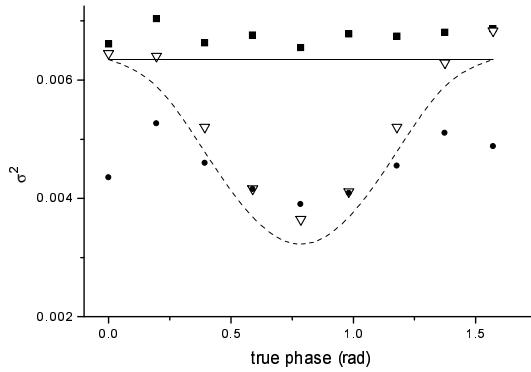
The ML estimator was found to be significantly more accurate (by many standard deviations) than its NFM semi-classical counterpart. This was confirmed by evaluating the difference of the rectangular costs (10.13), see the left panel in Fig. 10.3. Here, the chosen input total energy roughly fits the maximum,  $N \approx 7.5$ , of the curve seen in Fig. 10.2.

A significant difference between the effectiveness of semi-classical and optimal treatments is apparent in Fig. 10.3. The optimal treatment provides an improvement in estimation procedure, and the difference is more than 10 standard deviations beyond the statistical error. High stability and visibility of interference fringes in the optical interferometer along with a high repetition rate of pulsed lasers made the improvement of the semi-classical phase prediction more evident than in a similar comparison that had been done with thermal neutrons [19], see the right panel in Fig. 10.3.

An experimental comparison of three different phase estimations – NFM, ML', and ML estimators – in the asymptotic regime is shown in Fig. 10.4. The experiment was done with photons. For comparison, the theoretical values of dispersions given in Table 10.1 are also shown. Several important conclusions can be drawn from Fig. 10.4. (i) We can see that the uncertainty of the constrained ML estimation is definitely below the uncertainty of the unconstrained estimation in agreement with theory presented in Table 10.1. It means that insisting on the physical constraints (here the non-negativity of



**Fig. 10.3.** Experimentally obtained  $\Delta E$  (squares) compared to theoretical values (circles). Left panel: experiment with photons,  $N = 10$  photons, visibility was better than 99.5%; error bars correspond to 7500 measured samples. Right panel: experiment with neutrons. The mean number of  $N = 8.54$  incoming neutrons was asymmetrically split between the ordinary ( $N_o = 2.21$ ) and extraordinary ( $N_h = 6.33$ ) channels. The average visibility was about 31%; error bars correspond to 690 registered samples.



**Fig. 10.4.** Asymptotic dispersions of the semi-classical NFM estimator; theory (solid line) and experimentally obtained values (squares). Asymptotic dispersions of the unconstrained ML estimator; theory (dashed line) and experimentally obtained values (triangles). Experimentally obtained dispersions of the ML estimation on the physical space of parameters (circles). The corresponding input mean number of photons and the estimated visibility were  $N = 160$  and  $V = 99.2\%$ , respectively.

the intensity) is important not only for the sake of interpretation, but it also makes the estimation more efficient. Of course, both ML estimations beat the phase resolution of the semi-classical NFM theory. (ii) The observed values of dispersions exhibit a systematic error. The additional noise above the theoretical uncertainty is caused by inherent phase fluctuations in the experimental setup. Hence our statistically motivated evaluation of experimental data can be used for inferring the amount of fluctuations, and therefore it provides an independent and nontrivial way of calibrating an interferometer.

### 10.3 ML neutron absorption tomography

Though the ML reconstruction has been proposed in quantum domain, it can be applied advantageously to any statistical theory, particularly to absorption tomography. This classical problem also illuminates the tomographic aspects of the quantum-state reconstruction.

The standard reconstruction method in present computerized tomographic (CT) imaging is the filtered back-projection (FBP) algorithm based on the inverse Radon transformation [24]. Technically, this transformation is implemented using fast Fourier transformation routines, which makes the reconstruction process fast and highly efficient. FBP gives satisfactory results in applications where (i) the intensity of the illuminating beam is so large that its statistics can safely be ignored, and (ii) the sample can be scanned over the whole  $180^\circ$  angular interval in small steps. In X-ray medical CT imaging these two conditions are usually met.

Unfortunately FBP fails in case of missing projections and/or if strong statistical fluctuations of the counting numbers are present in the small detector pixels. The latter situation occurs e.g. in neutron tomography [25–28], if monochromatic neutron beams are applied in order to avoid beam artifacts [29] or at the investigation of strong absorbing materials [30]. As present neutron sources are thermal in nature, they generate weak beams. Particles that have passed through the studied object are counted one by one and their statistical fluctuations have strong influence on the reconstructed images.

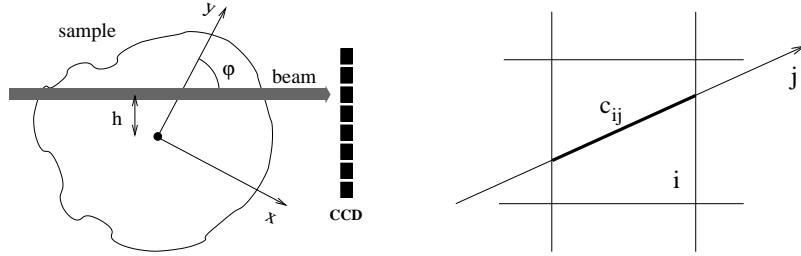
The case of missing projections or incomplete data sets is another important issue. If technical problems arise during the tomographic scan, FBP algorithm requires repeating the whole measurement. Partial or incomplete data cannot be inverted using this method. Sometimes it may be required to keep the sample in a cryostat during the measurement. The construction of such a cryostat may not allow turning it upside-down, so part of the measured angles may be missing, or there may not be enough space to rotate the sample in the full  $180^\circ$  interval. This was investigated in the past in detail by means of algebraic reconstruction techniques [31–33].

The maximum-likelihood reconstruction method described in this section can improve several tomographic applications in neutron optics which in many cases are limited by the weak intensity and the poor detector resolution [34,35]. Generally, it achieves better reconstruction results or reduce the scanning time in neutron optics and in medical and biological CT imaging.

In the following we will focus on the absorption tomography in neutron optics.

#### 10.3.1 LinPos tomography

Basic notions and the geometry of typical tomographic experiments are as follows. Let us assume that the sample is illuminated by parallel monochromatic pencil beams, see Fig. 10.5. Data consist of the number of particles



**Fig. 10.5.** Geometry of the experimental setup and the definition of coefficients  $c_{ij}$ .

counted behind the sample for  $M$  different scans – each scan being characterized by its horizontal position  $h$  and rotation angle  $\varphi$ . Alternatively, a broad illuminating beam combined with a position-sensitive detector (CCD camera) placed behind the sample can be used. In that case  $h$  labels the pixels of the camera. For the sake of simplicity a collective index  $j \equiv \{h, \varphi\}$  will be used, hereafter, to label the scans.

Mean number  $\bar{n}_j$  of particles (intensity) registered in  $j$ -th scan is given by the exponential attenuation law

$$\bar{n}_j = \bar{n}_0 \exp(- \int \mu(x, y) ds_j), \quad (10.16)$$

where  $\bar{n}_0$  is the intensity of the incoming beam,  $\mu(x, y)$  is the absorption index of the sample in position  $\{x, y\}$ , and the integration is the path integration along the pencil beam. This exponential attenuation law is a good approximation if scattering can be neglected. The beam hardening artifacts would also modify relation (10.16) but this complication can be avoided experimentally by the use of monochromatic beams [29].

Using neutron beams instead of X-ray photons has several advantages. Neutrons interact with the matter in a different way than photons do. Since neutrons are subject to the strong interaction, they can “see” and distinguish among different isotopes of the same element. Neutrons can also penetrate deeper into some materials such as metal surfaces, and many other objects like objects made from plastic materials often show considerably more detail when illuminated with neutrons rather than photons.

The absorption index for thermal neutron beams can be expressed as

$$\mu(x, y) = \Sigma_{th}(x, y) \lambda / \lambda_{th}, \quad (10.17)$$

where  $\Sigma_{th}$  is the macroscopic thermal cross section,  $\lambda$  is the wavelength of the illuminating beam, and  $\lambda_{th}$  is the thermal wavelength. Eq. (10.17) presumes a linear dependence of the involved cross sections on  $\lambda$  [36].

For practical purposes, it is convenient to discretize Eq. (10.16) as follows,

$$\bar{n}_j = \bar{n}_0 \exp\left(-\sum_{i=0}^N \mu_i c_{ij}\right). \quad (10.18)$$

The sample is now represented by a 2D mesh whose cells are assumed to be homogeneous. The variables are now  $N$  numbers  $\mu_i$  specifying absorption in those elementary cells. Matrix  $\{c_{ij}\}$  defines the overlaps of beams and cells, see Fig. 10.5.

Let us first ignore the statistics of the illuminating beam, and assume that the counted numbers of particles  $\{n_j\}$  do not fluctuate,  $n_j = \bar{n}_j, \forall j$ . Taking the logarithms of both sides of Eq. (10.18), one obtains a system of  $M$  linear algebraic equations for the  $N$  unknown absorption coefficients  $\mu_i$ :

$$f_j = p_j, \quad j = 1 \dots M, \quad (10.19)$$

where we defined,

$$f_j = -\ln \frac{n_j}{n_0}, \quad p_j = \sum_i \mu_i c_{ij}. \quad (10.20)$$

Notice that the problem (10.19) is a linear and positive (LinPos) problem. Its linearity is obvious, and positivity follows from the fact that no new particles are created in the sample. The importance of LinPos problems for experimental physics stems from the fact that many physical quantities, for instance mass, density, intensity, and so on, are intrinsically positive quantities that very often depend on the parameters of interest in a linear way. Linearized absorption tomography (10.19) is just one typical example taken from this wide family of problems including, among others, the measurement of focal intensity distribution for polarized input fields [37], or the characterization of the photon content of light pulses [38] to give some examples.

The inversion of Eq. (10.19) is provided by the FBP algorithm,

$$\boldsymbol{\mu} \propto \sum_{\varphi} \mathbf{f} \star \mathbf{g}, \quad (10.21)$$

where  $\star$  denotes a discrete convolution of data with the regularized singular transformation kernel [39]. However, by linearity of (10.21), the positivity of the reconstructed absorption index is not guaranteed when the data is noisy. A negative value of a reconstructed  $\mu_i$  would then suggest that particles were being created in the  $i$ -th cell in the course of the experiment, which would obviously be a wrong conjecture. Another problem arises when there are only a few projections available. In such a case, the summation in (10.21) no longer approximates the original integral and unwanted artifacts appear in the reconstructions. Both these drawbacks can be avoided if the problem (10.19) is solved in the sense of maximum likelihood on the space of physically allowed absorption coefficients. In this approach one considers the data  $\mathbf{f}$  and the prediction of the theory  $\mathbf{p}$  as two probability distributions. One looks for absorption coefficients  $\{\mu_i\}$  minimizing the Kullback-Leibler “distance”

$$D(\mathbf{f}, \mathbf{p}) = - \sum_j f_j \ln \frac{p_j}{f_j} \quad (10.22)$$

between the data  $\mathbf{f}$  and theory  $\mathbf{p}$ . Here, a little more care is needed since  $\mathbf{p}$  and  $\mathbf{f}$  are generally not normalized to unity. The minimum of the Kullback-Leibler distance corresponds to the maximum of the likelihood functional [15]

$$\mathcal{L} = \prod_j \left( \frac{p_j}{\sum_k p_k} \right)^{f_j}, \quad (10.23)$$

that quantifies the likelihood of the given distribution  $\{\mu_i\}$  in view of the registered data. We seek the maximum-likely distribution of the absorption indices. A convenient way to find it is the Expectation-Maximization (EM) iterative algorithm [40, 41],

$$\boldsymbol{\mu}^{n+1} = \mathbf{R}(\boldsymbol{\mu}^n) \cdot \boldsymbol{\mu}^n, \quad (10.24)$$

where

$$R_i = \frac{1}{\sum_{j'} c_{ij'}} \sum_j \frac{f_j c_{ij}}{p_j(\boldsymbol{\mu})}, \quad (10.25)$$

and  $\boldsymbol{\mu}^0$  is some initial strictly positive distribution  $\mu_i^{(0)} > 0$ ,  $i = 1 \dots N$ . A nice feature of EM algorithm is that its convergence is guaranteed for any input data  $f_j$  [42]. For this reason it became a valuable tool in many inverse problems that can be reduced to the form of Eq. (10.19), e.g. in positron emission tomography [42–44]. The original derivation of EM algorithm is based on alternating projections on specially chosen convex sets of vectors. However, one could directly use the calculus of variations to derive the necessary condition for the extreme of the functional (10.23). Iterating these, one eventually arrives at the EM algorithm again. An advantage of this alternative derivation is that it can be also applied to more realistic physical models of the actual absorption experiment. One such possible generalization will be shown in the following subsection.

### 10.3.2 Tomography with Poissonian signals

Real signals are not composed of a sharp number of particles. For instance, two kinds of signals often used in experiments —beams of thermal neutrons and laser light— both exhibit Poissonian fluctuations of the number of particles. Monochromatic neutron beams also are correctly described by the Poissonian statistics if the detected count events occur in a mutually independent manner [45]. The knowledge of the true character of the signal illuminating the sample is a useful piece of prior information, which can be utilized for improving the performance of the tomographic imaging.

As the Poissonian character of the signal is preserved during the attenuation, the counted numbers of particles behind the sample are random Poissonian variables. The corresponding likelihood functional reads,

$$\mathcal{L} = \prod_j \frac{\bar{n}_j^{n_j}}{n_j!} e^{-\bar{n}_j}. \quad (10.26)$$

This is the joint probability of counting  $\{n_j\}$  particles. Their mean values  $\{\bar{n}_j\}$  obey the exponential law (10.16) as before. They depend on the absorption in the sample  $\{\mu_j\}$  that is to be inferred from the data. The necessary condition for the maximum of the likelihood (10.26) can be derived using the calculus of variations. The extremal equation can be shown to have the same vector form as the extremal equation of the LinPos problem (10.24) with the vector  $\mathbf{R}$  replaced by

$$R_i^{(\text{Poisson})} = \frac{\bar{n}_0}{\sum_{j'} c_{ij'} n_{j'}} \sum_j c_{ij} \exp(-\sum_{i'} \mu_{i'} c_{i'j}). \quad (10.27)$$

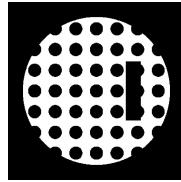
When the input intensity  $\bar{n}_0$  is not known, it should be estimated too:

$$\bar{n}_0 = \sum_j n_j / \sum_j \exp(-\sum_i \mu_i c_{ij}). \quad (10.28)$$

As seen, the Poissonian tomography is intrinsically a nonlinear problem. This has serious consequences for the convergence properties of the iterative algorithm (10.24) and (10.27). Instead of converging to a stationary point it might end up in oscillations. Typically, such convergence problems arise in the presence of very noisy data. When this happens, it is necessary to decrease the size of the iteration step as follows:  $R_i \rightarrow R_i^\alpha$ ,  $i = 1 \dots M$ ,  $0 < \alpha < 1$ . Of course, any solution to the regularized problem is also a solution to the original problem.

### 10.3.3 Comparison with standard methods

In real experiments there are many factors that could influence the quality of the measured data and therefore also the result of the tomography. Misalignments present in the experimental setup, instability of the illuminating beam, white spots and damaged detector pixels can be such factors, to name a few. To avoid this problem let us first show a few simulations. The data were generated on a computer from the artificial object shown in Fig. 10.6. It is a circle made of a homogeneous material with many small round holes drilled through it. One additional rectangular piece of material was removed from the circle to make it less symmetric. The absorption index of the material was chosen in such a way that the maximum attenuation along a beam was close to 50% of the input intensity.



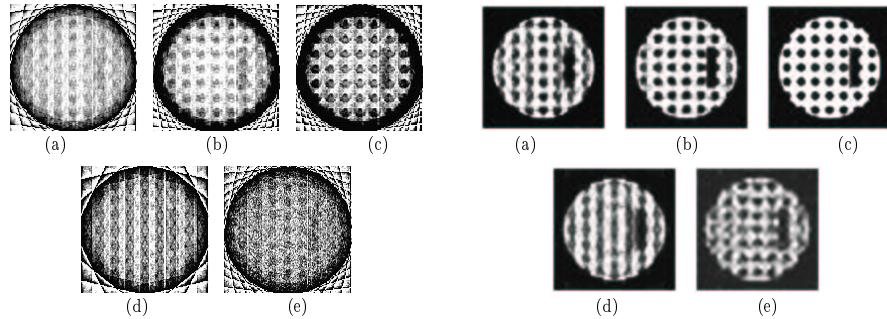
**Fig. 10.6.** The object.

**Table 10.2.** Quality of the input data. The last column shows the mean number of counted particles per pixel in the incident beam.

	reconstruction angles	pixels	intensity
a	13	161	$\infty$
b	19	101	$\infty$
c	20	101	$\infty$
d	7	301	$\infty$
e	15	161	2000

In the simulations, the object was subject to five different experiments. Their parameters are summarized in Table 10.2. First four experiments correspond to the ideal situation of a very high beam intensity where the Poissonian detection noise can safely be ignored. The last reconstruction simulates more realistic conditions with 2000 counts per pixel in the open beam. Notice that a relatively small number of rotations is chosen for all five experiments. In this regime the inverse Radon transformation is expected to yield bad results and the improvement of the maximum-likelihood tomography upon the standard technique should be most prominent. This regime is also important from the practical point of view. Doing more rotations implies a longer measurement time and more radiation absorbed by a sample. The latter may be an important factor if the imaging of biological samples is considered. Provided the improvement of the reconstruction technique gives comparable resolution with less data, imaging costs and damage done to a sample due to radiation might be reduced.

Reconstructions from the simulated data are shown in Fig. 10.7 [46]. The simulated data were first processed using the IDL imaging software (Research Systems Inc.) which implements the standard FBP algorithm (inverse Radon transform), see the left panel in Fig. 10.7. This software is one of the industrial standards in the computer assisted tomography. The same data were then processed using the iterative algorithm based on the maximization of the Poissonian likelihood function, see the right panel Fig. 10.7. In the absence of noise, see cases (a)-(d), the fidelity of a reconstruction depends on two main factors—the spatial resolution of the detector, and the number of rotations used. It is apparent from Fig. 10.7 that the latter factor is the more



**Fig. 10.7.** Left panel: IDL reconstructions from the simulated data, for parameters see Tab. 10.2. Right panel: ML reconstructions from the same data; the iterative algorithm Eqs. (10.24) and (10.27) has been used for reconstruction.

important of the two. Very small number of angles cannot be compensated by an increased spatial resolution of the detector, compare e.g. cases (c) and (d), and reconstruction (d) is by far the worst one. However, ML tomography is much less sensitive to the number of angles than the standard filtered back-projection. Even the large rectangular hole in the object is hardly perceptible in the IDL reconstruction (d) in Fig. 10.7, whereas it nicely shows in the ML reconstruction from the same data. ML reconstructions are superior to the standard ones also in cases (a)-(c); notice that the ML reconstruction (c) that is based on as few as 20 different angles is nearly perfect.

Benefits of the ML tomography are fully revealed when the detected data are noisy. This is case (e) in Tab. 10.2. Standard filtered back-projection applied to noisy data faces serious difficulties. This is due to ill-posedness of the Radon transformation where data are integrated with a singular filter function. Obviously such deconvolution greatly amplifies any noise present in the data. Having little or no prior information about the object it is difficult to tell true details of the object from artifacts. ML tomography gives much better results. Since noises are incorporated into the algorithm in a natural and statistically correct way, artificial smoothing is not needed. Notice in simulation Fig. 10.7e that the noisy data yields through the ML algorithm a little distorted but otherwise clear image, unlike the corresponding very noisy standard reconstruction. This is a nice feature of the intrinsically non-linear ML algorithm, which, in the course of reconstruction, self-adapts to the registered data and always selects the most likely configuration.

#### 10.3.4 Strongly absorbing materials: an experiment

One specific application of neutron CT imaging is the quantitative analysis and three-dimensional visualization of the  $^{10}\text{B}$  isotope distribution in boron alloyed steel. Boron alloyed steel is used in nuclear engineering as neutron shielding for the radioactive waste disposal equipment, such as components

for the compact fuel storage racks and transportation baskets. The main demand on the sheets for these applications is, besides mechanical stability and corrosion resistance, the largest possible thermal neutron attenuation, which has to be uniform over the whole volume. The attenuation in the steel depends mainly on the  $^{10}\text{B}$  isotope, which has a large attenuation cross section for thermal neutrons  $\sigma_{th}(^{10}\text{B}) = 3838.1(10) \times 10^{-24}\text{cm}^2$  [47].

Imaging of strongly absorbing samples suffers from the beam hardening effect. The wavelength dependence of scattering cross section (10.17) causes a spectral change of the neutron flux in the sample, where preferably low energy neutrons are absorbed so that the remaining beam becomes richer in high energy neutrons. Besides this effect, the following specific difficulties occur with strong absorbing materials:

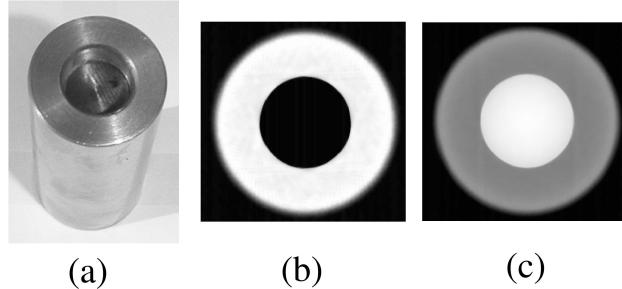
- The exposure time for one projection is several minutes and the total measurement time for one tomographic data set lasts several hours. If the data are analyzed with the FBP algorithm, one depends on a stable operation of the neutron source and CCD detector.
- The low count numbers lead to increased statistical fluctuations which get amplified in the FBP reconstruction.
- The imaging quality depends strongly on low background conditions. The contribution of scattered thermal neutrons, fast neutrons, gammas, and light penetration in the detector box has to be suppressed as far as possible.
- The stronger the attenuation in the sample, the more the result will interfere with the choice of input parameters in the FBP routine.

All these reasons clearly favor ML statistical inversion over the deterministic FBP algorithm.

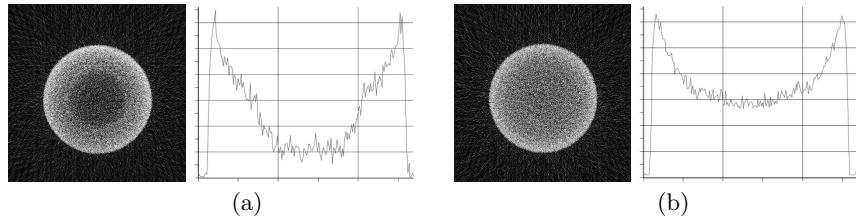
Let us show some examples of the tomographic investigations of strong absorbers that were performed at the 250 kW TRIGA reactor of the Atominstitut in Vienna, where at a well-thermalized beamline, a neutron-tomography facility had been implemented [48]. Surprisingly, neutron-tomography of strong absorbers is still possible with weak beam intensities of about  $10^5\text{n/cm}^2\text{s}$ . High resolution camera optics with the nominal resolution of  $80\mu\text{m}$  was used [48]. During the experiment, only a few neutrons per second per pixel were registered. Naturally, the discrete character of the quantum signal plays an important role at such low intensities.

The measured sample, see Fig. 10.8 was a two-component system consisting of a ring with an outer diameter of 2 cm and a hole of 1 cm diameter. The hole was partially filled with a second rod of 1 cm diameter with somewhat larger  $^{10}\text{B}$  content.

First we will show some typical results obtained with the standard FBP algorithm, see Fig. 10.9. Notice, that despite a large number of projections (angles) that were used for the inversion, the reconstructions are still rather noisy. Also, the reconstructed absorption profiles are far from the expected



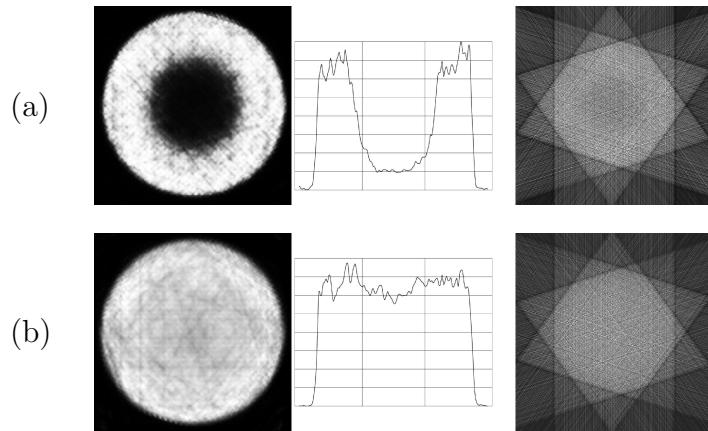
**Fig. 10.8.** Measured steel sample. This is a two component system, which consists of an outer ring ( $\emptyset = 2$  cm), partly filled with a second rod with 68% enlarged boron content ( $\emptyset = 1$  cm). Expected distribution of the absorption index in the upper hollow (b) and lower filled (c) regions are also shown [30].



**Fig. 10.9.** FBP reconstructions of the sample shown in Fig. 10.8. 50 different projections were used. (a) A typical reconstructed cut through upper region, where instead of the second rod only an air hole is inside the ring; (b) a typical reconstructed cut through the middle region, where the second steel rod with higher boron content is inside the ring. The noise in the profile plots was suppressed by averaging over several tens of reconstructed slices.

ones. This is caused partly by the previously mentioned beam hardening effect that was not taken into account here.

Figure 10.10 illustrates the usefulness of the ML technique in cases where only a few projections are available. The reconstructions are based on the same experimental data as the corresponding Figures 10.9a and 10.9b. However, the number of projections was reduced from 50 to just 10 (!) projections. In this extreme case the filtered back-projection fails completely. It is interesting to notice that although no correction of the beam hardening effect was done during the ML reconstruction, the reconstructed density profiles resemble the true density profiles more closely than the corresponding FBP profiles shown in Fig 10.9. Also the quality of the ML reconstruction from 10 projections is not inferior to FBP results obtained from data sets that are five times larger. A proper statistical treatment extracts more information from the measured data than the standard reconstruction methods do. In this way, the measurement time can be significantly reduced without loss of



**Fig. 10.10.** Reconstructions of the sample of Fig. 10.8 from only 10 (!) projections; panels (a) and (b) correspond to slices shown in Figs. 10.8b and 10.8c, respectively. Left: ML reconstructions; middle: ML profiles; right: standard FBP interpretation of the same data is shown for comparison.

resolution. Still better results can be expected provided the beam hardening is incorporated into the physical model.

## 10.4 ML neutron phase tomography

### 10.4.1 Experimental setup

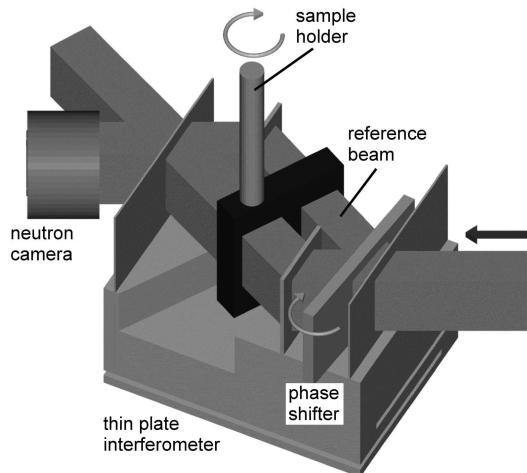
The absorption tomography discussed in the previous section provides only partial information about the object of interest. Optical properties of objects are more completely described by the distribution of the complex index of refraction  $n(x, y)$ , whose real part called simply index of refraction, and imaginary part called index of absorption are responsible for the phase shift and absorption of the illuminating beam, respectively. It is clear that for the 3D visualization of the index of refraction some kind of interferometric technique is needed. In a sense, phase tomography is a combination of both the previously mentioned inverse problems — the phase estimation discussed in Sec. 10.2 and tomographic imaging discussed in Sec. 10.3.

As has been already mentioned, in neutron optics one is often confronted with low count numbers because the phase space density of present neutron beams is 30 orders of magnitude below that of laser beams and many orders below X-ray sources. This intensity problem has dramatically arisen in the recently developed neutron phase contrast tomography (nPCT) [49]. PCT was originally invented in X-ray tomography with much higher coherent intensities available [50–52]. In order to utilize nPCT it is necessary to

develop an advanced reconstruction technique, which can be applied to very low count numbers. For instance, the typical count number in present nPCT setups is around 200 n/2h in a  $50 \times 50 \mu\text{m}^2$  pixel. In principle, focusing techniques, e.g. asymmetric Bragg reflections, can enhance the density of quasi-monochromatic neutron beams, but such hypothetic gains will be used to reduce the measurement time rather than to raise the count numbers. Therefore the low numbers of detected neutrons have to be accepted as the limiting factor of nPCT.

A strong motivation for developing nPCT is its extreme sensitivity, which is at least three orders of magnitude higher than in the conventional absorption tomography. The nPCT method proves its strength in extreme applications where other methods fail: (i) 3D investigation of non or weak-absorbing substances, (ii) analysis of isotope distributions with high sensitivity, (iii) investigation of magnetic domains in bulk materials [53], and (iv) energy and momentum exchange free analysis of magnetic (axial) and scalar potentials.

The experimental setup of nPCT is schematically shown in Fig. 10.11. The sample is inserted into one arm of a perfect crystal interferometer while



**Fig. 10.11.** Scheme of an nPCT experiment.

an object of known characteristics placed in the other arm compensates the large overall phase shift introduced by the thick sample. The output beam is then registered by a CCD camera with the spatial resolution of  $50 \mu\text{m}$ . Like in the absorption tomography, the sample is rotated around the vertical axes and up to several tens of scans are registered. Angle  $\varphi$  together with the position  $h$  of a CCD pixel specify the path of the particles registered by that pixel through the sample, see Fig. 10.5. For the sake of brevity, they will again be represented by a single collective index  $j$ .

To get an unambiguous value of the reconstructed phase, a set of auxiliary phases  $\delta_j$ , controlled by the experimenter, is needed for the estimation. They are provided by an auxiliary phase shifter, see Fig. 10.11, and the resulting interferograms are simply called scans or phase projections [54]. In nPCT their number should be chosen as less as possible so as to minimize the measurement time.

The conceptual difference between the standard deterministic and ML statistical inversions can be nicely illustrated on the example of nPCT. Standard nPCT consist of two separate steps: First, each set of interferograms is processed to get the distribution of the total phase accumulated in the  $j$ -th scan; this total phase is an integral of the unknown index of refraction along the  $j$ -th beam path. In this way, the original problem is reduced to the conventional tomographic imaging, and hence the inverse Radon transformation of the accumulated phase yields the distribution of the index of refraction within the object.

This straightforward procedure suffers from all known shortcomings of the standard FBP routine, which are further accented in nPCT by still smaller measured intensities and thus increased fluctuations. Additional problems arising in the standard nPCT are caused by using the accumulated phases as the starting point for tomographic reconstruction. Due to the high phase sensitivity it is very likely that some of the projected (accumulated) phases will exceed  $2\pi$  phase interval. The correction of these “phase jumps” becomes virtually impossible if the data are very noisy.

Most of these problems can be avoided when the ML statistical inversion is adopted. Here, the object is decomposed into many elementary cells each shifting the phase of the overlapping beam by only a fraction of the  $2\pi$  phase window. Thus the phase value of each cell is uniquely defined. Then we look for such distribution of those elementary phases that is most likely from the point of view of registered data. In this way, phase estimation and tomographic reconstruction are done simultaneously.

#### 10.4.2 ML reconstruction algorithm

Let us first discuss phase estimation in the context of nPCT. Consider an interferometric measurement with mean intensity  $N$  and amplitude  $V$ . The interference pattern,

$$\bar{n}_\alpha = N + V \cos(\theta + \delta_\alpha), \quad (10.29)$$

will be scanned with  $L$  different settings of the auxiliary phase shifter uniformly distributed over the  $2\pi$  phase window,

$$\delta_\alpha = \alpha \frac{2\pi}{L}, \quad \alpha = 0, 1, \dots, L - 1. \quad (10.30)$$

Interference pattern Eq. (10.4) is just a special case of Eq. (10.30) corresponding to  $L = 4$  phase shifts.

When the measurement is over, all the accumulated information can be expressed as an *a-posteriori* likelihood function. It is essential that the likelihood includes all measured data, and, together with the physical model for the detection probabilities, all experimental evidence. Thus, the likelihood is the optimum starting point for a complete tomographic analysis.

Ideally, the only fluctuating quantity in the tomographic measurement is the counted number of particles. The fluctuations produced by thermal sources such as nuclear reactors are well described by the Poissonian statistics, which was confirmed in several experiments [45,55]. Since the detections with different settings  $\delta_\alpha$  are independent, the joint probability of registering data  $\mathbf{n}$  is simply a product

$$\mathcal{L} = \prod_{\alpha} (\bar{n}_\alpha^{n_\alpha} e^{-\bar{n}_\alpha} / n_\alpha!). \quad (10.31)$$

This is also the likelihood of the given value  $\theta$  of the unknown phase shift.

In accordance with the maximum-likelihood principle we will take the maximum likely phase as the inferred value of  $\theta$ . Since the amplitude  $V$  and the total mean number of particles  $N$  are not under experimenter's control and may vary from one pixel to another, these parameters should be estimated together with phase. Their values are found by maximizing function (10.31), or its logarithm. The latter in the case of uniformly distributed auxiliary phase shifts simplifies to

$$\log \mathcal{L} \propto \sum_{\alpha} n_{\alpha} \log[N + V \cos(\theta + \delta_{\alpha})] - NL + \text{const..} \quad (10.32)$$

Now it is convenient to introduce new variables  $x = V \cos \theta$  and  $y = V \sin \theta$ . We are looking for the point where the likelihood has zero slope:  $\partial \mathcal{L} / \partial x = \partial \mathcal{L} / \partial y = \partial \mathcal{L} / \partial N = 0$ . From (10.32) we get the following extremal equations:

$$\sum_{\alpha} \frac{n_{\alpha} \mathbf{X}}{N + x \cos \delta_{\alpha} - y \sin \delta_{\alpha}} = 0, \quad \mathbf{X} = (\cos \delta_{\alpha}, \sin \delta_{\alpha}, 1). \quad (10.33)$$

In general, these equations must be solved numerically. Closed-form solution can be found only in some special cases such as  $\alpha = 3$  and  $\alpha = 4$ .

A particularly simple solution exists for three auxiliary phases, when the maximum of the Poissonian likelihood (10.31) coincides with the maximum of its Gaussian approximation,

$$\mathcal{L} \propto \exp \left[ -\frac{1}{2\sigma^2} \sum_{\alpha} (n_{\alpha} - \bar{n}_{\alpha})^2 \right], \quad (10.34)$$

yielding

$$x = \frac{2}{L} \sum_{\alpha=0}^2 n_{\alpha} \cos(-\delta_{\alpha}), \quad y = \frac{2}{L} \sum_{\alpha=0}^2 n_{\alpha} \sin(-\delta_{\alpha}), \quad N = \frac{1}{L} \sum_{\alpha=0}^2 n_{\alpha}. \quad (10.35)$$

These, going back to the original variables, can be written in the following compact form [19]

$$V = 2|R|/L, \quad e^{i\theta} = R/|R|, \quad (10.36)$$

where

$$R = \sum_{\alpha=0}^2 n_{j\alpha} e^{-i\delta_{j\alpha}}. \quad (10.37)$$

This quantity can be interpreted as the first coefficient of the discrete Fourier transformation of the registered counts  $n_{j\alpha}$  [23].

Therefore it is particularly useful to use three auxiliary phase shifts, for in that case, the optimal phase estimation is easily handled by means of the simple formula (10.36).

In nPCT, phase sensitive data  $n_{j\alpha}$  are registered. Subscripts  $j$  and  $\alpha$  label scans (i.e. pixels of the CCD camera and rotations of the sample) and auxiliary phases, respectively. As each scan contributes likelihood (10.34) and different scans are independent observations, the total log-likelihood reads

$$\log \mathcal{L} \propto \sum_j \sum_{\alpha} (n_{j\alpha} - \bar{n}_{j\alpha})^2 + \text{const.}, \quad (10.38)$$

where the mean number of particles detected in the  $j$ -th projection is given by

$$\bar{n}_{j\alpha} = N_j + V_j \cos(\theta_j + \delta_{j\alpha} + \theta_j^r). \quad (10.39)$$

Here  $\theta_j$  is the total phase accumulated along the  $j$ -th projection,  $\theta_j = \sum_i c_{ji} \mu_i$ . Coefficients  $c_{ji}$  are the overlaps between the  $j$ -th projection and the  $i$ -th elementary cell of the reconstruction mesh, as before, see Fig. 10.5, and  $\theta_j^r$  are the reference phases describing the phase properties of the empty interferometer. The latter can be estimated from the same set of projections measured without the sample. Likelihood (10.38) is to be maximized over the distribution  $\mu_i$  of the optical density of the sample.

In neutron phase imaging,  $\mu$  is composed of the sum of all scattering length densities ( $Nb$ ) of the isotopes contained in the sample:

$$\mu = -\lambda \sum_l N_l b_l = -\lambda \sum_l \frac{N_A \rho_l b_l}{A_l} \quad (10.40)$$

$N_l$  represents the number of isotopes  $l$  per unit volume,  $\rho_l$  the isotope density,  $N_A$  the Avogadro constant ( $6.02214199(47) \times 10^{23} \text{ mol}^{-1}$ ) and  $A_l$  the atomic weight. Parameter  $\lambda$  is the mean wavelength of the quasi-monochromatic beam, and  $b_l$  the coherent scattering length, which is a constant interaction parameter. Most isotopes have a positive coherent scattering length but some are known with negative  $b_l$ . The coherent scattering length is defined positive for repulsive optical potentials  $\bar{V} > 0$  with the index of refraction less than unity:  $n = \sqrt{1 - \bar{V}/E} = 1 - \lambda^2 Nb_l/2\pi$ . The existence of positive and negative

phase shifts is a speciality of neutron optics and can be utilized for fading out unwanted phase contributions.

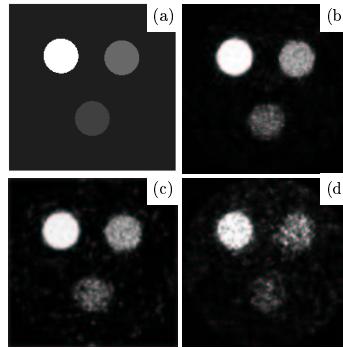
A necessary condition for the maximum of  $\log \mathcal{L}$ ,

$$\frac{\partial \log \mathcal{L}}{\partial \mu_i} = 0, \quad \forall \mu_i, \quad (10.41)$$

yields on using Eqs. (10.38) and (10.39) the following set of extremal equations,

$$\mu_i = \mu_i \frac{\sum_j V_j c_{ji} \sin \theta_j \text{Im}\{R_j\}}{\sum_j V_j c_{ji} \cos \theta_j \text{Re}\{R_j\}}. \quad (10.42)$$

These can be solved numerically by repeated iterations.

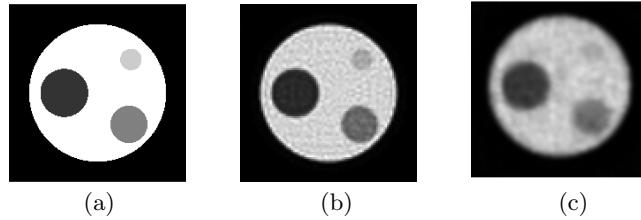


**Fig. 10.12.** Simulated phase tomography with a weak neutron signal. The maximal accumulated  $\theta_j$  in the three cylinders making up the object (a) are 150 deg., 50 deg, and 30 deg for white, light gray, and dark gray, respectively. The mean count numbers per pixel and visibilities are (b)  $N = 450$ ,  $V = 33\%$ ; (c)  $N = 150$ ,  $V = 33\%$ ; (d)  $N = 30$ ,  $V = 33\%$ .

Figure 10.12 shows a simulation of a nPCT experiment with various intensities and visibilities of the illuminating beam. The artificial object was scanned from 31 different angles with a resolution of 81 pixels.

Case (d) is the most interesting one. Here the incident beam has such a low intensity that its Poissonian fluctuations are comparable to the intensity changes caused by the maximal phase shifts in the light gray and dark gray cylinders. The useful phase information is thus almost lost in the background noise, yet all three cylinders nicely show in the reconstruction.

Another example of the ML phase tomography is shown in Fig. 10.13. The parameters of the simulated experimental setup are comparable to that of Fig. 10.12 but now the maximal phase shift in the sample is well in excess of  $4\pi$  radians. A priori knowledge about the shape and high index of refraction of the white container could be easily incorporated into the reconstruction, and as a result the internal structure of the object was nicely resolved.



**Fig. 10.13.** Simulated phase tomography with a weak neutron signal,  $N = 150$  and  $V = 33\%$ . The maximal accumulated  $\theta_j$  in the object is  $4.2\pi$  rad. (a) the artificial object; the ratio of the index of refraction in the white, light gray, dark gray, and black cylinder is  $1 : 0.8 : 0.5 : 0.2$ . (b) ML reconstruction from 31 angles and 81 pixels. (c) ML reconstruction from 21 angles and 41 pixels.

A similar isotope gauge, a mixture of S33+S34 and H<sub>2</sub>O+D<sub>2</sub>O isotopes, has recently been investigated at the nPCT setup ILL-S18 in Grenoble in order to verify the sensitivity and spatial resolution of present nPCT.

#### 10.4.3 Reference phase measurement

No interferometers are perfect. Already an empty interferometer shows a nonuniform transversal distribution of phase difference between its two arms. To get rid of this background phase the measurement is done in two steps: with and without the sample. The most simple way how to subtract the background phase is to perform two separate phase reconstructions, then subtract the reconstructed background phase from the reconstructed phase of the sample. This procedure is simple, but not optimal. It is not difficult to see why. The inspection of Eq. (10.42) shows that the reconstructed indexes of refraction  $\mu_i$  depend on the visibilities  $V_j$  of the registered interference fringes. This is natural since the reconstruction is a synthesis of many phase measurements and phases measured under higher visibility are less affected by the noise and hence more credible. For the same reason the phase introduced by an empty interferometer is measured more accurately than phase introduced by interferometer **and** sample. This additional knowledge should be incorporated into the reconstruction routine. Denoting  $\theta = \theta_s + \theta_r$  the total phase measured with the sample, which is the sum of the reference phase and phase introduced by the sample  $\theta_s$ , and using Eqs. (10.36) and (10.37) in Eq. (10.34) we can rewrite the posterior distributions of  $\theta$  and  $\theta_r$  in the following compact form:

$$P(\theta) \propto e^{V \cos(\theta - \theta_{\text{NFM}})} \quad (10.43)$$

$$P(\theta_r) \propto e^{V_r \cos(\theta_r - \theta_{r,\text{NFM}})}. \quad (10.44)$$

NFM denotes Gaussian (semi-classical) phase estimates that maximize posterior distributions (10.43) and (10.44), which are also known as von Mises

normal distributions defined on the unit circle. As has been mentioned above their widths are determined by visibilities. As we are interested only in phase  $\theta_s$  introduced by the sample alone, let us calculate its posterior distribution,

$$P(\theta_s) = \iint P(\theta)P(\theta_r)\delta(\theta - \theta_r - \theta_s)d\theta_r d\theta. \quad (10.45)$$

The double integrations can be easily carried out, and the result expressed in terms of the Bessel function  $I$ ,

$$P(\theta_s) \propto I_0 \left( \sqrt{V^2 + V_r^2 + VV_r \cos(\theta_s - \theta_{-,NFM})} \right), \quad (10.46)$$

where  $\theta_{-,NFM} = \theta_{NFM} - \theta_{r,NFM}$ . Optical density of the sample can now be estimated by maximizing the posterior distribution Eq. (10.46) with respect to indexes  $\mu_i$ . Such procedure accounts for the reference phases in an optimal way.

#### 10.4.4 Beyond phase tomography

Phase tomography discussed in the previous section is a simple example of the quantum-process tomography. In this case the process is a unitary operation that can be represented by a 2D rotation. The rotations vary in space, and the detected particles probe their spatial distribution via quantum interference.

Obviously, to “see” more complicated transformations, more complex probes are needed. It would be logical to proceed the exposition with the case of spin 1/2 systems whose Hilbert space could be associated with the 3D Bloch sphere. The synthesis of measurements - projections, is capable to determine an unknown quantum state, and such procedure posses all the features of quantum tomography of the internal spin state. ML approach according to the general receipt given in chapter by Hradil *et al.* in this volume can be straightforwardly applied here. This was worked out explicitly in Ref. [56] for the representation of the spin using the polarization vector.

Since magnetic momentum of particles is coupled to the magnetic field, spin state tomography can be used as a means e.g. for non-destructive investigations of the magnetic domains of bulk materials. Here the beam of polarized neutrons goes through the specimen interacting with its magnetic domains. Due to this interaction, the spin state of neutrons changes, and this depolarization can be utilized for the visualisation of the magnetic domains in the specimen just like in the case of phase tomography. However, the probed operations are now represented by rotations in 3D space, which unlike 2D rotations in phase tomography form a non-Abelian group of transformations. Consequently, the tensorial character of the corresponding depolarization observable together with the non-commutativity of rotation matrices makes the analysis rather involved. Though this is a challenging question with potential interesting applications, at present no analytical solution of this inverse scattering problem is known. Up to now there are several approaches based on deterministic techniques, but ML solution has not been devised yet.

## 10.5 Maximum-likelihood characterization of photonic qutrits

Apart from using spin 1/2 particles as probes for sensing and visualization of physical fields they can also be utilized for the physical realization of qubits in the newly emerging field of quantum information processing. Much attention has recently been turned to exploring the possibilities of applying quantum systems in communication and computing protocols. Usually, these protocols use the information encoded in such bidimensional systems. Nevertheless, some proposals show that higher dimensional systems are better suited for some purposes [57–64]. On a more fundamental level, higher dimensional spaces provide with counter-intuitive examples of the relationship between the quantum and the classical information, which cannot be found in two-dimensional systems [65].

Recently, some experimental implementations of higher dimensional quantum systems have appeared. Given the high technical status of photonic technology and the fact that photons will for a long time remain the only means for quantum communication, it is desirable to have qunits (systems with  $n$  different orthogonal states) to be carried by photons. Encoding qunits with photons has already been experimentally demonstrated using interferometric techniques, with time-bin schemes [66] and superpositions of spatial modes [67]. Up to now, the only non-interferometric technique of encoding qunits in photons is using the orbital angular momentum (OAM) of the photons or, equivalently, the transversal modes [68–74].

Before we apply ML technique to the reconstruction these higher dimensional states let us first review some basic facts about the quantum states of orbital angular momentum degrees of freedom of light.

On theoretical grounds, one convenient basis which describes the transversal modes of a light beam fulfilling the paraxial approximation is the Laguerre-Gaussian (LG) functions basis:  $LG_{p,m}(x, y)$ . Here  $m$  is the order of the phase dislocation characteristic of this set of functions and it accounts directly for the orbital angular momentum of the Laguerre-Gaussian mode in units of  $\hbar$  [75–77]. The other parameter  $p$  is a label which is related to the number of radial nodes of the mode and  $(x, y)$  refer to any point in a plane perpendicular to the beam propagation direction. The LG functions form a complete and orthonormal basis for any complex function in the transversal plane.

Holographic techniques, among other schemes, can be used to transform  $LG$  modes [78, 79]. Conveniently prepared holograms change the phase structure of the incoming beam, adding or removing the phase dislocations related with the orbital angular momentum. Whereas optical single mode fibers act as a filter for all higher  $LG$  modes, i.e. only the  $LG_{00}$  can be transmitted, the combination of holograms and single mode fibers project the incoming photon into different states. In this way we can define the basis of the experimentally accessible states as:

$$\langle \mathbf{x}|0\rangle = LG_{0,0}(x, y), \quad |m\rangle = H_m(\mathbf{0})|0\rangle, \quad (10.47)$$

where the vector  $|0\rangle$  represents the effect of using only a single mode fiber to detect the photon,  $\mathbf{x}$  is a shortcut to represent any point in the transversal space<sup>5</sup>,  $m$  is a positive or negative integer, and  $H_m(\mathbf{0})$  is the operator which describes the action of the  $m$ -th order hologram when it's centered, relative to the fiber. Explicitly, the expression of this operator is

$$\langle \mathbf{x}_1|H_m(\mathbf{0})|\mathbf{x}_2\rangle = \exp(-im \arctan(\frac{y_1}{x_1}))\delta(x_1 - x_2, y_1 - y_2). \quad (10.48)$$

Albeit by construction, any of the modes in our basis posses orbital angular momentum of  $m$ , they are not pure  $LG$ -modes. However, they can be described as coherent superpositions of different modes with the same  $m$ , but different  $p$ 's. In this sense, the base we have constructed in (10.47) is orthonormal<sup>6</sup>, although not complete, since it does not expand the complete  $LG$  basis. In the following we refer to all modes belonging to the subspace (10.47) as "inner" modes and the rest of the modes will be addressed as "outer" modes.

Thus, any displaced hologram and, in general, any linear operator which acts on our Hilbert space can be expressed like,

$$H_m(a, b) = \sum_{i=-\infty}^{+\infty} c_i(a, b) H_i(0) + \gamma(a, b) \Gamma, \quad (10.49)$$

where  $a, b$  are the displacements of the hologram relative to its centered position. The operator  $\Gamma$  accounts for the possibility that the displaced hologram is performing transformations between "outer" and "inner" states, i.e. transforming any "inner" state into an "outer" one, or the other way round.

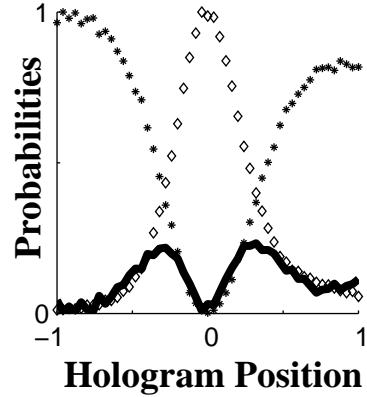
The value of  $\gamma$  can be estimated by

$$|\gamma(a, b)|^2 \sim 1 - \sum_{i=-\infty}^{+\infty} |\langle i|H_m(a, b)|0\rangle|^2, \quad (10.50)$$

which is a feasible observable. In Fig. 10.14 we present an example of such a measurement. The data was taken by preparing a photon in the  $|0\rangle$  state, and transforming its state by means of a hologram of order  $m = +1$ , which could be displaced along the horizontal  $x$ -axis. Afterwards, the transformed photon was projected onto three dimensional OAM Hilbert space consisting of  $| -1 \rangle$ ,  $| 0 \rangle$  and  $| 1 \rangle$ . The other  $|m\rangle$  states have a much smaller contribution than the one due to "outer" states. In Fig.10.14 there are two positions where the contribution of the "outer" modes is specially high. These positions roughly correspond with an equal superposition of the two higher modes, i.e.  $|c_0| = |c_1|$ .

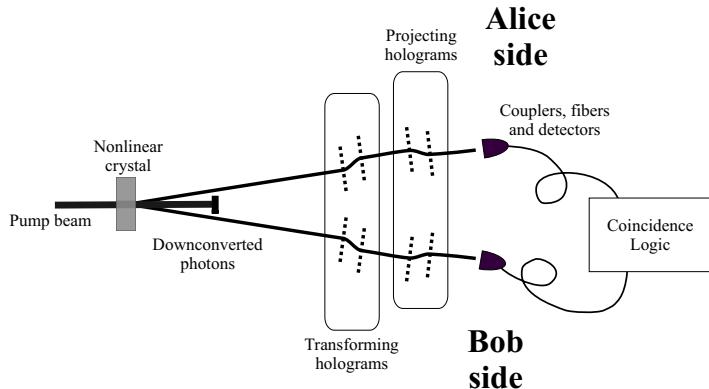
<sup>5</sup> The transversal position of a photon can only be accounted under the paraxial approximation.

<sup>6</sup> The normality of the functions comes from the fact that  $\langle 0|0\rangle = 1$  and that the hologram operations are unitary.



**Fig. 10.14.** Experimentally measured transfer of the  $|0\rangle$  mode to “outer” modes depending on the position of the hologram. The solid line represents the measured normalized value of  $1 - \sum_{i=-1}^{+1} |\langle i|H_m(x,0)|0\rangle|^2$ , where  $x$  is the position of the hologram in the  $x$ -axis. Other  $|m\rangle$  modes only contribute slightly to this value. The dashed lines represent the normalized measured values of  $|\langle 0|H_m(x,0)|0\rangle|^2$  and  $|\langle 1|H_m(x,0)|0\rangle|^2$ , and are presented for the sake of comparison. The maximum transfer to “outer” modes happens around the position where  $|\langle 0|H_m(x,0)|0\rangle|^2 = |\langle 1|H_m(x,0)|0\rangle|^2$ .

### 10.5.1 Experimental set-up



**Fig. 10.15.** Experimental set-up. A 351nm wavelength laser pumps a BBO crystal. The two generated 702nm down-converted photons are send to Alice and Bob’s sides respectively. Before being detected each photon propagates through four different holograms.

The experimental set-up is shown in Fig. 10.15. A 351nm wavelength Argon-ion laser pumps a 1.5-mm-thick BBO ( $\beta$ -barium-borate) crystal cut for Type I phase matching condition. The crystal is positioned to produce down-converted photons at a wavelength of 702nm emitted at an angle of  $4^\circ$  off the pump direction. These photons are directly entangled in the orbital angular momentum degree of freedom. In the actual protocol, an observer which we call Alice, possess the source and can manage one of the down-converted photons, meanwhile the other downconverted photon is sent to another observer, which, following a long tradition, we call Bob. Before being detected, each of the two down-converted photons traverses four holograms which we divide in two sets. Each set consists of one hologram with charge  $m = 1$  and another with charge  $m = -1$ . The first set of holograms provides the means of a rotation in the three dimensional space expanded by the states  $| -1 \rangle$ ,  $| 0 \rangle$ , and  $| 1 \rangle$ . The second set, together with a single mode fibers and a detector, act as projectors onto the three different basis states.

In the experiment, the tomographic reconstruction of Bob's received qutrit state was realized in two independent steps trying to avoid any bias from ‘a priori’ information. First, the team from University of Vienna projected the photons in Bob side. The minimum number of measurements to reconstruct the three dimensional state is 9. This number increases to 121 because of our enlargement to a 11-dimensional Hilbert space. In the end, to explode the power of the ML reconstruction and to minimize errors, the number of different projections was set around 2400.

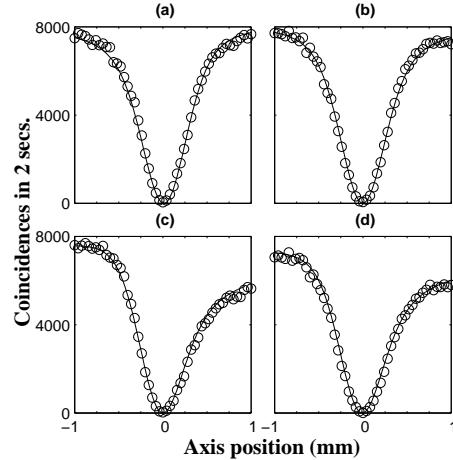
As will be shown below in all the cases, the reconstructed three dimensional state was a coherent superposition of the three “inner” vectors, which relative weights and phases could be effectively controlled. The noise and incoherence corresponded to a probability of less than 2%.

### 10.5.2 Experimental measurements

Prior to the measurement taking procedure, the rotating set of holograms was analyzed to properly describe the transformation done. From the description of each single hologram, we could express the action of each transformation set in the following way:

$$\begin{aligned} \langle \mathbf{x}_1 | H_1(a_{+1}, b_{+1}) H_{-1}(a_{-1}, b_{-1}) | \mathbf{x}_2 \rangle &= \exp(-i \arctan(\frac{y_1 - a_1}{x_1 - b_1})) \\ &\quad + i \arctan(\frac{y_1 - a_{-1}}{x_1 - b_{-1}}) - ik_x x_1 - ik_y y_1) \delta(x_1 - x_2, y_1 - y_2), \end{aligned} \quad (10.51)$$

where  $k_x$  and  $k_y$  are free parameters which depend on the alignment procedure, and  $a_{\pm 1}, b_{\pm 1}$  represent the displacements of the two holograms. Each set of holograms is then described completely by eight parameters: the number of maximum coincidences, the width of the beam, four numbers to determine the centered position of each hologram and the two parameters  $k_x$  and  $k_y$ .



**Fig. 10.16.** Curve fitting for a set of holograms. First row and second row correspond to a hologram of charge  $m = +1$  and  $m = -1$  respectively. Left and right columns correspond to scans of the  $x$  and  $y$  axis respectively. Circles are experimental measurements. Solid line correspond to the estimated result using Eq. (10.51) and  $|0\rangle$  as initial and final states.

The estimation of these eight parameters was performed by fitting four different experimental curves as shown in Fig. 10.16. The data which conformed the curves were taken by sending to Bob a photon prepared in the  $|0\rangle$  state. Bob fixed one of his holograms in one determined position and performed a scan on one of the axis of the other hologram. The resulting state was again projected to the  $|0\rangle$  state, i.e. one of this curves can be described with the following number  $\langle 0|H_1(x, 0)H_{-1}(1, 1)|0\rangle$  as a function of  $x$ . Each of the four curves corresponds to the scan of the four different axis of the holograms.

The projections measurements were made by moving the set of two holograms into around 2400 different positions and counting the number of coincident detections which took place in two seconds. The complete time for each of this measurements was around six hours, due mainly to the software to access the different motors. After this time some slight misalignments where detected, which could be compensated mainly due to the large number of different projections taken and the reconstruction process.

### 10.5.3 Results

The registered data were processed using the maximum-likelihood reconstruction algorithm. Notice that the measurement is incomplete,  $\sum_j |j\rangle\langle j| \neq 1$ , where  $|j\rangle = H_1(a_{+1}^j, b_{+1}^j)H_{-1}(a_{-1}^j, b_{-1}^j)|0\rangle$  are the projections done. This has been taken into account, see Hradil *et al.* in this volume for more details.

From the technical point of view, the ML reconstruction was done by iterating the extremal equation [80],

$$R\rho R = G\rho G, \quad R = \sum_j \frac{n_j}{p_j} |j\rangle\langle j|, \quad G = \frac{\sum_j n_j}{\sum_j p_j} \sum_j |j\rangle\langle j|, \quad (10.52)$$

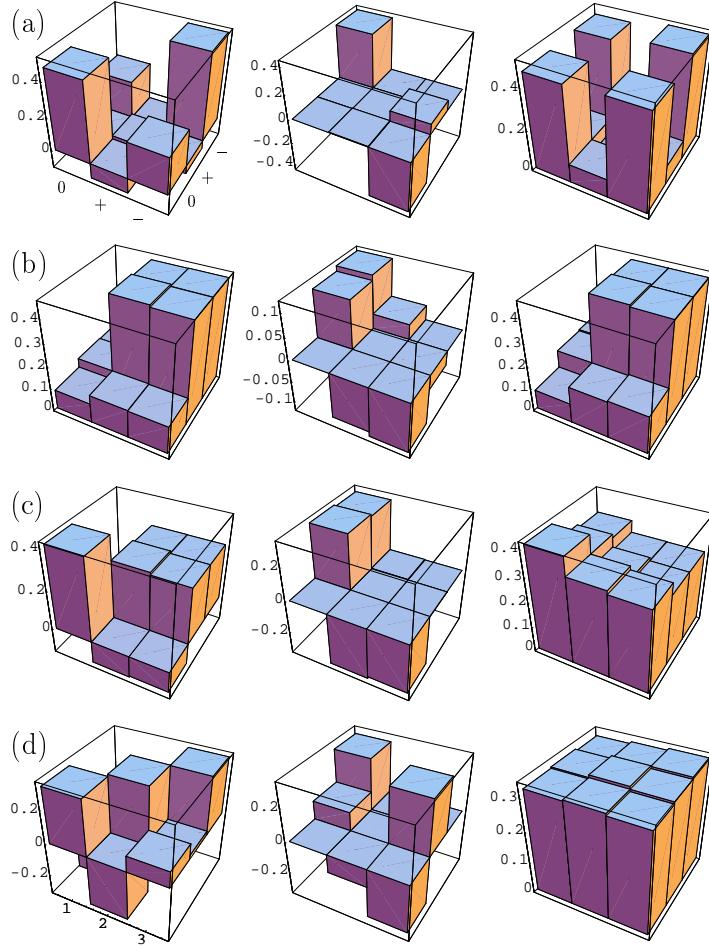
starting from the maximally mixed state. Here  $n_j$  is the number of photon pairs found in state  $|j\rangle$  and  $p_j = \text{Tr}\{|j\rangle\langle j|\rho_{\text{Bob}}\}$  are the corresponding probabilities.

Let us mention that though the reconstruction is done on the full  $3 + 8$  dimensional Hilbert space spanned by the “inner” and “outer” states, we are interested only in the former subspace. Therefore all the reconstructed states are projected to this subspace to simplify the discussion.

As we have already mentioned, different projections done by Alice translate through the entanglement to different state preparations on Bob’s side. Four such remote preparations are shown in Fig. 10.17. All of them were found to be very nearly pure states, their largest eigenvalues and corresponding eigenvectors being (a)  $\lambda_{\max} = 0.99$ ,  $|e_{\max}\rangle = (0.31 + 0.61i)|0\rangle + 0.71|-1\rangle$ ; (b)  $\lambda_{\max} = 0.97$ ,  $|e_{\max}\rangle = (0.2 + 0.16i)|0\rangle + 0.68|1\rangle + (0.68 - 0.05i)|-1\rangle$ ; (c)  $\lambda_{\max} = 0.99$ ,  $|e_{\max}\rangle = 0.65|0\rangle + (-0.18 - 0.52i)|1\rangle + (-0.18 - 0.5i)|-1\rangle$ ; (d)  $\lambda_{\max} = 0.99$ ,  $|e_{\max}\rangle = (-0.54 + 0.2i)|0\rangle + 0.58|1\rangle + 0.58i|-1\rangle$ . In case (a) Alice tried to prepare an equal-weight superposition of  $|0\rangle$  and  $|-1\rangle$  basis states. Utilizing the conservation of the orbital momentum in downconversion, this was easily done by projecting her qutrit along the ray  $|0\rangle + |1\rangle$ : Her hologram with the positive charge was taken out of the beam path and the center of the other other one was displaced with respect to the beam by a calculated translation vector. In case (b) the remotely prepared state supposed to be an equal-weight superposition of  $|-1\rangle$  and  $|1\rangle$  states. Here both holograms had to be used to remove the unwanted  $|0\rangle$  contribution via destructive interference. Due to the greater complexity of this operation the acutually prepared state was found to still contain a small contribution of the  $|0\rangle$  component. In cases (c) and (d) the goal was to prepare an equal-weight superpositions of all three basic “inner” states with different phases. The tomography reveals that we have come very close to achieving this goal. The  $|0\rangle$  component of state (c) had only a slightly greater weight than the other two, while state (d) was almost perfect.

## 10.6 Maximum-likelihood quantum homodyne tomography

In the final section of this chapter we discuss applications of the likelihood maximization technique to quantum homodyne tomography. Homodyne characterization of a quantum state of an optical mode is based on multiple phase-sensitive measurements of its electric field. The field quantum noise statistics



**Fig. 10.17.** Results of quantum state tomography applied to three different remotely prepared states of Bob’s qutrits. Left and middle panels show real and imaginary parts of the reconstructed density matrices; right panels show the absolute values of those elements for getting picture of how large the contributions are of the three basic states.

collected at various optical phases provide enough information to reconstruct the density matrix of the ensemble of interest.

Since its proposal [82] and first experimental implementation [83] in the early 1990s, quantum homodyne tomography has become a robust and versatile tool of quantum optics and has been applied in many different experimental settings. As any statistical method, it is suitable for combining with the likelihood maximization technique. Indeed, a homodyne measurement run yields a set of statistical data, whose likelihood with respect to a

particular quantum ensemble can be evaluated easily. The goal of this section is to describe the way of finding the hypothetical ensemble which maximizes the probability of acquiring a particular data set. In writing this section we have assumed the reader to be familiar with the basics of quantum homodyne tomography; otherwise we recommend to refer to chapter by Raymer and Beck of this volume for a detailed review.

We consider a homodyne tomography experiment performed on an optical mode prepared repeatedly in the same quantum state  $\hat{\rho}$ . In an experimental run one measures the field quadrature at various phases of the local oscillator. Each measurement is associated with the observable  $\hat{X}_\theta = \hat{X} \cos \theta + \hat{P} \sin \theta$ , where  $\hat{X}$  and  $\hat{P}$  are the canonical position and momentum operators and  $\theta$  is the local oscillator phase.

For a given phase  $\theta$ , the probability to detect a particular quadrature value  $x$  is given by

$$\text{pr}_\theta(x) = \text{Tr}[\hat{\Pi}(\theta, x)\hat{\rho}], \quad (10.53)$$

where  $\hat{\Pi}(\theta, x) = |\theta, x\rangle\langle\theta, x|$  is the projector onto this quadrature eigenstate. In the Fock (photon number state) basis, the projection operator is expressed as

$$\Pi_{mn}(\theta, x) = \langle m | \hat{\Pi}(\theta, x) | n \rangle = \langle m | \theta, x \rangle \langle \theta, x | n \rangle, \quad (10.54)$$

where the overlap between the number and quadrature eigenstates is given by the well known stationary solution of the Schrödinger equation for a particle in a harmonic potential:

$$\langle m | \theta, x \rangle = e^{im\theta} \left( \frac{2}{\pi} \right)^{1/4} \frac{H_n(x)}{\sqrt{n!}} \exp(-q^2), \quad (10.55)$$

with  $H_n$  denoting the Hermite polynomials<sup>7</sup>.

Consider an experimental run consisting of  $N$  field quadrature measurements at various local oscillator phases  $\theta_i$ . For a given ensemble, the likelihood of obtaining a specific set of results  $\{x_i\}$  is the product of individual probabilities (10.53) for each result

$$\ln \mathcal{L} = \sum_i \ln \text{pr}_{\theta_i}(x_i). \quad (10.56)$$

We now solve the inverse problem of estimating the ensemble  $\hat{\rho}$  that maximizes the likelihood of a given data set  $\{\theta_i, x_i\}$  by means of the iterative algorithm described in chapter by Hradil *et al.* We apply the iteration operator

---

<sup>7</sup> The additional phase factor  $e^{im\theta}$  originates from the properties of the phase-space rotation operator [81]  $\hat{U}(\theta) = e^{-i\theta\hat{n}}$ . From  $\hat{U}^\dagger(\theta)\hat{a}\hat{U}(\theta) = \hat{a}e^{-i\theta}$  we find for the quadrature operator  $\hat{U}^\dagger(\theta)\hat{X}\hat{U}(\theta) = \hat{X}_\theta$  and for its eigenstate  $|\theta, x\rangle = \hat{U}^\dagger |0, x\rangle$ . From the first and last relations above, we obtain  $\langle m | \theta, x \rangle = e^{i\theta m} \langle m | 0, x \rangle$ . The quantity  $\langle m | 0, x \rangle$  is the energy eigenwavefunction of a harmonic oscillator.

$$\hat{R}(\hat{\rho}) = \sum_i \frac{\hat{\Pi}(\theta_i, x_i)}{\text{pr}_{\theta_i}(x_i)}, \quad (10.57)$$

to the density matrix according to Eq. (3.40) of chapter by Hradil *et al.* so each step preserves the positivity of the density matrix:

$$\hat{\rho}^{k+1} = \mathcal{N} \left[ \hat{R}(\hat{\rho}^k) \hat{\rho}^k \hat{R}(\hat{\rho}^k) \right], \quad (10.58)$$

where  $\mathcal{N}$  denotes normalization to a unitary trace.

In practice, the iteration algorithm is executed in the photon number (Fock) representation. Since the Hilbert space of optical states is of infinite dimension, the implementation of the algorithm requires its truncation so the Fock terms above a certain threshold are excluded from the analysis. This assumption conforms to many practical experimental situations in which the intensities of fields involved are *a priori* limited.

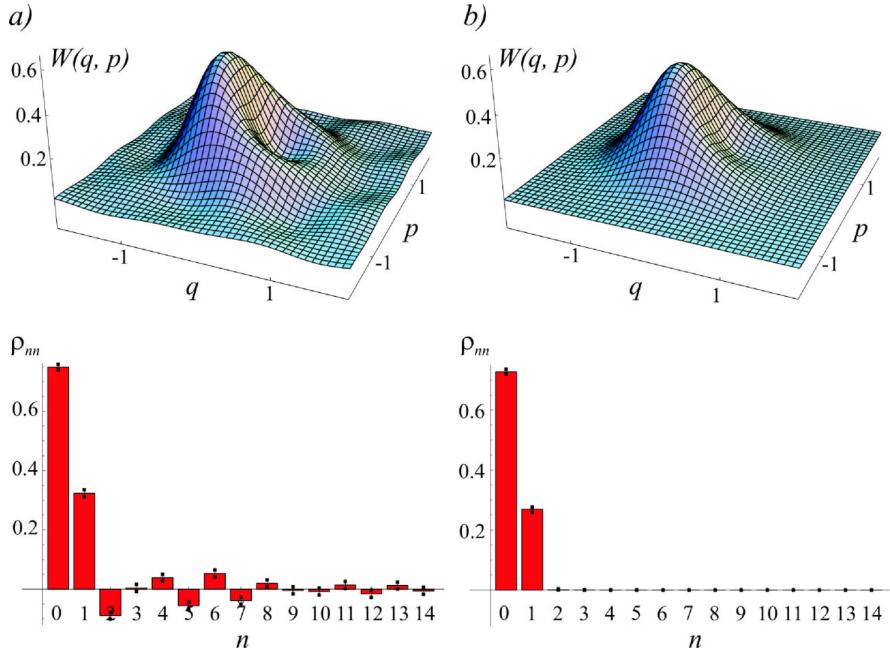
We note that the scheme above does not involve evaluating marginal distributions, i.e. histograms of the quadrature data  $\text{pr}_\theta(x)$  associated with particular phases. Elimination of this intermediate step from the reconstruction scheme allows one to avoid approximating the phase and quadrature values and thus further enhance the accuracy of the method.

It is instructive to compare the maximum likelihood quantum state estimation with the traditional methods: reconstruction of the Wigner function by means of the inverse Radon transformation [81] and evaluation of the density matrix using quantum state sampling (see Raymer and Beck in this volume and Refs. [84, 85]). Fig. 10.18 shows application of these two techniques to the experimental data from Ref. [86]. The data set consists of 14152 quadrature samples of an ensemble approximating a coherent superposition of the single-photon and vacuum states.

The reconstruction shown in Fig. 10.18 reveals the advantages of the ML technique in comparison with the standard algorithm. First, the finite amount and discrete character of the data available leads necessarily to statistical noise which prevents one from extracting complete information about a quantum state of infinite dimension. To deal with this issue, both techniques apply certain assumptions on the ensemble to be reconstructed. While the ML algorithm truncates the Fock space, the filtered back-projection imposes low pass filtering to the Fourier image of the Wigner function<sup>8</sup>, i.e. assumes the ensemble to possess a certain amount of “classicality” [82]. The latter assumption is dictated by mathematical convenience and is much less physically founded than the former. The ripples visible in the Wigner function reconstruction in Fig. 10.18(a) are a direct consequence of statistical noise and are associated with the unphysical high number terms in the density matrix. Such ripples are typical of the inverse Radon transformation [88].

---

<sup>8</sup> The pattern function reconstruction of the density matrix is free form this drawback as it does not involve spectral filtering and relies on truncating the Fock space instead.



**Fig. 10.18.** Estimation of an optical ensemble from a set of 14152 experimental homodyne measurements [86] by means of the inverse Radon transformation (a) and the likelihood maximization algorithm (b). The Wigner function and the diagonal elements of the reconstructed density matrix are shown. The inverse Radon transformation in (a) was performed by means of the filtered back-projection algorithm with the cutoff frequency of 6.3. The statistical uncertainties in (b) were determined by means of a Monte-Carlo simulation (see text).

Second, the back-projection algorithm does not impose any a priori restrictions on the reconstructed ensemble. This may lead to unphysical features in the latter, such as negative diagonal elements of the density matrix in Fig. 10.18(a). The ML technique, on the other hand, allows to incorporate the positivity and unity-trace constraints into the reconstruction procedure, thus always yielding a physically plausible ensemble [89, 90].

A third important advantage of the ML technique is the possibility to incorporate the detector inefficiencies. In a practical experiment, the photodiodes in the homodyne detector are not 100% efficient, i.e. they do not transform every incident photon into a photoelectron. This leads to a distortion of the quadrature noise behavior which needs to be adjusted for in the reconstructed ensemble.

A common model for an homodyne detector of non-unitary efficiency  $\eta$  is a perfect detector preceded by a fictitious beam splitter of transmission  $\eta$ . The reflected mode is lost so the transmitted density matrix undergoes a

so-called generalized Bernoulli transformation [81]:

$$\langle m | \hat{\rho}_\eta | n \rangle = \sum_{k=0}^{\infty} B_{m+k,m}(\eta) B_{n+k,n}(\eta) \langle m+k | \hat{\rho}_0 | n+k \rangle, \quad (10.59)$$

where  $\hat{\rho}_0$  and  $\hat{\rho}_\eta$  are the density matrices of the original and transmitted ensembles, respectively, and  $B_{n+k,n} = \sqrt{\binom{n+k}{n} \eta^n (1-\eta)^k}$ . Under these circumstances the probability (10.53) of detecting a quadrature value  $x$  becomes

$$\begin{aligned} \text{pr}_\theta^\eta(x) &= \langle \theta, x | \hat{\rho}_\eta | \theta, x \rangle \\ &= \sum_{m,n=0}^{\infty} \sum_{k=0}^{\infty} B_{m+k,m}(\eta) B_{n+k,n}(\eta) \langle n | \theta, x \rangle \langle \theta, x | m \rangle \langle m+k | \hat{\rho}_0 | n+k \rangle, \end{aligned} \quad (10.60)$$

so the projection operator  $\hat{\Pi}(\theta, x)$  becomes replaced by a POVM element given by [91]

$$\hat{E}_\eta(\theta, x) = \sum_{m,n,k} B_{m+k,m}(\eta) B_{n+k,n}(\eta) \langle n | \theta, x \rangle \langle \theta, x | m \rangle | n+k \rangle \langle m+k |. \quad (10.61)$$

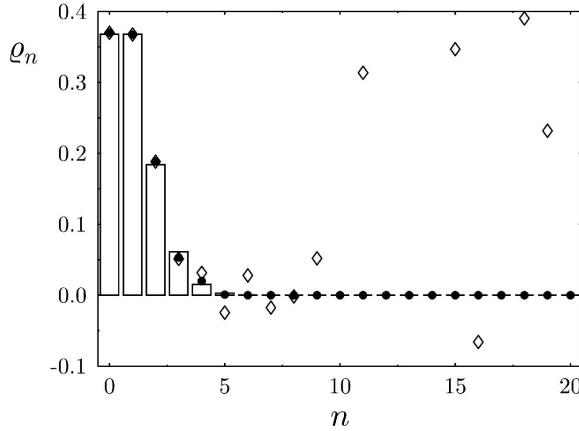
Aware of the homodyne detector efficiency  $\eta$ , one runs the iterative algorithm (10.57, 10.58) and reconstructs the original density matrix  $\hat{\rho}_0$ .

The performance of this technique is illustrated by Fig. 10.19 showing the reconstructed photon statistics of a coherent state measured with an inefficient detector. The estimated photon number distribution represents correctly the original state, in contrast to that obtained via the quantum state sampling followed by the inverted Bernoulli transformation [92].

Banaszek [93] has proposed an elegant modification of the ML state estimation method, applicable specifically to the evaluation of the Wigner functions. First, we notice that the Wigner function in the *origin* of the phase space can be determined using the diagonal elements of the density matrix in the Fock basis [94]:

$$W(0,0) = \sum_{n=0}^{\infty} (-1)^n \rho_{nn}. \quad (10.62)$$

These elements can be evaluated by means of the ML algorithm, simplified in the following manner. Since we are interested only in the diagonal elements of the density matrix, we set all off-diagonal elements of  $\hat{\rho}$  to zero. This new diagonal density matrix  $\hat{\rho}'$  corresponds to a optical ensemble which is symmetric with respect to phase rotations. The phase-sensitive off-diagonal elements of the projection operator (10.54) can therefore be neglected. Furthermore, the iteration step (10.58) can be reduced to  $\hat{\rho}^{k+1} = \mathcal{N}[\hat{R}(\hat{\rho}'^n)\hat{\rho}'^n]$  with no risk of arriving at a nonpositive density matrix.



**Fig. 10.19.** (courtesy K. Banaszek, Oxford) Loss corrected reconstruction of the photon number distribution of an  $\alpha = 1$  coherent state from  $10^5$  simulated homodyne data samples with the detector quantum efficiency  $\eta = 0.85$ . Filled circles represent the maximum likelihood estimation, hollow diamonds represent the quantum state sampling reconstruction followed by an inverse Bernoulli transformation.

Repetitive application of the iteration above reconstructs the photon distribution and the value of the Wigner function at the phase space origin. To evaluate the Wigner function  $W(p, q)$  at any *arbitrary* point, one applies the phase space displacement  $\hat{D}(-p, -q)$  to the ensemble  $\hat{\rho}$  so the point  $(p, q)$  gets transferred to the origin. In terms of the quadrature data, this corresponds to the shift

$$x_i \rightarrow x_i - q \cos \theta_i - p \sin \theta_i. \quad (10.63)$$

Upon this transformation, we apply the iterative algorithm again and calculate the Wigner function at the desired phase space location.

This scheme may appear more involved than the one discussed in the beginning of the section, as one needs to run a separate iteration series for every point in which the Wigner function is to be calculated. However, due to a smaller number of parameters and a simplified iteration step, each iteration takes less time and the series converges faster. The choice of a particular scheme depends on a specific task and on the chosen truncation threshold in the Fock space.

Finally, we address the issue of statistical errors in ML homodyne tomography. Hradil *et al.* in this volume described the way of estimating the error by diagonalizing the Fisher information matrix. Although this method is generally applicable here, it does not regard the density matrix positivity constraint. If one of the parameters lies near the boundary of the positivity region, the statistical error will be governed by the first rather than the sec-

ond derivative of the likelihood, and the actual uncertainty will be smaller than the one estimated.

A sensible alternative is offered by a clumsy, yet simple and robust technique of *simulating* the quadrature data that would be associated with the estimated ensemble  $\hat{\rho}_{ML}$  if it were the true state. One generates a large number of random sets of homodyne data according to Eq. (10.53), then applies the ML reconstruction scheme to each set and obtains a series of density matrices  $\hat{\rho}'_i$  each of which approximates the original matrix  $\hat{\rho}_{ML}$ . The average difference  $\langle |\hat{\rho}_{ML} - \hat{\rho}'_i| \rangle_i$  evaluates the statistical uncertainty associated with the reconstructed density matrix.

## Conclusion

The goal of this chapter was two-fold. First of all, considering several different experimental setups, the general principles of ML estimation discussed in chapter by Hradil *et al.* were shown “at work.” The second goal was to compare the outcomes of ML estimation to that of other, more standardly used, reconstruction techniques. In all discussed cases, the ML estimation was seen to provide better results. We hope we persuaded the reader that the ML estimation is a versatile and highly efficient tool that can be applied to almost any reconstruction problem. We also believe that especially in quantum theory, where the experiments are often very demanding, the high efficiency of ML estimation is worth of somewhat larger computing costs compared to standard techniques. The aim of our research could be paraphrased as our credo: *If the experiment deserves to be done, then the data certainly deserve to be treated properly.* Anybody familiar with the rules of logic would agree that this statement could also be rephrased as the negative affirmation: *If the data are not treated properly, the experiment did not deserve to be done.*

## Acknowledgments

The authors would like to acknowledge support by the following grants and projects. J. Řeháček and Z. Hradil: Grant No. LN00A015 of the Czech Ministry of Education and East-West program of the Austrian Academy of Sciences; G. Badurek, H. Rauch, and M. Zawisky: Project No P14229-PHY of the Austrian Science Foundation and Böhler Bleche GmbH, Mürzzuschlag, Austria; G. Molina-Terriza: Marie Curie Fellowship, project number SFB 015 P06 of Austrian Science Foundation (FWF), and contract No. IST-2001-38864, RAMBOQ of the European Commission; A. I. Lvovsky: Emmy Noether research grant of the Deutsche Forschungsgemeinschaft.

## References

1. M. M. Nieto: Phys. Scr. T **48**, 5 (1993).
2. A. Royer: Phys. Rev. A **53**, 70 (1996).
3. R. Lynch: Phys. Reports **256**, (1995).
4. V. Peřinová and A. Lukš: *Phase in Optics* (World Scientific, Singapore, 1998).
5. A. Luis and L. L. Sánchez-Soto: In *Progress in Optics*, vol 41, ed by E. Wolf (Elsevier, Amsterdam, 2000).
6. C. W. Helstrom: *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
7. B. Yurke, S. L. McCall, and J. R. Klauder: Phys. Rev. A **33**, 4033 (1986).
8. A. S. Lane, S. L. Braunstein, and C. M. Caves: Phys. Rev. A **47**, 1667 (1993).
9. M. J. Holland and K. Burnett: Phys. Rev. Lett. **71**, 1355 (1993).
10. B. C. Sanders and G. J. Milburn: Phys. Rev. Lett. **75**, 2944 (1995).
11. A. V. Chizhov, V. De-Renzi, and M. G. A. Paris: Phys. Lett. A **237**, 201 (1998).
12. L. Susskind and J. Glogower: Physics **1**, 49 (1964).
13. S. M. Barnett and D. T. Pegg: Phys. Rev. Lett. **76**, 4148 (1996).
14. Z. Hradil: Phys. Rev. A **55**, R1561 (1997).
15. M. G. Kendall and A. Stuart: *Advanced Theory of Statistics*, vol 2 (Charles Griffin, London, 1961).
16. K. R. W. Jones: Annals of Phys. **207**, 140 (1991).
17. Z. Hradil et al: Phys. Rev. Lett. **76**, 4295 (1996).
18. M. Zawisky et al: J. Phys. A: Math. Gen. **31**, 551 (1998).
19. J. Řeháček, Z. Hradil, M. Zawisky, S. Pascazio, H. Rauch, and J. Peřina: Phys. Rev. A **60**, 473 (1999).
20. J. Řeháček, Z. Hradil, M. Dušek, O. Haderka, and M. Hendrych: J. Opt. B: Quantum Semiclass. Opt. **2**, 237 (2000).
21. J. W. Noh, A. Fougères, and L. Mandel: Phys. Rev. Lett. **67**, 1426 (1991).
22. J. W. Noh, A. Fougères, and L. Mandel: Phys. Rev. Lett. **71**, 2579 (1993).
23. J. F. Walkup and J. W. Goodman: J. Opt. Soc. Am. **63**, 399 (1973).
24. A.C. Kak, and M.Slaney: *Principles of Computerized Tomographic Imaging*, ed by R.F. Cotellessa (IEEE Press, 1987).
25. B. Schillinger, W. Blümlhuber, A. Fent, and M. Wegner: Nucl. Instr. and Methods, **A424**, 58 (1998).
26. B. Schillinger, E. Lehmann, and P. Vontobel: Physica B **276**, 59 (2000).
27. S. Koerner, B. Schillinger, P. Vontobel, and H. Rauch: Nucl. Instrum. Meth. A **471**, 69 (2001).
28. P.J. McMahon, B.E. Allman, K.A. Nugent, D.L. Jacobson, M. Arif, and S.A. Werner: Appl. Phys. Lett. **78**, 1011 (2001).
29. F. Dubus, U. Bonse, T. Biermann, M. Baron, F. Beckmann, and M. Zawisky: SPIE Proc. **4503**, 259 (2002).
30. M. Zawisky, M. Bastürk, J. Rehacek, Z. Hradil: Neutron Tomographic Investigations of Strong Absorbing Materials. In: *7th World Conference on Neutron Radiography* (European Commission Special Book, Roma, 2002), in print.
31. W. Treimer, P. Maaß, H. Strothmann, and U. Feye-Treimer: Physica B **17**, 532 (1991).
32. P. Maaß, W. Treimer, and U. Feye-Treimer: Impact of Computing in Science and Engineering **4**, 250 (1992).
33. W. Treimer and U. Feye-Treimer: Physica B **241**, 1228 (1998).

34. B.E. Allman, P.J. McMahon, K.A. Nugent, D. Paganin, D.L. Jacobson, M. Arif, and S.A. Werner: *Nature* **408**, 158 (2000).
35. U. Bonse, R. Nusshardt, F. Busch, R. Pahl, J.H. Kinney, Q.C. Johnson, R.A. Saroyan, M.C. Nichols: *Journal of Materials Science* **26**, 4076 (1991).
36. V.F. Sears: *Neutron Optics*, (Oxford University Press, Oxford, 1989).
37. S. Quabis, R. Dorn, M. Eberler, O. Glöckl, and G. Leuchs: *Appl. Phys. B* **72**, 109 (2001)
38. J. Řeháček, Z. Hradil, O. Haderka, J. Perina, and M. Hamar: *Phys. Rev. A*, **67**, 061801 (2003).
39. U. Janicke and M. Wilkens: *J. Mod. Optics* **42**, 2183 (1995).
40. A.P. Dempster, N.M. Laird, and D.B. Rubin: *J. R. Statist. Soc. B* **39**, 1 (1977).
41. Y. Vardi and D. Lee: *J. R. Statist. Soc. B* **55**, 569 (1993).
42. L.A. Shepp and Y. Vardi: *IEEE Trans. Med. Imag.* **1**, 113 (1982).
43. Y. Vardi, L.A. Shepp, and L. Kaufman: *Statist. Assoc.* **80**, 8 (1985).
44. B.A. Mair, M. Rao, and J.M.M. Anderson: *Inverse Problems* **12**, 965 (1996).
45. H. Rauch, J. Summhammer, M. Zawisky, and E. Jericha: *Phys. Rev. A* **42**, 3726 (1990).
46. J. Rehacek, Z. Hradil, M. Zawisky, W. Treimer, and M. Strobl: *Europhysics Letters* **59** 694 (2002).
47. V.F. Sears: *Neutron News*, **3**, 26 (1992).
48. S. Koerner, B. Schillinger, P. Vontobel, and H. Rauch: *Nucl. Instrum. Meth. A* **471**, 69 (2001).
49. F. Dubus, U. Bonse, M. Zawisky, M. Baron, and R. Loidl: *First Phase-Contrast Tomography with thermal neutrons*, 7<sup>th</sup> World Conference on Neutron Radiography, Roma (2002), European Commission Special Book, in print; *IEEE Transactions of Nuclear Science*, Special Issue, in print.
50. A. Momose: *Nucl. Instr. and Methods A* **352**, 622 (1995).
51. A. Momose, T. Takeda, Y. Itai, and K. Hirano: *Nature Med* **2**, 473 (1996).
52. F. Beckmann, U. Bonse, F. Busch, and O. Gnnewig: *J. Comput. Assist. Tomogr.* **21**, 539 (1997).
53. G. Badurek, R.J. Buchelt, and H. Leeb: *Physica B* **276**, 588 (2000).
54. H. Rauch and S.A. Werner: *Neutron Interferometry* (Oxford University Press, Oxford, 2000).
55. M. Zawisky, H. Rauch, and Y. Hasegawa: *Phys. Rev. A* **50** 5000 (1994).
56. Z. Hradil, J. Summhammer, G. Badurek, and H. Rauch: *Phys. Rev. A* **62** 014101 (2000).
57. M. Bourennane, A. Karlsson, and G. Björk: *Phys. Rev. A* **64**, 012306 (2001).
58. H. Bechmann-Pasquinucci and A. Peres: *Phys. Rev. Lett.* **85**, 3313 (2000).
59. H. Bechmann-Pasquinucci and W. Tittel: *Phys. Rev. A* **61**, 62308 (2000).
60. D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal: arXiv:quant-ph/9908070 (2000).
61. S. D. Bartlett, H. de Guise, and B. C. Sanders: arXiv:quant-ph/0011080 (2000).
62. V. Karimipour, S. Bagherinezhad, and A. Bahraminasab: arXiv:quant-ph/0112050 (2001).
63. Andris Ambainis: Proceedings of STOC, 2001, pp 134–142.
64. C. Brukner, M. Zukowski and A. Zeilinger: *Phys. Rev. Lett.* **89**, 197901 (2002).
65. R. Jozsa, and J. Schlienz: *Phys. Rev. A* **62**, 012301 (2000).
66. H. de Riedmatten, I. Marcikic, H. Zbinden and N. Gisin: *Quantum Information and Computation* **2**, 425 (2002).

67. M. Zukowski, A. Zeilinger, M. A. Horne: Phys. Rev. A **55**, 2564 (1997).
68. A. Mair, A. Vaziri, G. Weihs, and A. Zeilinger: Nature **412**, 313 (2001).
69. A. Vaziri, J. W. Pan, T. Jennewein, G. Weihs and A. Zeilinger: arXiv:quant-ph/0303003 (2003).
70. A. Vaziri, G. Weihs and A. Zeilinger: Phys. Rev. Lett. **89**, 240401 (2002).
71. G. Molina-Terriza, J. P. Torres and L. Torner: Phys. Rev. Lett. **88**, 013601 (2002).
72. A. Vaziri, G. Weihs, and A. Zeilinger: J. Opt. B: Quantum Semiclass. **4**, S47 (2002).
73. J. P. Torres, Y. Deyanova, L. Torner and G. Molina-Terriza: Phys. Rev. A **67**, 052313 (2003).
74. J. P. Torres, A. Alexandrescu and L. Torner: arXiv:quant-ph/0306105 (2003).
75. L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman: Phys. Rev. A **45**, 8185 (1992).
76. H. He, M.E. Friese, N.R. Heckenberg, and H. Rubinsztein-Dunlop: Phys. Rev. Lett. **75**, 826 (1995).
77. M. E. J. Friese, J. Enger, H. Rubinsztein-Dunlop, and N.R. Heckenberg: Phys. Rev. A **54**, 1593 (1996).
78. V. Yu Bazhenov, M. V. Vasnetsov, and M. S. Soskin: JETP Lett. **52** 429 (1990).
79. N. R. Heckenberg, R. McDu., C. P. Smith, and A. G. White: Opt. Lett. **17**, 221 (1992).
80. J. Řeháček, Z. Hradil, and M. Ježek: Phys. Rev. A **63**, 040303(R) (2001).
81. Ulf Leonhardt: *Measuring the Quantum State of Light* (Cambridge University Press, Cambridge, 1997).
82. K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989)
83. D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993);
84. G. M. D'Ariano, U. Leonhardt, H. Paul: Phys. Rev. A **52**, 1801(R) (1995).
85. U. Leonhardt, M. G. Raymer: Phys. Rev. Lett. **76**, 1985 (1996).
86. A. I. Lvovsky and J. Mlynek: Phys. Rev. Lett **88** 250401 (2002).
87. W. Vogel: Phys. Rev. Lett. **84**, 1849 (2000).
88. G. Breitenbach, S. Schiller, and J. Mlynek: Nature **387**, 471 (1997); H. Hansen et al: Opt. Lett. **26**, 1714 (2001).
89. K. Banaszek: Phys. Rev. A **57**, 5013 (1998).
90. K. Banaszek: Acta Phys. Slov. **48**, 185 (1998).
91. K. Banaszek et al: Phys. Rev. A **61**, 010304 (2000).
92. T. Kiss, U. Herzog, U. Leonhardt: Phys. Rev. A **52**, 2433 (1995).
93. K. Banaszek: Phys. Rev. A **59**, 4797 (1999).
94. A. Royer: Phys. Rev. A **15**, 449 (1977); B.-G. Englert: J. Phys. A **22**, 625 (1989); H. Moya-Cessa and P. L. Knight: Phys. Rev. A **48**, 2479 (1993).

## Part II

### **Quantum Decision**



# 11 Discrimination of Quantum States

János A. Bergou<sup>1</sup>, Ulrike Herzog<sup>2</sup>, and Mark Hillery<sup>3</sup>

<sup>1</sup> Department of Physics and Astronomy, Hunter College of The City University of New York, 695 Park Avenue, New York, NY 10021  
`jbergou@hunter.cuny.edu`

<sup>2</sup> Institut für Physik, Humboldt-Universität Berlin, Newtonstrasse 15, D-12489 Berlin, Germany `ulrike.herzog@physik.hu-berlin.de`

<sup>3</sup> Department of Physics and Astronomy, Hunter College of The City University of New York, 695 Park Avenue, New York, NY 10021  
`mhillery@hunter.cuny.edu`

## 11.1 Introduction

In quantum information and quantum computing the carrier of information is some quantum system and information is encoded in its state [1]. The state, however, is not an observable in quantum mechanics [2] and, thus, a fundamental problem arises: after processing the information - i.e. after the desired transformation is performed on the input state by the quantum processor - the information has to be read out or, in other words, the state of the system has to be determined. When the possible target states are orthogonal, this is a relatively simple task if the set of possible states is known. But when the possible target states are not orthogonal they cannot be discriminated perfectly, and optimum discrimination with respect to some appropriately chosen criteria is far from being trivial even if the set of the possible nonorthogonal states is known. Thus the problem of discriminating among nonorthogonal states is ubiquitous in quantum information and quantum computing, underlying many of the communication and computing schemes that have been suggested so far. It is the purpose of this article to review various theoretical schemes that have been developed for discriminating among nonorthogonal quantum states. The corresponding experimental realizations will be mentioned only in a cursory manner as they are reviewed elsewhere in this volume by Chefles.

The field of discriminating among nonorthogonal quantum states has been around for quite some time now [3]. Stimulated by the rapid developments in quantum information theory of the 90's the question of how to discriminate between nonorthogonal quantum states *in an optimum way* has gained renewed interest. The developments until about the late 90's are reviewed in an excellent review article by Chefles [4]. Therefore, in this review we will mainly focus our attention to recent advances not contained in [4] and will cite earlier results only when they are necessary for the understanding of the newer ones.

In order to devise an optimum state-discriminating measurement, strategies have been developed with respect to various criteria. In this review article we restrict ourselves to the two most obvious criteria for optimizing a mea-

surement scheme that is designed for discriminating between different states of a quantum system, being either pure states or mixed states. The two methods are optimum unambiguous discrimination of the states, on the one hand, and state discrimination with minimum error, on the other hand. They will be outlined in detail in the next two sections of this review. In particular, at the beginning of these sections we give a tutorial introduction to the two main strategies by considering simple examples, namely unambiguous discrimination of two pure states in Sec. 11.2, and minimum-error discrimination of two mixed states in Sec. 11.3. This will allow us to introduce the concept of generalized measurements along with the other typical theoretical tools employed in problems of this sort. Selected applications of the unambiguous discrimination strategy, one chosen from quantum communication and the other from quantum computing will be reviewed at the end of Sec. 11.2. Section 11.4 is devoted to the recently emerging sub-field of discriminating among multipartite states by means of local operations and classical communication (LOCC). We conclude with a brief outlook in Sec. 11.5. Throughout the article we assume that for each measurement only a single copy of the quantum system is available. However, the case of multiple copies could be easily accounted for, in the measurement schemes we consider, if the states are replaced by their corresponding multi-fold tensor products. We note that apart from the two optimization schemes we consider, state-distinguishing measurements can also be optimized with respect to other criteria, such as requiring a maximum of the mutual information [5] or of the fidelity [6]. In particular, the former approach is of importance for the transmission of quantum information. From a mathematical point of view, however, these other criteria pose much bigger problems, since they rely on optimizing a nonlinear functional of the given states and, therefore, are beyond the scope of the current review.

## 11.2 Unambiguous discrimination

In this Section we will review schemes for unambiguous discrimination. Although the simplest case of two pure states is well known and has been reviewed extensively (see, for example, Ref. [4]), from a pedagogical point of view we find it useful to include it here as well, because many of the techniques employed later can be best understood on this simple example. We also want to mention that the two main discrimination strategies evolved rather differently from the very beginning. On the one hand, unambiguous discrimination started with pure states and only very recently turned its attention to discriminating among mixed quantum states. At the end of this section we will review recent progress in this area. On the other hand, minimum-error discrimination addressed the problem of discriminating among two mixed quantum states from the very beginning and the results for two pure states followed as

special cases. Each strategy has its own advantages and drawbacks. While unambiguous discrimination is relatively straightforward to generalize for more than two states it is difficult to treat mixed states. The error-minimizing approach, initially developed for two mixed states, is hard to generalize for more than two states.

Before we begin our systematic study of the optimal unambiguous discrimination of two pure states in the next subsection, we want to gain some physical insight first. To this end, let us describe the procedure of optimum unambiguous discrimination between two nonorthogonal single photon polarization states in terms of classical optics. We consider a series of weak pulses containing, on the average, much less than one photon. Each pulse is linearly polarized, with equal probability, either in the direction  $\mathbf{e}_1$  or in the direction  $\mathbf{e}_2$  with  $\mathbf{e}_{1,2} = \cos\Theta \mathbf{e}_x \pm \sin\Theta \mathbf{e}_y$ , where we assume that  $\cos\Theta \geq \sin\Theta$ . If the pulses pass through a linear optical device and undergo polarization-selective linear attenuation in the  $x$  direction, their polarization vectors can be made orthogonal. For this purpose the attenuation process has to be designed in such a way that the amplitude of the  $x$ -component is reduced by a factor of  $\tan\Theta$ . Due to the attenuation, the initial electric field vectors  $\mathbf{E}_{1,2} = E_0 \mathbf{e}_{1,2}$  of the respective pulses are then transformed into the vectors  $\mathbf{E}'_{1,2} = E_0 \sqrt{2} \sin\Theta (\mathbf{e}_x \pm \mathbf{e}_y) / \sqrt{2}$ . Hence, after leaving the linear optical device, the polarization directions of the two kinds of pulses are orthogonal and, therefore, can be discriminated unambiguously even when the pulses contain only a single photon. However, the intensity that can be used for unambiguous polarization state discrimination is reduced by a factor of  $2 \sin^2\Theta$  relative to the total initial intensity. Since a classical intensity ratio corresponds to the probability ratio of detecting a photon, every incoming photon will yield an unambiguous result in the state discrimination process only with the probability

$$P_D = 2 \sin^2\Theta = 1 - \cos(2\Theta) = 1 - |\mathbf{e}_1 \cdot \mathbf{e}_2|. \quad (11.1)$$

Here we introduced the modulus of the scalar product since, when the direction of one of the vectors,  $\mathbf{e}_1$  or  $\mathbf{e}_2$ , is reversed, the linear polarization state remains the same. The probability that the discrimination procedure fails, i. e. that the polarization state of the photon cannot be determined unambiguously, is therefore  $Q_F = 1 - P_D = |\mathbf{e}_1 \cdot \mathbf{e}_2|$ . Although orthogonalization of the polarization states is also possible when polarization-selective attenuation affects a polarization direction that differs from the  $x$ -axis, it is easy to see that the symmetric procedure described above is optimal in the sense that it yields the maximum achievable value of  $P_D$ , or the minimum value of  $Q_F = 1 - P_D$ , respectively.

From the pioneering investigations of unambiguous discrimination between general nonorthogonal quantum states it follows that for any two states,  $|\psi_1\rangle$  and  $|\psi_2\rangle$ , occurring with equal a priori probability, the optimum probability of obtaining an unambiguous result is given by Eq. (11.1) when the

scalar product  $e_1 e_2$  is replaced by the overlap  $\langle \psi_1 | \psi_2 \rangle$ , as we will show in the following (see Eq. (11.2), in particular).

### 11.2.1 Unambiguous discrimination of two pure states

Unambiguous discrimination started with the work of Ivanovic [7] who studied the following problem. A collection of quantum systems is prepared so that each single system is equally likely to be prepared in one of two known states,  $|\psi_1\rangle$  or  $|\psi_2\rangle$ . Furthermore, the states are not orthogonal,  $\langle \psi_1 | \psi_2 \rangle \neq 0$ . The preparer then hands the systems over to an observer one by one whose task is to determine which one of the two states has actually been prepared in each case. All the observer can do is to perform a single measurement or perhaps a series of measurements on the individual system. Ivanovic came to the conclusion that if one allows inconclusive detection results to occur then in the remaining cases the observer can conclusively determine the state of the individual system.

It is rather easy to see that a simple von Neumann measurement can accomplish this task. Let us denote the Hilbert space of the two given states by  $\mathcal{H}$  and introduce the projector  $P_1$  for  $|\psi_1\rangle$  and  $\bar{P}_1$  for the orthogonal subspace, such that  $P_1 + \bar{P}_1 = 1$ , the unity in  $\mathcal{H}$ . Then we know for sure that  $|\psi_2\rangle$  was prepared if in the measurement of  $\{P_1, \bar{P}_1\}$  a click in the  $\bar{P}_1$  detector occurs. A similar conclusion for  $|\psi_1\rangle$  can be reached with the roles of  $|\psi_1\rangle$  and  $|\psi_2\rangle$  reversed. Of course, when a click along  $P_1$  (or  $P_2$ ) occurs then we learn nothing about which state was prepared thus corresponding to inconclusive results. Ivanovic's startling observation was that a sequence of measurements can sometimes do better than a single von Neumann measurement described here. Dieks [8] then found that this sequence of measurements can be realized with a single generalized measurement (POVM) and Peres subsequently showed that this POVM is optimal in the sense that its failure probability, the probability that an inconclusive outcome occurs, is minimum [9]. The probability of the inconclusive outcome, or failure, is  $Q_{IDP} = |\langle \psi_1 | \psi_2 \rangle|$  and the probability of success is then given by what is called the Ivanovic-Dieks-Peres (IDP) limit,

$$P_{IDP} = 1 - Q_{IDP} = 1 - |\langle \psi_1 | \psi_2 \rangle|. \quad (11.2)$$

This result can be generalized for the case when the preparation probabilities of the states,  $\eta_1$  and  $\eta_2$ , are different,  $\eta_1 \neq \eta_2$ . The preparation probabilities are also called a priori probabilities. The IDP result thus corresponds to the case of equal a priori probabilities for the two states,  $\eta_1 = \eta_2 = 1/2$ , and the generalization for arbitrary a priori probabilities is due to Jaeger and Shimony [10]. Here we briefly review the derivation of the general result following [10] and the very readable account by Ban [11].

### 11.2.2 Optimal POVM and the complete solution

The von Neumann projective measurement described above has two outcomes. It can correctly identify one of the two states at the expense of missing the other completely and occasionally missing the identifiable one, as well. If we want to do better we would like to have a measurement with three - not just two - outcomes,  $|\psi_1\rangle$ ,  $|\psi_2\rangle$  and failure. In the two-dimensional Hilbert space  $\mathcal{H}$  the number of possible outcomes for a von Neumann measurement cannot exceed two, since it is always restricted by the dimensionality of the Hilbert space. We have to turn to generalized measurements that allow greater flexibility than simple projective measurements [12]. In particular, the number of distinguishable outcomes can exceed the dimensionality of the corresponding Hilbert space. For our case this means that we replace the projector  $\bar{P}_2$  by the quantum detection operator  $\Pi_1$ ,  $\bar{P}_1$  by  $\Pi_2$  and introduce  $\Pi_0$  for the inconclusive results in such a way that  $\langle\psi_1|\Pi_1|\psi_1\rangle = p_1$  is the probability of successfully identifying  $|\psi_1\rangle$ ,  $\langle\psi_1|\Pi_0|\psi_1\rangle = q_1$  is the probability of failing to identify  $|\psi_1\rangle$ , (and similarly for  $|\psi_2\rangle$ ). For unambiguous discrimination we then require  $\langle\psi_2|\Pi_1|\psi_2\rangle = \langle\psi_1|\Pi_2|\psi_1\rangle = 0$ . We want these possibilities to be exhaustive,

$$\Pi_1 + \Pi_2 + \Pi_0 = I , \quad (11.3)$$

where  $I$  is the unity operator in  $\mathcal{H}$ . The probabilities are always real and non-negative which implies that the quantum detection operators are Hermitean and non-negative or, in other words, positive semi-definite.

Clearly, Eq. (11.3) does not correspond to orthogonal measurements when all detection operators are different from zero. It describes an operation called *positive operator-valued measure* (POVM) or simply a generalized measurement with the detection operators as its elements.

We now want to determine these operators explicitly. Consider the operator  $A_k = U_k \Pi_k^{1/2}$ , where  $U_k$  is an arbitrary unitary operator ( $k = 0, 1, 2$ ). From this expression we immediately obtain  $\Pi_k = A_k^\dagger A_k$  and the detection probability can be expressed as  $\langle\psi_i|A_k^\dagger A_k|\psi_i\rangle = \|A_k \psi_i\|^2 \geq 0$  where  $\|...\|$  stands for the norm. This expression also helps us to identify the so far arbitrary operator  $A_k$ . The expression  $A_k |\psi_i\rangle$  corresponds to the post-detection state. Because of the positivity of the norm, the condition of unambiguous discrimination is equivalent to the requirement

$$A_1 |\psi_2\rangle = A_2 |\psi_1\rangle = 0 . \quad (11.4)$$

If we introduce  $|\psi_i^\perp\rangle$  as the vector orthogonal to  $|\psi_{i'}\rangle$  ( $i \neq i'$ ) - a notation that will become obvious in Sec. 11.2.2 - then  $A_1 = c_1 |\bar{\psi}_1\rangle \langle \psi_1^\perp|$  and  $A_2 = c_2 |\bar{\psi}_2\rangle \langle \psi_2^\perp|$ . Here  $c_i$  are complex coefficients to be determined from the condition of optimum and  $|\bar{\psi}_1\rangle$  and  $|\bar{\psi}_2\rangle$  are the post-detection states, normalized to unity. For perfect distinguishability of the post-detection states,

corresponding to optimal discrimination, we have to require their orthogonality,  $\langle \psi_1 | \bar{\psi}_2 \rangle = 0$ , so they can be represented by a pair of arbitrarily directed orthogonal vectors in  $\mathcal{H}$ .

With the help of these expressions we can write the detection operators as  $\Pi_1 = A_1^\dagger A_1 = |c_1|^2 |\psi_1^\perp\rangle\langle\psi_1^\perp|$  and  $\Pi_2 = A_2^\dagger A_2 = |c_2|^2 |\psi_2^\perp\rangle\langle\psi_2^\perp|$ . Inserting these expressions in the definition of  $p_1$  and  $p_2$  gives  $|c_1|^2 = p_1 / |\langle \psi_1 | \psi_1^\perp \rangle|^2$  and a similar expression for  $|c_2|^2$ . Finally, introducing  $\cos \Theta = |\langle \psi_1 | \psi_2 \rangle|$  and  $\sin \Theta = |\langle \psi_1 | \psi_1^\perp \rangle|$ , we can write the detection operators as

$$\begin{aligned}\Pi_1 &= \frac{p_1}{\sin^2 \Theta} |\psi_1^\perp\rangle\langle\psi_1^\perp|, \\ \Pi_2 &= \frac{p_2}{\sin^2 \Theta} |\psi_2^\perp\rangle\langle\psi_2^\perp|.\end{aligned}\quad (11.5)$$

Now,  $\Pi_1$  and  $\Pi_2$  are positive semi-definite operators by construction. However, there is one additional condition for the existence of the POVM which is the positivity of the inconclusive detection operator,

$$\Pi_0 = I - \Pi_1 - \Pi_2. \quad (11.6)$$

This is a simple 2 by 2 matrix in  $\mathcal{H}$  and the corresponding eigenvalue problem can be solved analytically. Non-negativity of the eigenvalues leads, after some tedious but straightforward algebra, to the condition

$$q_1 q_2 \geq |\langle \psi_1 | \psi_2 \rangle|^2, \quad (11.7)$$

where  $q_1 = 1 - p_1$  and  $q_2 = 1 - p_2$  are the failure probabilities for the corresponding input states.

Eq. (11.7) represents the constraint imposed by the positivity requirement on the optimum detection operators. The task we set out to solve can now be formulated as follows. Let

$$Q = \eta_1 q_1 + \eta_2 q_2 \quad (11.8)$$

denote the average failure probability for unambiguous discrimination. We want to minimize this failure probability subject to the constraint, Eq. (11.7). Due to the relation,  $P = \eta_1 p_1 + \eta_2 p_2 = 1 - Q$ , the minimum of  $Q$  also gives us the maximum probability of success. Clearly, for optimum the product  $q_1 q_2$  should be at its minimum allowed by (11.7), and we can then express  $q_2$  with the help of  $q_1$  as  $q_2 = \cos^2 \Theta / q_1$ . Inserting this expression in (11.8) yields

$$Q = \eta_1 q_1 + \eta_2 \frac{\cos^2 \Theta}{q_1}, \quad (11.9)$$

where  $q_1$  can now be regarded as the independent parameter of the problem. Optimization of  $Q$  with respect to  $q_1$  gives  $q_1^{POVM} = \sqrt{\eta_2 / \eta_1} \cos \Theta$  and  $q_2^{POVM} = \sqrt{\eta_1 / \eta_2} \cos \Theta$ . Finally, substituting these optimal values into Eq. (11.8) gives the optimum failure probability,

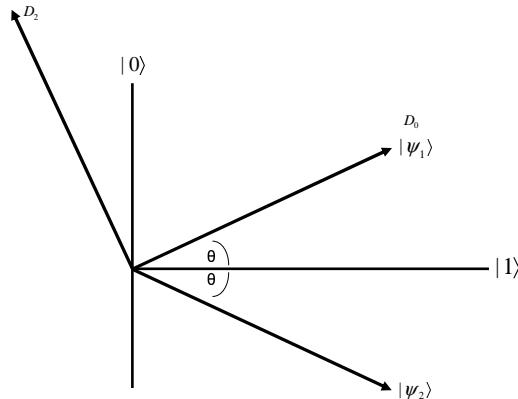
$$Q^{POVM} = 2\sqrt{\eta_1 \eta_2} \cos \Theta . \quad (11.10)$$

For  $\eta_1 = \eta_2 = 1/2$  this reproduces the IDP result, Eq. (11.2), as it should.

Let us next see how this result compares to the average failure probabilities of the two possible unambiguously discriminating von Neumann measurements that were described at the beginning of this section. The average failure probability for the first von Neumann measurement, with its failure direction along  $|\psi_1\rangle$ , can be written by simple inspection as

$$Q_1 = \eta_1 + \eta_2 |\langle \psi_1 | \psi_2 \rangle|^2 , \quad (11.11)$$

since  $|\psi_1\rangle$  gives a click with probability 1 in this direction but it is only prepared with probability  $\eta_1$  and  $|\psi_2\rangle$  gives a click with probability  $|\langle \psi_1 | \psi_2 \rangle|^2$  but it is only prepared with probability  $\eta_2$ . The corresponding detector set-up, yielding  $Q_1$  for the failure probability, is depicted in Fig. 11.1.

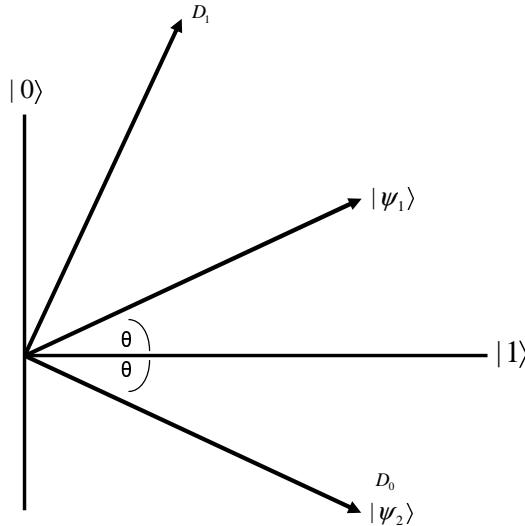


**Fig. 11.1.** A von Neumann measurement that discriminates  $|\psi_2\rangle$  unambiguously. The detector  $D_0 = P_1$  is set up along  $|\psi_1\rangle$  and the detector  $D_2 = \bar{P}_1$  is set up along the orthogonal direction. When a click in the  $D_2$  detector occurs we know for certain that  $|\psi_2\rangle$  was prepared as the input state since it is the only one that has a component along this direction. When a click in the detector  $D_0$  occurs we learn nothing about which state was prepared as the input since both have a component along this direction. This measurement outcome then corresponds to the inconclusive result so  $D_0$  is the failure detector and the probability that it clicks is  $Q_1$ .

By entirely similar reasoning, the average failure probability for the second von Neumann measurement, with its failure direction along  $|\psi_2\rangle$ , is given by

$$Q_2 = \eta_1 |\langle \psi_1 | \psi_2 \rangle|^2 + \eta_2 . \quad (11.12)$$

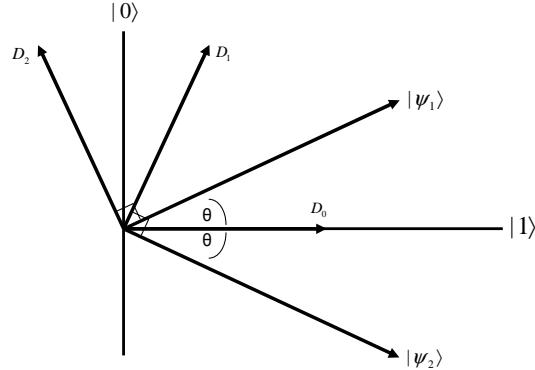
The corresponding detector set-up, yielding  $Q_2$  for the failure probability, is depicted in Fig. 11.2.



**Fig. 11.2.** A von Neumann measurement that discriminates  $|\psi_1\rangle$  unambiguously. The detector  $D_0 = P_2$  is set up along  $|\psi_2\rangle$  and the detector  $D_1 = \bar{P}_2$  is set up along the orthogonal direction. When a click in the  $D_1$  detector occurs we know for certain that  $|\psi_1\rangle$  was prepared as the input state since it is the only one that has a component along this direction. When a click in the detector  $D_0$  occurs we learn nothing about which state was prepared as the input since both have a component along this direction. This measurement outcome then corresponds to the inconclusive result so  $D_0$  is the failure detector and the probability that it clicks is  $Q_2$ .

What we can observe is that  $Q_1$  and  $Q_2$  are given as the arithmetic mean of two terms and  $Q^{POVM}$  is the geometric mean of the same two terms for either case. So, one would be tempted to say that the POVM performs better always. This, however, is not quite the case, it does so only when it exists. The obvious condition for the POVM solution to exist is that both  $q_1^{POVM} \leq 1$  and  $q_2^{POVM} \leq 1$ . Using  $\eta_2 = 1 - \eta_1$ , a little algebra tells us that the POVM exists in the range  $\cos^2 \Theta / (1 + \cos^2 \Theta) \leq \eta_1 \leq 1 / (1 + \cos^2 \Theta)$ . If  $\eta_1$  is smaller than the lower boundary, the POVM goes over to the first von Neumann measurement and if  $\eta_1$  exceeds the upper boundary the POVM goes over to the second von Neumann measurement. This can be easily seen from Eqs. (11.5) and (11.6) since  $p_1 = 1 - q_1 = 0$  for  $q_1 = 1$  and  $\Pi_0$  becomes

a projection along  $|\psi_1\rangle$  (and correspondingly for  $p_2 = 0$ ). The set-up of the detection operators, yielding  $Q^{POVM}$  for the failure probability, is depicted in Fig. 11.3.



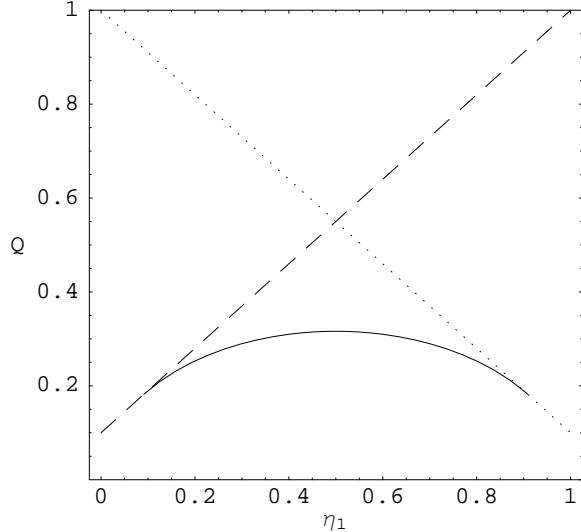
**Fig. 11.3.** Optimal POVM that discriminates  $|\psi_1\rangle$  and  $|\psi_2\rangle$  unambiguously. The detector  $D_0 = \Pi_0$  is set up symmetrically between  $|\psi_1\rangle$  and  $|\psi_2\rangle$ , the detector  $D_1 = \Pi_1$  is set up along  $\bar{P}_2$  and the detector  $D_2 = \Pi_2$  is set up along  $\bar{P}_1$ . When a click in the  $D_i$  detector occurs we know for certain that  $|\psi_i\rangle$  was prepared ( $i = 1, 2$ ) as the input state since it is the only one that has a component along this direction. When a click in the detector  $D_0$  occurs we learn nothing about which state was prepared as the input since both have a component along this direction. This measurement outcome then corresponds to the inconclusive result so  $D_0$  is the failure detector and the probability that it clicks is  $Q^{POVM}$ . (The figure is for illustrative purposes only. Arrows representing POVM detection operators *do not correspond* to simple projections along their respective directions. They are drawn shorter than arrows representing state vectors which, in turn, have unit length. For an implementation of the POVM, see Sec. 11.2.3.)

These findings can be summarized as follows. The optimal failure probability,  $Q^{opt}$ , is given as

$$Q^{opt} = \begin{cases} Q^{POVM} & \text{if } \frac{\cos^2 \theta}{1+\cos^2 \theta} \leq \eta_1 \leq \frac{1}{1+\cos^2 \theta}, \\ Q_1 & \text{if } \eta_1 < \frac{\cos^2 \theta}{1+\cos^2 \theta}, \\ Q_2 & \text{if } \frac{1}{1+\cos^2 \theta} < \eta_1. \end{cases} \quad (11.13)$$

Fig. 11.4 displays the failure probabilities vs.  $\eta_1$  for a fixed value of the overlap,  $\cos^2 \theta$ .

The above result is very satisfying from a physical point of view. The POVM delivers a lower failure probability in its entire range of existence than either of the two von Neumann measurements. At the boundaries of this range it merges smoothly with the von Neumann measurement that has a lower failure probability at that point. Outside this range the state



**Fig. 11.4.** Failure probability,  $Q$ , vs. the prior probability,  $\eta_1$ . Dashed line:  $Q_1$ , dotted line:  $Q_2$ , solid line:  $Q_{POVM}$ . For the figure we used the following representative value:  $|\langle\psi_1|\psi_2\rangle|^2 = 0.1$ . For this the optimal failure probability,  $Q_{opt}$ , is given by  $Q_1$  for  $0 < \eta_1 < 0.09$ , by  $Q_{POVM}$  for  $0.09 \leq \eta_1 \leq 0.9$  and by  $Q_2$  for  $0.9 < \eta_1$ .

preparation is dominated by one of the states and the optimal measurement becomes a von Neumann projective measurement, using the state that is prepared less frequently as its failure direction.

Next we turn our attention to the physical implementation of the POVM.

### 11.2.3 Neumark's theorem and the realization of the POVM

The problem with the POVM is that it is very hard to realize in the original system Hilbert space since it involves detection operators that do not correspond to orthogonal projectors. In fact, they are not projectors at all. Fortunately, there is an essential result, known as Neumark's theorem [13]. It states that a POVM can be realized by the following constructive procedure, also known as generalized measurement. The system is embedded in a larger Hilbert space where the extra degrees of freedom are customarily called the ancilla. Then a unitary transformation entangles the system degrees of freedom with those of the ancilla. Finally, after this interaction, projective von Neumann measurements can be carried out on this larger system. As a consequence of the entanglement between the original system and the ancilla, a projective measurement on the larger system will also transform the system state in the original Hilbert space. One can choose the unitary transformation and the subsequent von Neumann measurement in such a way that if outcome  $k$  is found in the von Neumann measurement on the larger system,

the resulting transformation on the original system state is  $A_k|\psi\rangle$ , i. e. it corresponds to an element of the POVM in the original system Hilbert space.

We illustrate the power of this theorem on an alternative derivation of the condition on the individual failure probabilities, Eq. (11.7). The joint Hilbert space  $\mathcal{K}$  of the ‘system plus ancilla’ is a tensor product of the two Hilbert spaces,  $\mathcal{H}$  of the system and  $\mathcal{A}$  of the ancilla,  $\mathcal{K} = \mathcal{H} \otimes \mathcal{A}$ . This means that a state in  $\mathcal{K}$  is a superposition of product states where, in each product, the first member is from  $\mathcal{H}$  and the second is from  $\mathcal{A}$ . Specifically, the two inputs now correspond to  $|\psi_1\rangle|\phi_0\rangle$  and  $|\psi_2\rangle|\phi_0\rangle$ , where  $|\phi_0\rangle$  describes the initial state of the ancilla. We choose the unitary transformation as

$$U(|\psi_1\rangle|\phi_0\rangle) = \sqrt{p_1}|\psi'_1\rangle|\phi_0\rangle + \sqrt{q_1}|\psi_0\rangle|\phi_1\rangle , \quad (11.14)$$

$$U(|\psi_2\rangle|\phi_0\rangle) = \sqrt{p_2}|\psi'_2\rangle|\phi_0\rangle + \sqrt{q_2}e^{i\theta}|\psi_0\rangle|\phi_1\rangle , \quad (11.15)$$

where  $|\phi_1\rangle$  is chosen to be orthogonal to  $|\phi_0\rangle$ , and  $|\psi'_1\rangle$  and  $|\psi'_2\rangle$  correspond to orthogonal vectors in the original Hilbert space. If we now perform a von Neumann measurement on the ancilla then a click along the  $|\phi_1\rangle$  direction collapses both inputs onto the same output,  $|\psi_0\rangle$ , and all information about the inputs is lost. The probability that this happens for input  $i$  is  $q_i$  ( $i = 1, 2$ ). Obviously, this outcome corresponds to the inconclusive result so  $q_i$  are the failure probabilities of the corresponding input states  $i$ . On the other hand, a click along the  $|\phi_0\rangle$  direction transforms the original inputs into orthogonal outputs in the system Hilbert space. The probability that this happens for input  $i$  is  $p_i$  ( $i = 1, 2$ ). Obviously, this outcome corresponds to full distinguishability in the original system Hilbert space, so  $p_i$  are the probabilities of success for discriminating the corresponding input states  $i$ . From unitarity we obtain  $p_i + q_i = 1$  for  $i = 1, 2$  and by taking the inner product of Eq. (11.14) and Eq. (11.15) we obtain Eq. (11.7). Thus Neumark’s theorem delivers the positivity condition in a few lines and from here the rest of the derivation of the optimal failure probability goes along the same lines as for the POVM method. An optical implementation, based on the tensor product extension of the Hilbert space and employing linear optical elements (beam splitters and phase shifters) only, has been proposed in [14]. The states to be discriminated are represented by a single photon that can be in a superposition between two input ports of a six port interferometer with three input and three output ports. The third input corresponds to the ancilla, initially in the vacuum state. The desired unitary transformation is carried out by this six port constructed from the appropriate linear optical elements. At the output side detectors are placed in front of each output port. If the photon emerges from the ancilla port 3, the measurement failed. We can construct the device so that if the photon emerges from port  $i$  ( $i = 1, 2$ ), and we know that the input state was  $i$ . Thus, detector clicks in the first two output ports correspond to successful measurements. Measurements of

this type can easily be generalized higher dimensional systems, using more general multiports. An explicit example will be given in section 3.2.

Before moving on to more uncharted territory we should note that the example of unambiguous discrimination between two nonorthogonal polarization states by using a polarization-sensitive absorber that was considered at the beginning of this section does not conform to the scheme described here, although it clearly accomplishes its goal. Rather, it corresponds to another kind of possible extension of the original system Hilbert space via the direct sum method. In this case we append the ancilla space,  $\mathcal{A}$ , to the system space  $\mathcal{H}$  to form the joint Hilbert space  $\mathcal{K}$  of the ‘system plus ancilla’.  $\mathcal{K}$  is now a direct sum,  $\mathcal{K} = \mathcal{H} \oplus \mathcal{A}$ , of  $\mathcal{H}$  and  $\mathcal{A}$ , meaning that a state in  $\mathcal{K}$  is a superposition of two terms where the first one lies entirely in  $\mathcal{H}$  and the second one in  $\mathcal{A}$ . In particular, the two input states,  $|\psi_1\rangle$  and  $|\psi_2\rangle$ , have no components in the ancilla space. In the example with the two single photon polarization states spanning the two-dimensional system Hilbert space this means that we simply include a third state, the vacuum, as the ancilla space. We can say that our two-dimensional objects, the qubits, secretly live in the three-dimensional space of qutrits. The described unambiguous discrimination procedure between two single-photon polarization states has been successfully performed experimentally, implementing the polarization-selective attenuation either by polarization-dependent absorption in a fiber [15] or with the help of polarizing beamsplitters [16]. The transformation corresponding to the linear absorber redistributes the population among the three basis states in such a way that the parts of the input qubit states that remain in the original qubit Hilbert space become orthogonal there. In this method we do not keep track of what happens to the photon in the absorption process, whereas in the tensor product method, including a complete description of the ancilla (environment), we do. For the simple examples that we consider, one can introduce equivalent notations for the two alternative methods. Therefore, in the next section we will employ the somewhat simpler notations of the direct sum method.

This completes our tutorial review of the unambiguous discrimination of two pure states and illustrates the two possible approaches, the direct POVM method and the method of generalized measurements based on Neumark’s theorem. We now turn our attention to more complicated problems and briefly review recent progress in dealing with them.

#### 11.2.4 More than two pure states

So far we have only considered discriminating between two states, and we have seen that in that case a complete solution can be given. For more than two states, however, there are only a few general results, and explicit solutions exist only for special cases. Here we shall review what is known.

Two general results apply to the case of unambiguous discrimination. The first, due to Chefles, is that only linearly independent states can be

unambiguously discriminated [17]. This can be seen as follows. Let the POVM for discriminating the  $N$  states  $|\psi_1\rangle, \dots, |\psi_N\rangle$  be given by the operators  $A_1, \dots, A_N$ , and  $A_I$ , where the operators act on the vectors in the space  $\mathcal{H}$ , which is the span of the vectors  $|\psi_1\rangle, \dots, |\psi_N\rangle$ , and

$$A_I^\dagger A_I + \sum_{j=1}^N A_j^\dagger A_j = I, \quad (11.16)$$

which is an obvious generalization of Eq. (11.3) to  $N$  states. The operator  $A_I$  again corresponds to the inconclusive outcome, and the operator  $A_j$ , for  $j = 1 \dots N$  corresponds to identifying the state as  $|\psi_j\rangle$ . Because there must be no errors, we must have

$$\langle\psi_k|A_j^\dagger A_j|\psi_k\rangle = p_j \delta_{jk}, \quad (11.17)$$

where  $0 \leq p_j \leq 1$  is the probability of successfully identifying  $|\psi_j\rangle$ . Now suppose that the states are linearly dependent so that they can be expressed in terms of each other

$$|\psi_j\rangle = \sum_{k=1}^N c_{jk} |\psi_k\rangle. \quad (11.18)$$

Substituting this into the above equation we have that

$$\sum_{m,n=1}^N c_{km}^* c_{kn} \langle\psi_m|A_j^\dagger A_j|\psi_n\rangle = p_j \delta_{jk}. \quad (11.19)$$

This can be simplified by noting that

$$|\langle\psi_m|A_j^\dagger A_j|\psi_n\rangle|^2 \leq \langle\psi_m|A_j^\dagger A_j|\psi_m\rangle \langle\psi_n|A_j^\dagger A_j|\psi_n\rangle, \quad (11.20)$$

which gives us that

$$\langle\psi_m|A_j^\dagger A_j|\psi_n\rangle = p_j \delta_{mn} \delta_{jm}. \quad (11.21)$$

Substituting this into Eq. (11.19) we find that  $|c_{kj}|^2 = \delta_{jk}$ , which implies that the states are not linear combinations of each other and are, hence, linearly independent.

It is also possible to quickly draw some conclusions about the form of the operator  $A_j$ . Because we have that

$$A_j|\psi_k\rangle = 0, \quad (11.22)$$

for  $j \neq k$ , it annihilates the subspace  $\mathcal{H}_k$ , which is the span of the vectors  $|\psi_1\rangle, \dots, |\psi_N\rangle$  with  $|\psi_k\rangle$  omitted. Let  $|\psi_j^\perp\rangle$  be the unit vector orthogonal to  $\mathcal{H}_k$  which, by the way, explains the notation introduced in the previous section. We can then choose

$$A_j = \frac{\sqrt{p_j}}{\langle\psi_j^\perp|\psi_j\rangle} |\psi_j'\rangle\langle\psi_j^\perp|, \quad (11.23)$$

where  $|\psi'_j\rangle$ ,  $j = 1, \dots, N$  are arbitrary orthogonal unit vectors. The remaining problem is to find the values of  $p_j$ . Let us denote the *a priori* probability of the state  $|\psi_j\rangle$  by  $\eta_j$ . The values of  $p_j$  should be chosen to maximize the average success probability,  $P$ , where

$$P = \sum_{j=1}^N \eta_j p_j . \quad (11.24)$$

In addition, they must be chosen so that the operator

$$\begin{aligned} A_I^\dagger A_I &= I - \sum_{j=1}^N A_j^\dagger A_j \\ &= I - \sum_{j=1}^N \frac{p_j |\psi_j^\perp\rangle\langle\psi_j^\perp|}{|\langle\psi_j|\psi_j^\perp\rangle|^2}, \end{aligned} \quad (11.25)$$

is positive. This is a nontrivial problem.

The second general result states that there do exist upper and lower bounds on the success probability. Here we shall just present the results; the interested reader can consult the original papers for the derivations. An upper bound is given by [18]

$$P \leq 1 - \frac{1}{N-1} \sum_{j=1}^N \sum_{k=1, k \neq j}^N \sqrt{\eta_j \eta_k} |\langle\psi_j|\psi_k\rangle|. \quad (11.26)$$

Building on work by Duan and Guo [19], X. Sun, *et al.* derived a lower bound [20]. Consider the  $N \times N$  matrix whose elements are  $\langle\psi_j|\psi_k\rangle$ , and let  $\lambda_N$  be the smallest eigenvalue of this matrix. They showed that  $P \geq \lambda_N$ .

The problem of discriminating among three nonorthogonal states was first considered by Peres and Terno [21]. They developed a geometric approach and applied it numerically to several examples. A different method was considered by Duan and Guo [19] and Y. Sun and ourselves [22]. We considered the three vectors to be discriminated,  $|\psi_j\rangle$ ,  $j = 1, 2, 3$ , to lie in the space  $\mathcal{H}$ . To this a “failure” space,  $\mathcal{A}$ , is appended so that the whole problem takes place in the space obtained by the direct sum extension,  $\mathcal{K} = \mathcal{H} \oplus \mathcal{A}$ . If the procedure fails, the vector  $|\psi_j\rangle$  is mapped into a vector in the failure space,  $|\phi_j\rangle$ , and if it succeeds it is mapped onto a vector in the original space,  $\sqrt{p_j}|\psi'_j\rangle$ , where  $\|\psi'_j\| = 1$ , and  $0 \leq p_j \leq 1$ . The vectors  $|\psi'_j\rangle$  are mutually orthogonal, so that they can be perfectly distinguished. Chefles showed that the set of failure vectors must be linearly dependent for the optimal procedure [17], so that the dimension of  $\mathcal{A}$  will be one or two. One way of understanding this result is that if the failure vectors were linearly independent, then we could perform a further unambiguous state discrimination procedure on them and, with some probability, tell which state we were originally given. This

would imply that the original procedure was not optimal. Therefore, the optimal procedure produces linearly dependent failure vectors, which cannot be further discriminated.

Making this more explicit, we assume that there is a unitary operator,  $U$ , acting on  $\mathcal{K}$ , such that

$$U|\psi_j\rangle = \sqrt{p_j}|\psi'_j\rangle + |\phi_j\rangle. \quad (11.27)$$

It should be noted that, unlike in Eq. (11.15), the vector  $|\phi_j\rangle$  is not normalized to unity. Instead, we use  $\langle\phi_j|\phi_j\rangle = q_j$  here. After  $U$  has been performed, we measure the projection operator onto the space  $\mathcal{H}$ . If we obtain 1, the procedure has succeeded, and we know what the input state was. If the input was  $|\psi_j\rangle$ , the procedure succeeds with probability  $p_j$ . If we obtain 0, the procedure has failed, and this happens with probability  $q_j = 1 - p_j = \|\phi_j\|^2$ , if the input state was  $|\psi_j\rangle$ . The above equation implies that

$$\langle\phi_j|\phi_k\rangle = \langle\psi_j|\psi_k\rangle - \delta_{jk}p_j. \quad (11.28)$$

Defining the matrix  $C_{jk} = \langle\phi_j|\phi_k\rangle$ , we see by its definition that it must be positive definite. Therefore, the problem of finding the optimal unambiguous state discrimination procedure reduces to finding the values of  $p_j$  that optimize the success probability

$$P = \sum_{j=1}^3 \eta_j p_j, \quad (11.29)$$

subject to the constraint that the  $3 \times 3$  matrix, whose elements are  $\langle\psi_j|\psi_k\rangle - \delta_{jk}p_j$ , is positive.

This can be solved in some special cases. We shall assume that all of the a priori probabilities are the same, so that they are all  $1/3$ . If all of the overlaps are the same, i.e.

$$\langle\psi_1|\psi_2\rangle = \langle\psi_1|\psi_3\rangle = \langle\psi_2|\psi_3\rangle = s, \quad (11.30)$$

where  $s$  is real and positive, then  $q_j = s$ , for  $j = 1, 2, 3$ , and  $Q = 1 - P = s$  as well.

There is also an explicit solution if

$$\begin{aligned} \langle\psi_1|\psi_2\rangle &= \langle\psi_1|\psi_3\rangle = s_1, \\ \langle\psi_2|\psi_3\rangle &= s_2, \end{aligned} \quad (11.31)$$

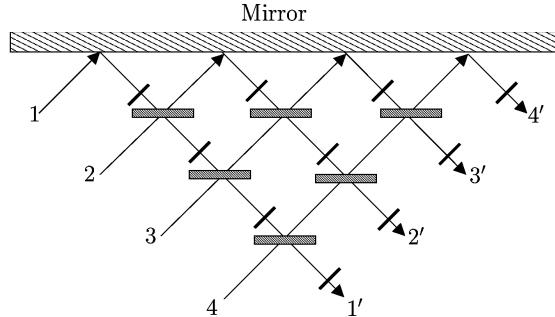
where both  $s_1$  and  $s_2$  are real and positive. We first note that for a fixed value of  $s_1$  there is a restriction on how large  $s_2$  can be. The largest the angle between  $|\psi_2\rangle$  and  $|\psi_3\rangle$  can be is twice the angle between  $|\psi_1\rangle$  and  $|\psi_2\rangle$  (this maximum is achieved when the vectors are coplanar). This implies that  $s_2 \geq 2s_1^2 - 1$ . The solution to the state discrimination problem depends on whether  $s_1/s_2 < 2$  or not. If it is, we have

$$\begin{aligned} q_1 &= \frac{s_1^2}{s_2}, \quad q_2 = q_3 = s_2, \\ Q &= \frac{1}{3} \left[ \frac{s_1^2}{s_2} + 2s_2 \right]. \end{aligned} \quad (11.32)$$

If  $s_1/s_2 \geq 2$ , we have

$$\begin{aligned} q_1 &= 2s_1, \quad q_2 = q_3 = s_1 + s_2, \\ Q &= \frac{2}{3}(2s_1 - s_2). \end{aligned} \quad (11.33)$$

This approach lends itself naturally to an optical implementation [22]. The states to be discriminated are represented by a single photon that can be in one of three modes. Additional modes (one or two, depending on the dimension of the failure space) that represent the ancilla, are initially in the vacuum state. Fig. 11.5 displays the simpler case, that of an eight port. In general, the unitary transformation is carried out by an optical  $N$  port, where  $N$  is either 8 or 10, depending on the number of vacuum ports needed. It should be noted that an  $N$  port is a linear optical device with  $N/2$  inputs and  $N/2$  outputs, that can be constructed from beam splitters, phase shifters, and mirrors. At the output detectors are placed to determine from which of the ports the photon emerges. If it emerges from one of the failure ports, the measurement has failed, but if it emerges from one of the other three, we know what the input state was.



**Fig. 11.5.** An optical eight-port. The beams are straight lines, a suitable beam splitter is placed at each point where two beams intersect, phase shifters are at one input of each beam splitter and at each output.

In particular, we can construct the device so that the three output ports that correspond to a successful measurement are numbered 1 through 3, and a photon emerging from port  $j$  means that the input state was  $|\psi_j\rangle$ . Measurements of this type have been carried out by Mohseni, *et al.* [23].

In the case that the failure space is two-dimensional, it is sometimes possible to obtain some information about the input state even if the initial measurement fails [21, 22]. Sun, *et al.* presented an example of an optical network that does the following [22]. It consists of an optical 10 port followed by a 6 port. The first three inputs of the 10 port are where the state  $|\psi_j\rangle$ ,  $j = 1, 2, 3$  is sent in, and the other two are in the vacuum state. The failure space for this particular situation is two-dimensional, and if the photon emerges from outputs 4 or 5, the measurement has failed. If it emerges from outputs 1, 2, or 3, we know what the input state was. The 6 port has as its three inputs the two failure outputs from the 10 port, and the vacuum. It is constructed so that if the photon comes out of the first output, we know that the input state was not  $|\psi_3\rangle$ , if it comes out of the second output, the input was not  $|\psi_2\rangle$ , and if it comes out of the third output, the measurement has failed. Therefore, even if the initial measurement (the 10 port) fails, there is some possibility of gaining information about the input state by further processing. Note that this depends on the failure space having two dimensions; with a one-dimensional failure space, no further processing is possible.

There is another case where the exact solution to the unambiguous discrimination problem is known, and to which we now turn our attention. First we note that a set of  $N$  states is called symmetric [3, 5] if there exists a unitary operator,  $V$ , such that, for  $j = 1, \dots, N - 1$ ,

$$V|\psi_j\rangle = |\psi_{j+1}\rangle, \quad V|\psi_N\rangle = |\psi_1\rangle. \quad (11.34)$$

This implies that  $|\psi_j\rangle = V^{j-1}|\psi_1\rangle$ . The case of unambiguous state discrimination for  $N$  symmetric states was analyzed byChefles and Barnett [24]. They found an analytical expression for the optimal success probabilities for the case when the *a priori* probabilities of the states are the same. The vectors  $|\psi_j\rangle$ ,  $j = 1, \dots, N$  are now assumed to be linearly independent, and to span the entire space. Because the states  $|\psi_j\rangle$  form a basis for the space, Eq. (11.34) now implies that  $V^N = I$ . This, in turn, means that  $V$  can be expressed as

$$V = \sum_{k=0}^{N-1} e^{2\pi i k/N} |\gamma_k\rangle \langle \gamma_k|, \quad (11.35)$$

where  $|\gamma_k\rangle$  is the eigenstate of  $V$  with eigenvalue  $e^{2\pi i k/N}$ . The states we are trying to distinguish among can now be expanded as

$$|\psi_j\rangle = \sum_{k=0}^{N-1} e^{2\pi i k(j-1)/N} c_k |\gamma_k\rangle. \quad (11.36)$$

The optimal success probability was found to be

$$P = N \min |c_k|^2, \quad (11.37)$$

where the minimum is taken over  $k$ , and throughout the derivation it was assumed that the *a priori* probabilities of the states were the same.

### 11.2.5 The discrimination of mixed states

We begin this part by introducing some terminology first. This will greatly facilitate the presentation of the material to follow and help us to interpret some of the existing statements about this topics in the literature.

The *support* of a mixed state, described by a density matrix, is the space spanned by its eigenvectors with nonzero eigenvalues. The *rank* of a mixed state is the dimension of its support. The *kernel* of a mixed state is the space orthogonal to its support. With these definitions at hand, we are ready to interpret the statement that “... one cannot unambiguously discriminate mixed states (the reason is that the IDP scheme does not work for linearly dependent states)” [25]. This is not the only statement of this sort but is a representative one that is obviously correct. But we have to elaborate it further. It clearly refers to mixed states that have the same support and excludes mixed states that have a nonzero overlap with the kernels of the others. In the following we will focus our attention on precisely those cases, admittedly a small subset of all mixed states of a system, where the density operators to be discriminated have different supports. It is only within this subset that unambiguous discrimination of mixed states is possible.

By the time of the publication of Ref. [25], several other works have been published that can be interpreted as special instances of unambiguous discrimination of mixed states. Below we give a brief overview of recent progress in this area.

#### Unambiguous filtering

In [26], we introduced the problem of unambiguous discrimination between sets of states. The set of  $N$  given states is divided into subsets, and we want to determine to which subset a particular input state, known to be prepared in one of the  $N$  given states, belongs. In the simplest case, the division is into two sets only, the first containing the first  $M$  states and the second the remaining  $N - M$  states. We want to unambiguously assign a given input state to one of these two subsets. Clearly, the discrimination of two pure states corresponds to  $N = 2$  and  $M = 1$ , there is one state in each set. The next simplest case is the one with  $N = 3$  and  $M = 1$ . This is the case of unambiguously discriminating whether a state is  $|\psi_1\rangle$  or whether it is in the set  $\{|\psi_2\rangle, |\psi_3\rangle\}$  with a priori probabilities  $\eta_1, \eta_2$ , and  $\eta_3$ , respectively. Since in this case all we are interested in is whether a particular input is  $|\psi_1\rangle$  or not we termed this case unambiguous quantum state filtering. (Actually, we introduced the term quantum state filtering first in the context of minimum-error discrimination [63] and this work will be discussed in Sec. 11.3.4.) First, it is straightforward to see that filtering is a particular instance of mixed state discrimination and, by a simple extension of the following considerations, set discrimination, in general, is equivalent to the discrimination of mixed

states. Indeed, since we do not want to resolve the states within a set, we can introduce the density operators

$$\rho_\alpha = |\psi_1\rangle\langle\psi_1|, \quad \rho_\beta = \frac{\eta_2}{\eta_2 + \eta_3}|\psi_2\rangle\langle\psi_2| + \frac{\eta_3}{\eta_2 + \eta_3}|\psi_3\rangle\langle\psi_3|, \quad (11.38)$$

with a priori probabilities  $\eta_\alpha = \eta_1$ ,  $\eta_\beta = \eta_2 + \eta_3$ . Filtering is, thus, the discrimination between these two density operators, one a rank one mixed state which is, in fact, a pure state, and the other a rank two mixed state.

Key to the solution is the observation that the ancilla space is one dimensional in the Neumark implementation of the optimal POVM. This immediately yields the condition analogous to Eq. (11.7),

$$q_1 q_i = |\langle\psi_1|\psi_i\rangle|^2, \quad (11.39)$$

where  $q_1$  and  $q_i$  ( $i = 2, 3$ ) are the failure probabilities for the corresponding input states. The derivation of the optimum average failure probability is then very similar to the derivation of the IDP result for two pure states. The optimal failure probability,  $Q^{opt}$ , is given as

$$Q^{opt} = \begin{cases} Q^{POVM} & \text{if } \frac{F^2}{1+F^2} \leq \eta_\alpha \leq \frac{1}{1+F^2}, \\ Q_\alpha & \text{if } \eta_\alpha < \frac{F^2}{1+F^2}, \\ Q_\beta & \text{if } \frac{1}{1+F^2} < \eta_\alpha. \end{cases} \quad (11.40)$$

Here  $F = \sqrt{\langle\psi_1|\rho_\beta|\psi_1\rangle}$  is the fidelity between a pure and a mixed state [1], and

$$\begin{aligned} Q^{POVM} &= 2\sqrt{\eta_\alpha\eta_\beta}F, \\ Q_\alpha &= \eta_\alpha + \eta_\beta F^2, \\ Q_\beta &= \eta_\alpha \|\psi_1^\parallel\|^2 + \eta_\beta \frac{F^2}{\|\psi_1^\parallel\|^2}. \end{aligned} \quad (11.41)$$

In the last line  $|\psi_1^\parallel\rangle$  is the component of  $|\psi_1\rangle$  in the support of  $\rho_\beta$ . Written in terms of the fidelity between a pure and a mixed state, the solution remains valid for an arbitrary number of states in the second set and, so, it represents the solution for a rank one vs. rank  $N$  unambiguous discrimination problem where  $N$  is arbitrary [27]. An interesting application of this result for a probabilistic quantum algorithm will be presented in Sec. 11.2.6. We note that unambiguous discrimination between multiple sets of pure states is a straightforward generalization and was independently investigated by Zhang and Ying in [28]. It, too, can be recast in a form of discriminating between mixed states. The problem of filtering a mixed state out of many was addressed by Takeoka, *et al.* in Ref. [29].

### State comparison

In Ref. [30], Barnett, Chefles, and Jex introduced the problem of state comparison that can be stated as follows. Given two systems, each of which is in one of two, in general, nonorthogonal states  $\{|\psi_1\rangle, |\psi_2\rangle\}$ , what is the optimum probability to determine whether the two systems were prepared in the same state or in different states? One of the surprising aspects of their result is that one can decide with a certain probability whether the two systems are not in the same state even if there is no prior knowledge of the possible states. In this context we note that discrimination between non perfectly known states has also been addressed by Ježek in Ref. [31].

More importantly for our purpose, the comparison of states can be cast to a form that corresponds to the discrimination between mixed states described by rank two density matrices. Obviously, we now want to discriminate between the sets  $\{|\psi_1\rangle|\psi_1\rangle, |\psi_2\rangle|\psi_2\rangle\}$  and  $\{|\psi_1\rangle|\psi_2\rangle, |\psi_2\rangle|\psi_1\rangle\}$ . The first set contains the combined states where the two systems are in the same individual state with a priori probabilities  $\eta_1$  and  $\eta_2$  for the respective combined states and the second set contains the states where the two systems are in different individual states with a priori probabilities  $\eta_3$  and  $\eta_4$  for the respective combined states. Introducing density operators for the sets, by employing a slight extension of the method in Eq. (11.38), we see that state comparison is equivalent to the discrimination of two rank two mixed states [32]. In Ref. [33], we derived the optimal POVM failure probability solution for the unambiguous discrimination between a rank two and a rank  $N$  density matrix,

$$Q^{POVM} = 2\sqrt{\eta_\alpha \eta_\beta} F , \quad (11.42)$$

where  $F$  is the fidelity between the two mixed states  $\rho_\alpha$  and  $\rho_\beta$  [1]. In the case of state comparison, the intersection between the two supports is one dimensional and the fidelity can easily be calculated to give  $F = |\langle\psi_1|\psi_2\rangle|$ . Thus, for the optimal failure probability of state comparison,  $Q_{SC}$ , we get

$$Q_{SC} = 2\sqrt{\eta_\alpha \eta_\beta} |\langle\psi_1|\psi_2\rangle| , \quad (11.43)$$

where  $\eta_\alpha = \eta_1 + \eta_2$  and  $\eta_\beta = \eta_3 + \eta_4$  are the a priori probabilities for the respective mixed states. This result slightly generalizes the result of [30], based on a previously unpublished result of our own [33]. In a subsequent work [34], Jex, Andersson, and Chefles extended the state comparison results from two systems to the comparison of many. By employing similar methods to the ones here, this can be shown to be equivalent to the discrimination between density matrices of higher rank.

### Two arbitrary mixed states: General considerations

Generalizing the ideas of Refs. [26]-[34], two important works have appeared recently. In the first, Rudolph, *et al.* [35] established lower and upper bounds

on the minimum failure probability for the unambiguous discrimination of two mixed states. Based on Uhlmann's theorem [1] they found that the lower bound is given by the fidelity, and the upper bound is based on the geometrical invariants between the kernels. They showed that for all known solutions the upper and lower bounds coincide and found numerically in other cases that the two bounds are very close. As an application of their method, they also provided the general solution for the one-dimensional kernel problem and applied it to the special case of two rank  $N - 1$  density matrices in an  $N$ -dimensional Hilbert space.

Raynal, *et al.* [36], introduced reduction theorems for the problem of optimal unambiguous discrimination of two general density matrices of rank  $N$  and  $M$ . In particular, they showed that the problem can be reduced to the discrimination of two density matrices that have the same rank  $N_0$  where  $N_0$  is bounded by  $N_0 \leq \min(N, M)$ . Necessary and sufficient conditions for optimality were discussed also in [25] and [37], along with some numerical methods based on linear programming.

The upper and lower bounds on the failure probabilities by Ref. [35] and the reduction theorems by Ref. [36] bring us very close to a full solution of the unambiguous discrimination problem between two arbitrary mixed states and it is fully expected that a full solution will be found in the near future.

### 11.2.6 Selected applications

We shall conclude this Section with a discussion of two applications of unambiguous state discrimination. The first is a quantum cryptographic protocol, while the second is a quantum algorithm.

Bennett proposed using the unambiguous discrimination of two nonorthogonal states as the basis of a quantum-key-distribution protocol [38]. Alice and Bob want to establish a secure key that they can use to send encrypted messages to each other. To do so, Alice sends Bob a sequence of particles, where each particle is either in the state  $|\psi_1\rangle$  or  $|\psi_2\rangle$ . The state  $|\psi_1\rangle$  corresponds to a bit value of 0 and  $|\psi_2\rangle$  corresponds to a bit value of 1. These states are known to both Alice and Bob, and they are not orthogonal. Upon receiving a particle, Bob applies the optimal two-state unambiguous measurement procedure to it. He then tells Alice over a public channel whether the measurement succeeded or failed. If it succeeded, they keep the bit, and if it failed, they discard the bit. In this way they can establish a key.

An eavesdropper, Eve, who wants to determine the key without being discovered has a problem. Let us assume that she can intercept the particles Alice is sending to Bob, and that she knows the states  $|\psi_1\rangle$  and  $|\psi_2\rangle$ . Because these states are not orthogonal, she cannot tell with certainty which state a particular particle is in. One possibility is for her to apply the same procedure used by Bob, optimal two-state unambiguous state discrimination. If her measurement succeeds, all is well. She simply notes which state she found, and prepares another particle in this state and sends it on to Bob. She then

knows this key bit, and Alice and Bob do not know that she knows. However, if her measurement fails, she does not know which state Alice sent, and she has to make a guess which state to send on to Bob. That means that she will introduce discrepancies between the state that Alice sent and the state that Bob received. In some of the cases in which this happens, Bob's measurement will succeed, and this will cause errors in the key to appear. If Alice and Bob publicly share a subset of their good key bits (these bits have to then be discarded), and if they see discrepancies, then they know an eavesdropper was present, and that the key is insecure. If they find none, then with high probability (Eve could get lucky and have all of her measurements succeed, but this is very improbable) the key is secure.

Our second example is a probabilistic quantum algorithm to discriminate between sets of Boolean functions [27]. A Boolean function on  $n$  bits is one that returns either 0 or 1 as output for every possible value of the input  $x$ , where  $0 \leq x \leq 2^n - 1$ . The function is constant if it returns the same output on all of its arguments, i.e. either all 0's or all 1's; it is balanced if it returns 0's on half of its arguments and 1's on the other half; and it is biased if it returns 0's on  $m_0$  of its arguments and 1's on the remaining  $m_1 = 2^n - m_0$  arguments ( $m_0 \neq m_1 \neq 0$  or  $2^n - 1$ ). Classically, if one is given an unknown function and told that it is either balanced or constant, one needs  $2^{(n-1)} + 1$  measurements to decide which. Deutsch and Jozsa [39] developed a quantum algorithm that can accomplish this task in one step. To discriminate a biased Boolean function from an unknown balanced one,  $2^{(n-1)} + m_1 + 1$  measurements are needed classically, where, without loss of generality, we have assumed that  $m_1 < m_0$ . There is a probabilistic quantum algorithm, based on quantum state filtering, that can unambiguously discriminate a known biased Boolean function from a given set of balanced ones. There is a significant chance that only one function evaluation will be necessary.

The algorithm distinguishes between sets of Boolean functions. Let  $f(x)$ , where  $0 \leq x \leq 2^n - 1$ , be a Boolean function, i.e.  $f(x)$  is either 0 or 1. One of the sets we want to consider is a set of balanced functions. The second set has only two members, and we shall call it  $\mathcal{W}_k$ . A function is in  $\mathcal{W}_k$  if  $f(x) = 0$  for  $0 \leq x < [(2^k - 1)/2^k]2^n$  and  $f(x) = 1$  for  $[(2^k - 1)/2^k]2^n \leq x \leq 2^n - 1$ , or if  $f(x) = 1$  for  $0 \leq x < [(2^k - 1)/2^k]2^n$  and  $f(x) = 0$  for  $[(2^k - 1)/2^k]2^n \leq x \leq 2^n - 1$ . We now wish to distinguish between the balanced functions and functions in  $\mathcal{W}_k$ , that is, we are given an unknown function that is in one of the two sets, and we want to find out which set it is in. We note that the two functions in  $\mathcal{W}_k$  are biased functions, so that this is a special case of a more general problem of distinguishing a set of biased functions from balanced functions.

The Deutsch-Jozsa algorithm makes use of the unitary operation

$$|x\rangle|y\rangle \rightarrow |x\rangle|y + f(x)\rangle, \quad (11.44)$$

where the first state,  $|x\rangle$ , is an  $n$ -qubit state, the second state,  $|y\rangle$ , is a single qubit state, and the addition is modulo 2. The state  $|x\rangle$ , where  $x$  is an  $n$ -

digit binary number, is a member of the computational basis for  $n$  qubits, and the state  $|y\rangle$ , where  $y$  is either 0 or 1, is a member of the computational basis for a single qubit. In solving the Deutsch-Jozsa problem, this mapping is employed in the following way

$$\sum_{x=0}^{D-1} |x\rangle(|0\rangle - |1\rangle) \rightarrow \sum_{x=0}^{D-1} (-1)^{f(x)} |x\rangle(|0\rangle - |1\rangle), \quad (11.45)$$

where  $D = 2^n$ . This has the effect of mapping Boolean functions to vectors in the  $D$ -dimensional Hilbert space,  $\mathcal{H}_D$ , and we shall do the same. The final qubit is not entangled with the remaining  $n$  qubits and can be discarded. The vectors  $\sum_{x=0}^{D-1} (-1)^{f(x)} |x\rangle$  that are produced by balanced functions are orthogonal to those produced by constant functions. This is why the Deutsch-Jozsa problem is easy to solve quantum mechanically. In our case, the vectors produced by functions in  $\mathcal{W}_k$  are not orthogonal to those produced by balanced functions. However, unambiguous quantum state filtering provides a probabilistic quantum algorithm for the optimal solution of this problem.

In order to apply the filtering solution, we note that both functions in  $\mathcal{W}_k$  are mapped, up to an overall sign, to the same vector in  $\mathcal{H}_D$ , which we shall call  $|w_k\rangle$ . The vectors that correspond to balanced functions are contained in the subspace,  $\mathcal{H}_b$ , of  $\mathcal{H}_D$ , where  $\mathcal{H}_b = \{|v\rangle \in \mathcal{H}_D | \sum_{x=0}^{D-1} v_x = 0\}$ , and  $v_x = \langle x|v\rangle$ . This subspace has dimension  $2^n - 1 = D - 1$ , and it is possible to choose an orthonormal basis,  $\{|v_i\rangle | i = 2, \dots, D\}$ , for it in which each basis element corresponds to a particular balanced Boolean function [40].

Let us first see how the filtering procedure performs when applied to the problem of distinguishing  $|w_k\rangle (= |\psi_1\rangle)$  from the set of the  $D - 1$  orthonormal basis states,  $|v_i\rangle (= |\psi_i\rangle)$ , in  $\mathcal{H}_b$ . We assume their a priori probabilities to be equal, i.e.  $\eta_i = \eta = (1 - \eta_1)/(D - 1)$  for  $i = 2, \dots, D$ , where  $\eta_1$  is the a priori probability for  $|w_k\rangle$ . For  $\|\psi_1\|^2 = \|w_k\|^2 \equiv f_k$  we obtain  $f_k = (2^k - 1)/2^{2k-2}$ . Then the average overlap,  $S_k$ , between  $|w_k\rangle$  and the set of balanced basis vectors can be written as

$$S_k = \frac{1 - \eta_1}{D - 1} f_k, \quad (11.46)$$

in terms of  $f_k$  [40]. The failure probabilities are given by the filtering result, using  $S = S_k$  and, to good approximation, the POVM result holds when  $1/2^{k-2} \leq D\eta_1 \leq 2^{k-2}$ . For example, in the case in which all of the a priori probabilities are equal, i.e.  $\eta_1 = 1/D$ , we find that  $Q_1 = Q_2 \equiv Q^{SQM} = (1 + f_k)/D$  where SQM stands for Standard Quantum Measurement (or von Neumann projective measurement). To good approximation,  $Q^{POVM}/Q^{SQM} = 4/2^{k/2}$ , which, for  $k \gg 1$ , shows that the POVM can perform significantly better than the von Neumann measurements.

Now that we know how this procedure performs on the basis vectors in  $\mathcal{H}_b$ , we shall examine its performance on any balanced function, i.e. we apply it to the problem of distinguishing  $|w_k\rangle$  from the set of all states in  $\mathcal{H}_b$  that correspond to balanced functions. The number of such states is  $N =$

$D!/(D/2)!$ <sup>2</sup> and we again assume their a priori probabilities to be equal,  $\eta = (1-\eta_1)/N$ . It can be shown [40] that the average overlap between  $|w_k\rangle$  and the set  $\{|v\rangle\}$  is given by the same expression, Eq. (11.46), as in the previous case. Therefore, much of what was said in the previous paragraph remains valid for this case, as well, with one notable difference. The case  $\eta_1 = 1/D$  now does not correspond to equal a priori probability for the states but, rather, to a priori weight of the sets that is proportional to their dimensionality. In this case it is the POVM that performs best. In the case of equal a priori probability for all states,  $\eta_1 = 1/(N+1)$ , we are outside of the POVM range of validity and it is the first standard quantum measurement (SQM1) that performs best. Both the POVM and the SQM1 are good methods for distinguishing functions in  $\mathcal{W}_k$  from balanced functions. Which one is better depends on the a priori probabilities of the functions.

Classically, in the worst case, one would have to evaluate a function  $2^n[(1/2)+(1/2^k)]+1$  times to determine if it is in  $\mathcal{W}_k$  or if it is an even function. Using quantum information processing methods, one has a very good chance of determining this with only one function evaluation. This shows that Deutsch-Jozsa-type algorithms need not be limited to constant functions; certain kinds of biased functions can be discriminated as well.

## 11.3 State discrimination with minimum error

### 11.3.1 Introductory remarks

In the previous chapter we have required that, whenever a definite answer is returned after a measurement on the state, the result should be unambiguous, at the expense of allowing inconclusive outcomes to occur. For many applications in quantum communication, however, one wants to have conclusive results only. This means that errors are unavoidable when the states are non-orthogonal. Based on the outcome of the measurement, in each single case then a guess has to be made as to what the state of the quantum system was. This procedure is known as *quantum hypothesis testing*. The problem consists in finding the optimum measurement strategy that minimizes the probability of errors.

Let us state the optimization problem a little more precisely. In the most general case, we want to distinguish, with minimum probability of error, between  $N$  given states of a quantum system ( $N \geq 2$ ), being characterized by the density operators  $\rho_j$  ( $j = 1, 2, \dots, N$ ) and occurring with the given a priori probabilities  $\eta_j$  which sum up to unity. The measurement can be formally described with the help of a set of detection operators  $\Pi_j$  that refer to the possible measurement outcomes [3, 4]. They are defined in such a way that  $\text{Tr}(\rho_j \Pi_j)$  is the probability to infer the system is in the state  $\rho_j$  if it has been prepared in a state  $\rho$ . Since the probability is a real non-negative number, the detection operators have to be Hermitean and positive-

semidefinite. In the error-minimizing measurement scheme the measurement is required to be exhaustive and conclusive in the sense that in each single case one of the  $N$  possible states is identified with certainty and inconclusive results do not occur. This leads to the requirement

$$\sum_{j=1}^N \Pi_j = I_{D_S}, \quad (11.47)$$

where  $I_{D_S}$  denotes the unit operator in the  $D_S$ -dimensional physical state space of the quantum system. The overall probability  $P_{\text{err}}$  to make an erroneous guess for any of the incoming states is then given by

$$P_{\text{err}} = 1 - P_{\text{corr}} = 1 - \sum_{j=1}^N \eta_j \text{Tr}(\rho_j \Pi_j) \quad (11.48)$$

with  $\sum_j \eta_j = 1$ . Here we introduced the probability  $P_{\text{corr}}$  that the guess is correct. In order to find the minimum-error measurement strategy, one has to determine the specific set of detection operators that minimizes the value of  $P_{\text{err}}$  under the constraint given by Eq. (11.47). By inserting these optimum detection operators into Eq. (11.48), the minimum error probability  $P_{\text{err}}^{\min} \equiv P_E$  is determined. The explicit solution to the error-minimizing problem is not trivial and analytical expressions have been derived only for a few special cases.

### 11.3.2 Distinguishing two quantum states with minimum error

#### The Helstrom formula

For the case that only two states are given, either pure or mixed, the minimum error probability,  $P_E$ , was derived in the mid 70s by Helstrom [3] in the framework of quantum detection and estimation theory. We find it instructive to start by analyzing the two-state minimum-error measurement with the help of an alternative method (cf. [41, 42]) that allows us to gain immediate insight into the structure of the optimum detection operators, without applying variational techniques. Starting from Eq. (11.48) and making use of the relations  $\eta_1 + \eta_2 = 1$  and  $\Pi_1 + \Pi_2 = I_{D_S}$  that have to be fulfilled by the a priori probabilities and the detection operators, respectively, we see that the total probability to get an erroneous result in the measurement is given by

$$P_{\text{err}} = 1 - \sum_{j=1}^2 \eta_j \text{Tr}(\rho_j \Pi_j) = \eta_1 \text{Tr}(\rho_1 \Pi_2) + \eta_2 \text{Tr}(\rho_2 \Pi_1). \quad (11.49)$$

This can be alternatively expressed as

$$P_{\text{err}} = \eta_1 + \text{Tr}(\Lambda \Pi_1) = \eta_2 - \text{Tr}(\Lambda \Pi_2), \quad (11.50)$$

where we introduced the Hermitean operator

$$\Lambda = \eta_2 \rho_2 - \eta_1 \rho_1 = \sum_{k=1}^{D_S} \lambda_k |\phi_k\rangle\langle\phi_k|. \quad (11.51)$$

Here the states  $|\phi_k\rangle$  denote the orthonormal eigenstates belonging to the eigenvalues  $\lambda_k$  of the operator  $\Lambda$ . The eigenvalues are real, and without loss of generality we can number them in such a way that

$$\begin{aligned} \lambda_k < 0 &\quad \text{for} & 1 \leq k < k_0, \\ \lambda_k > 0 &\quad \text{for} & k_0 \leq k \leq D, \\ \lambda_k = 0 &\quad \text{for} & D < k \leq D_S. \end{aligned} \quad (11.52)$$

By using the spectral decomposition of  $\Lambda$ , we get the representations

$$P_{\text{err}} = \eta_1 + \sum_{k=1}^{D_S} \lambda_k \langle\phi_k|\Pi_1|\phi_k\rangle = \eta_2 - \sum_{k=1}^{D_S} \lambda_k \langle\phi_k|\Pi_2|\phi_k\rangle. \quad (11.53)$$

Our optimization task now consists in determining the specific operators  $\Pi_1$ , or  $\Pi_2$ , respectively, that minimize the right-hand side of Eq. (11.53) under the constraint that

$$0 \leq \langle\phi_k|\Pi_j|\phi_k\rangle \leq 1 \quad (j = 1, 2) \quad (11.54)$$

for all eigenstates  $|\phi_k\rangle$ . The latter requirement is due to the fact that  $\text{Tr}(\rho\Pi_j)$  denotes a probability for any  $\rho$ . From this constraint and from Eq. (11.53) it immediately follows that the smallest possible error probability,  $P_{\text{err}}^{\min} \equiv P_E$ , is achieved when the detection operators are chosen in such a way that the equations  $\langle\phi_k|\Pi_1|\phi_k\rangle = 1$  and  $\langle\phi_k|\Pi_2|\phi_k\rangle = 0$  are fulfilled for eigenstates belonging to negative eigenvalues, while eigenstates corresponding to positive eigenvalues obey the equations  $\langle\phi_k|\Pi_1|\phi_k\rangle = 0$  and  $\langle\phi_k|\Pi_2|\phi_k\rangle = 1$ . Hence the optimum detection operators can be written as

$$\Pi_1 = \sum_{k=1}^{k_0-1} |\phi_k\rangle\langle\phi_k|, \quad \Pi_2 = \sum_{k=k_0}^{D_S} |\phi_k\rangle\langle\phi_k|, \quad (11.55)$$

where the expression for  $\Pi_2$  has been supplemented by projection operators onto eigenstates belonging to the eigenvalue  $\lambda_k = 0$ , in such a way that  $\Pi_1 + \Pi_2 = I_{D_S}$ . Obviously, provided that there are positive as well as negative eigenvalues in the spectral decomposition of  $\Lambda$ , the minimum-error measurement for discriminating two quantum states is a von Neumann measurement that consists in performing projections onto the two orthogonal subspaces spanned by the set of states  $\{|\phi_1\rangle, \dots, |\phi_{k_0-1}\rangle\}$ , on the one hand, and  $\{|\phi_{k_0}\rangle, \dots, |\phi_{D_S}\rangle\}$ , on the other hand. An interesting special case arises when negative eigenvalues do not exist. In this case it follows that

$\Pi_1 = 0$  and  $\Pi_2 = I_{D_S}$  which means that the minimum error probability can be achieved by always guessing the quantum system to be in the state  $\rho_2$ , without performing any measurement at all. Similar considerations hold true in the absence of positive eigenvalues. We note that these findings are in agreement with the recently gained insight [44] that measurement does not always aid minimum-error discrimination. By inserting the optimum detection operators into Eq. (11.50) the minimum error probability is found to be [42]

$$P_E = \eta_1 - \sum_{k=1}^{k_0-1} |\lambda_k| = \eta_2 - \sum_{k=k_0}^D |\lambda_k|. \quad (11.56)$$

Taking the sum of these two alternative representations and using  $\eta_1 + \eta_2 = 1$ , we arrive at

$$P_E = \frac{1}{2} \left( 1 - \sum_k |\lambda_k| \right) = \frac{1}{2} (1 - \text{Tr}|\Lambda|), \quad (11.57)$$

where  $|\Lambda| = \sqrt{\Lambda^\dagger \Lambda}$ . Together with Eq. (11.48) this immediately yields the well-known Helstrom formula [3] for the minimum error probability in discriminating  $\rho_1$  and  $\rho_2$ ,

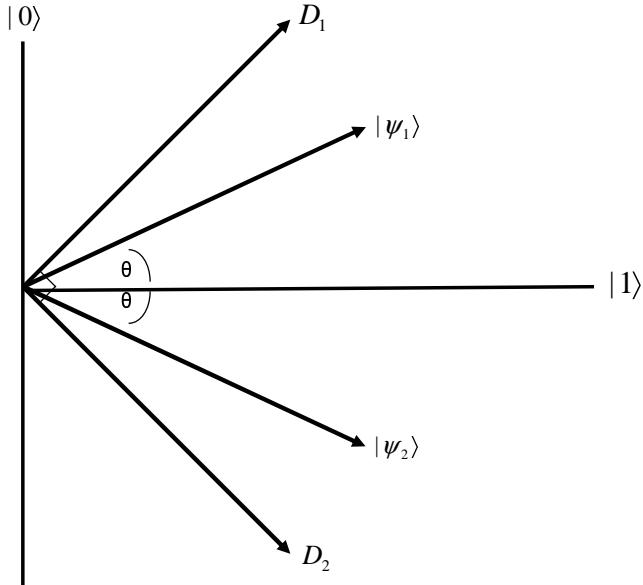
$$P_E = \frac{1}{2} (1 - \text{Tr}|\eta_2\rho_2 - \eta_1\rho_1|) = \frac{1}{2} (1 - \|\eta_2\rho_2 - \eta_1\rho_1\|). \quad (11.58)$$

In the special case that the states to be distinguished are the pure states  $|\psi_1\rangle$  and  $|\psi_2\rangle$ , this expression reduces to [3]

$$P_E = \frac{1}{2} \left( 1 - \sqrt{1 - 4\eta_1\eta_2|\langle\psi_1|\psi_2\rangle|^2} \right). \quad (11.59)$$

The set-up of the detectors that achieve the optimum error probabilities is particularly simple for the case of equal a priori probabilities. Two orthogonal detectors, placed symmetrically around the two pure states, will do the task, as shown in Fig. 11.6. The simplicity is particularly striking when one compares this set-up to that of Fig. 11.3, that displays the corresponding POVM set-up for optimal unambiguous discrimination.

In order to provide a simple and intuitive physical picture of an error-minimizing state discrimination measurement, let us consider the experiment performed by Barnett and Rijs [43] for distinguishing two equiprobable non-orthogonal single-photon polarization states, produced with the help of strongly attenuated light pulses. Imagine, we are given a series of very weak light pulses and we know in advance that, by preparation, each of the pulses is linearly polarized, with equal probability either in the direction  $e_1$  or in the direction  $e_2$  with  $e_{1,2} = \cos\theta e_x \pm \sin\theta e_y$ , where  $0 \leq \theta \leq \pi/4$ . Both kinds of pulses are assumed to have equal intensity and to contain on the average much less than one photon, so that the probability to obtain two photodetector clicks from a single pulse is negligible. Using a linear polarizer,



**Fig. 11.6.** Detector configuration for the optimum minimum-error discrimination of two pure states with equal a priori probabilities. A von Neumann measurement with two orthogonal detectors placed symmetrically around  $|\psi_1\rangle$  and  $|\psi_2\rangle$  will achieve the optimum.

we want to determine for each photon which polarization it had, by making the guess that it had polarization  $e_1$  when the photon is transmitted by the polarizer and had polarization  $e_2$  when the photon is absorbed in the polarizer. The problem of finding the optimum measurement strategy then amounts to determining the optimum orientation of the polarizer that minimizes the probability to make a wrong guess. It is easy to calculate that the error probability is smallest when the polarizer is oriented symmetrically with respect to the two given polarization directions, i. e. when its transmission direction is given by the unit vector  $e_p = (e_x + e_y)/\sqrt{2}$ . In this case the fraction of the intensity that is transmitted from the first pulse, which is also the probability, that the photon is indeed transmitted provided it is of the first kind, follows from the projection of  $e_1$  onto  $e_p$  and is given by  $\cos^2(\pi/4 - \theta)$ . The same quantity describes the probability that the photon is absorbed, provided it is of the second kind, as can be found by projecting  $e_2$  onto the vector perpendicular to  $e_p$ . Hence the resulting error probability is determined by

$$P_E = 1 - \cos^2\left(\frac{\pi}{4} - \theta\right) = \frac{1}{2}(1 - \sin 2\theta) = \frac{1}{2}(1 - \sqrt{1 - |\mathbf{e}_1 \mathbf{e}_2|^2}). \quad (11.60)$$

which reproduces Eq. (11.59) with  $\eta_1 = \eta_2 = 1/2$  when the scalar product  $\mathbf{e}_1 \mathbf{e}_2$  is replaced by the overlap  $\langle \psi_1 | \psi_2 \rangle$ .

### Minimum-error discrimination versus unambiguous discrimination for two mixed states

In the error-minimizing scheme for discriminating two different mixed states  $\rho_1$  and  $\rho_2$  of a quantum system, a non-zero probability of making a correct guess can always be achieved. However, it is clear that we can only distinguish the two mixed states unambiguously when, in the  $D_S$ -dimensional physical state space of the quantum system, there exists at least one state vector that has a non-zero probability of occurring in the first of the mixed states but will never be found in the second mixed state. Recently it has been shown that the maximum success probability for unambiguously discriminating the two given mixed states does not exceed a certain upper bound [35], depending on the states and their a priori probabilities. Consequently, the minimum failure probability,  $Q_F$ , cannot be smaller than a certain lower bound,  $Q_L$ . We now investigate the relation between this lower bound on the failure probability, on the one hand, and the minimum error probability for discriminating the same two states, on the other hand. The procedure that we shall apply is closely related to the derivation of inequalities between the fidelity and the trace distance [1].

Let us first consider the failure probability  $Q_F$  for unambiguously discriminating the two mixed states. From the result derived in [35] it follows that

$$Q_F \geq \begin{cases} 2\sqrt{\eta_1 \eta_2} F(\rho_1, \rho_2) \equiv Q_L & \text{if } F(\rho_1, \rho_2) < \sqrt{\frac{\eta_{\min}}{\eta_{\max}}} \\ \eta_{\min} + \eta_{\max}[F(\rho_1, \rho_2)]^2 \geq Q_L & \text{otherwise,} \end{cases} \quad (11.61)$$

where in the second line we have used the fact that the arithmetic mean cannot be smaller than the geometric mean. Here  $\eta_{\min}(\eta_{\max})$  is the smaller (larger) of the two a priori probabilities  $\eta_1$  and  $\eta_2$ , and  $F$  is the fidelity defined as  $F(\rho_1, \rho_2) = \text{Tr } O$ , where  $O = (\sqrt{\rho_2} \rho_1 \sqrt{\rho_2})^{1/2}$ . To allow for a comparison between  $P_E$  and  $Q_F$ , or  $P_E$  and  $Q_L$ , respectively, we need a suitable orthonormal basis. It has been shown [1, 41] that when the basis states are chosen to be the eigenstates  $\{|l\rangle\}$  of the operator  $\rho_2^{-1/2} O \rho_2^{-1/2}$ , the fidelity takes the form

$$F(\rho_1, \rho_2) = \sum_l \sqrt{\langle l | \rho_1 | l \rangle \langle l | \rho_2 | l \rangle} = \sum_l \sqrt{r_l s_l}. \quad (11.62)$$

Here  $\sum_l |l\rangle \langle l| = I$ , and we introduced the abbreviations  $r_l = \langle l | \rho_1 | l \rangle$  and  $s_l = \langle l | \rho_2 | l \rangle$ . The lower bound on the failure probability then obeys the equation

$$1 - Q_L = 1 - 2\sqrt{\eta_1 \eta_2} \sum_l \sqrt{r_l s_l} = \sum_l (\sqrt{\eta_1 r_l} - \sqrt{\eta_2 s_l})^2, \quad (11.63)$$

where the second equality sign is due to the relation  $\eta_1 + \eta_2 = 1$  and to the normalization conditions  $\text{Tr}\rho_1 = \sum_l r_l = 1$  and  $\text{Tr}\rho_2 = \sum_l s_l = 1$ .

Now we estimate the minimum error probability  $P_E$ , using the same set of basis states  $\{|l\rangle\}$ . From Eq. (11.57) and from the fact that  $\langle\phi_k|\phi_k\rangle = \sum_l |\langle\phi_k|l\rangle|^2 = 1$  it follows that

$$\begin{aligned} 1 - 2P_E &= \sum_k |\lambda_k| = \sum_l \sum_k |\lambda_k| |\langle\phi_k|l\rangle|^2 \\ &\geq \sum_l \left| \sum_k \lambda_k |\langle\phi_k|l\rangle|^2 \right| = \sum_l |\langle l|A|l\rangle|, \end{aligned} \quad (11.64)$$

where the last equality sign is due to the spectral decomposition (11.51). After expressing  $A$  in terms of the density operators describing the given states, we arrive at

$$\begin{aligned} 1 - 2P_E &\geq \sum_l |\langle l|\eta_1\rho_1 - \eta_2\rho_2|l\rangle| \\ &= \sum_l |\sqrt{\eta_1 r_l} - \sqrt{\eta_2 s_l}| |\sqrt{\eta_1 r_l} + \sqrt{\eta_2 s_l}|. \end{aligned} \quad (11.65)$$

By comparing the expressions on the right-hand sides of Eqs. (11.63) and (11.65), respectively, it becomes immediately obvious that  $1 - 2P_E \geq 1 - Q_L$ , or

$$P_E \leq \frac{1}{2}Q_L. \quad (11.66)$$

This means that for two arbitrary mixed states, occurring with arbitrary a priori probabilities, the smallest possible failure probability in unambiguous discrimination is at least twice as large as the smallest probability of errors in minimum-error discrimination of the same states.

### 11.3.3 The general strategy for minimum-error state discrimination

#### Formal solution for $N$ mixed states

We return now to the general problem of discriminating with minimum error between  $N$  given mixed states, where  $N$  is an arbitrary number. As has been outlined already in the introductory remarks, the problem amounts to finding the specific optimum detection operators that minimize the expression (11.48) under the constraint (11.47). It has been shown by Holevo [45] and Yuen *et al.* [46] that the set of detection operators  $\{\Pi_j\}$  determining the optimum measurement strategy must satisfy the necessary and sufficient conditions

$$\Pi_k(\eta_k \rho_k - \eta_j \rho_j) \Pi_j = 0 \quad (11.67)$$

$$\sum_j \eta_j \rho_j \Pi_j - \eta_k \rho_k \geq 0, \quad (11.68)$$

( $1 \leq j, k \leq N$ ), where the last equation expresses the fact that the eigenvalues of the operator on the left-hand side are non-negative. These conditions can be understood from the following considerations. In order to take the constraint (11.47) into account, it is possible to introduce a Lagrange operator  $\Gamma$ , in analogy to a Lagrangian multiplier. The optimization task is then equivalent to maximizing the operator functional

$$P_{\text{corr}}(\{\Pi_j\}, \Gamma) = \sum_j \text{Tr}[(\eta_j \rho_j - \Gamma) \Pi_j] + \text{Tr}\Gamma, \quad (11.69)$$

with  $\Gamma$  and each of the  $\Pi_j$  varying independently. Since  $P_{\text{corr}}$  is real,  $\Gamma$  has to be Hermitean. The necessary condition for an extremum,  $\delta P_{\text{corr}} = 0$ , yields for any  $j = 1, \dots, N$  the requirement that  $\text{Tr}[(\eta_j \rho_j - \Gamma) \delta \Pi_j] = 0$ . The detection operators can be varied by starting from their eigenstate expansion  $\Pi_j = \sum_l \pi_{jl} |\mu_{jl}\rangle \langle \mu_{jl}|$  and performing variations in the state vectors, yielding  $\delta \Pi_j = \sum_l \pi_{jl} |\mu_{jl}\rangle \langle \delta \mu_{jl}| + H.A.$ , where  $\pi_{jl} \geq 0$ . This leads to the necessary extremal condition [25]

$$(\eta_j \rho_j - \Gamma) \Pi_j = 0, \quad (11.70)$$

from which it follows that for any  $j, k$  the equations  $\eta_j \rho_j \Pi_j = \Gamma \Pi_j$  and  $\eta_k \Pi_k \rho_k = \Pi_k \Gamma^\dagger$  have to be fulfilled, where  $\Gamma$  is Hermitean. By multiplying the first equation from the left with  $\Pi_k$  and the second from the right with  $\Pi_j$  and taking the difference, the representation (11.67) for the necessary condition becomes immediately obvious. Now we discuss the second condition. First we note that

$$\Gamma = \sum_j \eta_j \rho_j \Pi_j, \quad (11.71)$$

which follows from Eq. (11.70) and from the constraint  $\sum_j \Pi_j = I$ . As can be seen from Eqs. (11.69) and (11.70), the extremal value is given by  $P_{\text{corr}} = \text{Tr}\Gamma$ . Certainly for the extremum of  $P_{\text{corr}}$  to be a maximum, its value cannot be smaller than the probability of correct guesses that would follow from other choices of the set of detection operators, in particular from the choice  $\{\Pi_j\} = \{I \ \delta_{j,k}\}$ , leading to  $P_{\text{corr}} = \text{Tr}(\eta_k \rho_k)$ . This means that for the extremum to be a maximum the relation  $\text{Tr}\Gamma \geq \text{Tr}(\eta_k \rho_k)$  has to be fulfilled for any  $k$ , which is indeed guaranteed by the requirement (11.68). The conditions (11.67) and (11.68) implicitly determine the solution of the general minimum-error discrimination problem.

### Survey of explicitly solvable special cases

In the following we give a brief overview that summarizes the cases where explicit analytical expressions for the optimum detection operators, and hence

for the minimum error probability, have been determined from the implicit general solution. The case that only two states are given,  $N = 2$ , has been already extensively discussed in the previous section. When the two given states are pure and occur with equal prior probability, the Helstrom bound derived there is a special case of a more general result, referring to the class of equiprobable and symmetric pure states. For these states the solution of the minimum-error discrimination problem, as derived by Ban *et al.* [47], will be discussed in a separate section. Recently Eldar *et al.* [48] and Chou and Hsu [49] obtained an extension of this solution to the case of  $N$  equiprobable states that are symmetric and mixed. A few other cases have been solved analytically, too. They include certain classes of linearly independent states [50] and also equiprobable states the projectors of which sum up to the identity [46]. Moreover, Barnett [51] found the minimum-error strategy for multiple symmetric pure states, and Andersson *et al.* [52] solved the case of three mirror symmetric pure states. With respect to the general problem of discriminating an arbitrary number of mixed states, Hunter [44] found the condition and gave the solution for those cases when the best strategy consists in making no measurement at all, but simply to guess always the state with the highest a priori probability. We encountered an example of such a case when considering the discrimination between two mixed states.

Apart from the analytical solutions, the minimum-error strategy has been also investigated numerically. In particular, Ježek *et al.* [53] proposed an algorithm for finding the optimum measurement by applying the theory of semi-definite programming.

### Generalized measurements

When the  $N$  given states of the quantum system are linearly independent, the minimum-error strategy for state discrimination is always a von Neumann measurement, as has been recently proved by Eldar [54]. This means that the detection operators are mutually orthogonal projection operators in the Hilbert space of the quantum system, fulfilling the relation  $\Pi_j \Pi_k = \delta_{jk} \Pi_j$ , where  $1 \leq j, k \leq N$ . We note that the quantum states are called linearly independent when the combined set of all the eigenvectors of the density operators  $\rho_j$  ( $j = 1, \dots, N$ ) forms a set of linearly independent state vectors. In this case the error-minimizing discrimination measurement is a projective measurement that can be realized by performing measurements on the original quantum system alone. Generally, however, this need not always be the case. In particular, when the number of different measurement outcomes, or the number of detection operators, respectively, exceeds the dimensionality of the physical state space of the quantum system, the detection operators cannot be represented by projection operators in that state space, as becomes immediately obvious from the constraint  $\sum_j \Pi_j = I_{D_S}$ . The state discrimination measurement then has to be described as a generalized measurement, based on positive-operator valued measures, and the detection operators  $\Pi_j$

are also called POVM-elements. According to Neumark's theorem [13], any generalized measurement can be realized with the help of a unitary transformation and a projective measurement in an extended Hilbert space.

In order to illustrate these general considerations we find it worthwhile to briefly consider a prominent example for a generalized minimum-error measurement, which consists in discriminating between the three states of a single qubit defined as

$$\begin{aligned} |\psi_1\rangle &= -\frac{1}{2} \left( |0\rangle + \sqrt{3} |1\rangle \right), \\ |\psi_2\rangle &= -\frac{1}{2} \left( |0\rangle - \sqrt{3} |1\rangle \right), \\ |\psi_3\rangle &= |0\rangle, \end{aligned} \quad (11.72)$$

where  $|0\rangle$  and  $|1\rangle$  are the orthonormal basis states of the qubit. The three given states form an overcomplete set of symmetric states that is known as the trine ensemble [55]. Provided that the occurrence of each of the three states is equally probable, the optimum detection operators for distinguishing among them with minimum error, or the optimum POVM-elements, respectively, are given by [55]

$$\Pi_j = A_j^\dagger A_j = \frac{2}{3} |\psi_j\rangle\langle\psi_j|, \quad (11.73)$$

for  $j = 1, 2, 3$ . The probability of correctly identifying any of the trine states in the error-minimizing measurement is  $2/3$  and the probability of making an error is  $1/3$ . Since the states  $|\psi_j\rangle$  are normalized, it is clear that the operators  $\Pi_j$  do not represent projection operators. For a physical implementation of the generalized measurement, means have to be found for extending the originally two-dimensional Hilbert space of the system. Then a unitary transformation on the extended system of basis states has to be performed in such a way that a final von Neumann measurement realizes the specific projective measurement that is necessary for applying Neumark's theorem.

As discussed before, there are two conceptually different ways of achieving an extension of the dimensionality of the Hilbert space. The first amounts to defining the original quantum system in such a way that auxiliary quantum states can be directly added. The extended Hilbert space is then the direct sum of the Hilbert space spanned by the states of the original system and of the Hilbert space spanned by the auxiliary states, being also called ancilla states. For the trine ensemble, it is possible to associate the three two-dimensional non-normalized detection states  $A_j = \sqrt{2/3} |\psi_j\rangle$  with three orthonormal states in three dimensions, given by

$$|\psi_j\rangle = \sqrt{\frac{2}{3}} |\psi_j\rangle + \sqrt{\frac{1}{3}} |2\rangle, \quad (11.74)$$

where the auxiliary normalized state  $|2\rangle$  is chosen to be orthogonal to the two basis states  $|0\rangle$  and  $|1\rangle$ . By performing the von-Neumann measurement

that consists of the three projections  $|\psi_j\rangle\langle\psi_j|$  ( $j = 1, 2, 3$ ) in the enlarged, i. e. three-dimensional Hilbert space, the required generalized measurement is realized in the original two-dimensional Hilbert space of the qubit [56]. To implement this scheme, an original atomic qubit system can be defined to consist of only two electronic states of a multi-level atom. The unitary transformation necessary for the generalized measurement can then be accomplished with the help of a third level, by appropriately redistributing the population of the three atomic levels using sequences of Raman transitions induced by specially taylored classical pulses [56]. Finally the resulting level population is detected for performing the projective measurement. Explicit theoretical proposals have been made for using this scheme in order to realize both optimum unambiguous discrimination between two nonorthogonal pure qubit states, and minimum-error discrimination for the trine ensemble [56]. Similarly, a single-photon qubit system could be represented by only two out of three possible input modes, or input ports, respectively, of an optical network built of beam splitters and phase shifters. The unitary transformation then would be implemented by this linear network, and the projective measurement could be performed by detecting the photon at one of the output ports. It is obvious that the direct-sum representation of extending the Hilbert space relies on assuming that the original qubit secretly consists of two components of a qutrit [57].

The second way of enlarging the Hilbert space relies on coupling the original system to an auxiliary system, or ancilla system, with the help of a physical interaction. The Hilbert space of the combined system is the tensor product of the Hilbert spaces of both subsystems. As has been stated by Jozsa, *et al.* [58], from a physical point of view the adjoining of an additional ancilla system is the only available means of extending a space while retaining the original system intact. The formalism of (nonorthogonal) POVMs is then a mathematical artifice that expresses the residual effect on the original system when a von Neumann measurement is performed on the combined system after the interaction [58]. Clarke, *et al.* demonstrated experimentally the minimum-error discrimination for the trine ensemble represented by polarization states of a single photon, using an interferometric set-up [55]. They also performed the optimal error minimizing measurements on the tetrad ensemble, which, as its name implies, consists of four states [55]. The points corresponding to these states on the Bloch sphere lie on the corners of a tetrahedron.

### **Error-minimizing discrimination with a fixed number of inconclusive results**

As an extension of the measurement strategy described so far, the error minimizing discrimination strategy has been also studied under the condition that inconclusive results are now allowed to occur, but with a fixed prescribed probability. This probability is assumed to be smaller than the

minimum failure probability resulting from the measurement scheme for optimum unambiguous discrimination, where errors do not occur at all. The optimum measurement minimizing the error under this prescribed condition is then intermediate between generic minimum-error discrimination and optimum unambiguous discrimination. It was first investigated for pure states byChefles and Barnett [59] and by Zhang et al. [60]. Later the method was generalized to the case of mixed states by Fiurášek and Ježek [25] and by Eldar [37]. The possible occurrence of inconclusive results has to be accounted for in the basic equations by introducing an additional detection operator  $\Pi_0$  such that  $\text{Tr}(\rho\Pi_0)$  describes the probability to get an inconclusive outcome provided the system is in a state  $\rho$ . The given fixed value of this probability, or the given failure probability, respectively, modifies the constraint on the detection operators. By optimization, the minimum probability for the occurrence of errors in the conclusive outcomes can be determined.

#### 11.3.4 Selected problems of minimum-error discrimination

##### Distinguishing $N$ symmetric pure states

We conclude our review of minimum-error discrimination by a more detailed consideration of a few special problems. In particular, in this connection we also present the results of some of our own recent investigations. To begin with, we recall a pure-state discrimination problem that is exactly solvable and has found wide application in quantum communication. It consists in the so called *square-root measurement* that discriminates with minimum error between  $N$  equally probable symmetric states. Symmetric pure states are defined in such a way that each state results from its predecessor by applying a unitary operator  $V$  in a cyclic way [47],

$$|\psi_j\rangle = V|\psi_{j-1}\rangle = V^{j-1}|\psi_1\rangle, \quad |\psi_1\rangle = V|\psi_N\rangle, \quad (11.75)$$

implying that  $V^N = I$ . For the case that the states occur with equal a priori probability, i. e. that  $\eta_j = 1/N$  for each of the states, Ban et al. found that the optimum detection operators for minimum-error discrimination are given by [47]

$$\Pi_j = A_j^\dagger A_j = B^{-1/2}|\psi_j\rangle\langle\psi_j|B^{-1/2} \equiv |\mu_j\rangle\langle\mu_j|, \quad (11.76)$$

where

$$B = \sum_{j=1}^N |\psi_j\rangle\langle\psi_j|. \quad (11.77)$$

The states  $|\mu_j\rangle = B^{-1/2}|\psi_j\rangle$  are in general non-normalized and are called detection states. It is obvious that the special structure of the detection operators, or of the detection states, respectively, suggests the name "square-root-measurement". The minimum error probability  $P_E$  for this measurement is [47]

$$P_E = 1 - \frac{1}{N} \sum_{j=1}^N |\langle \mu_j | \psi_j \rangle|^2, \quad (11.78)$$

in accordance with the fact that in the corresponding optimized measurement scheme the quantum system is inferred to have been prepared in the state  $|\psi_j\rangle$  provided that the state  $|\mu_j\rangle$  is detected. When the detection states  $|\mu_j\rangle$  are orthonormal, the detection operators are projection operators and the minimum-error measurement is a von-Neumann measurement, otherwise it is a generalized measurement. The latter always holds true when the number of states exceeds the dimensionality of the physical state space of the quantum system, as can be immediately seen from the fact that the detection operators have to sum up to the unit operator in that space. In this case the given states are linearly dependent and form an overcomplete set in the Hilbert space of the system.

Let us apply the general solution in order to investigate minimum-error discrimination for the set of the  $N$  symmetric states

$$|\psi_j\rangle = \sum_{k=1}^D c_k e^{i\frac{2\pi}{N} j(k-1)} |\gamma_k\rangle, \quad (N \geq D), \quad (11.79)$$

where the coefficients  $c_k$  are arbitrary non-zero complex numbers with  $\sum_k |c_k|^2 = 1$ , and the states  $|\gamma_k\rangle$  ( $k = 1, \dots, D$ ) form a  $D$ -dimensional orthonormal basis. The given symmetric states are non-orthogonal except for the case that both the conditions  $N = D$  and  $|c_k|^2 = 1/N$  are fulfilled. For distinguishing them with minimum error, provided that they occur with equal a priori probability, we obtain the optimum detection states

$$|\mu_j\rangle = \frac{1}{\sqrt{N}} \sum_{k=1}^D \frac{c_k}{|c_k|} e^{i\frac{2\pi}{N} j k} |\gamma_k\rangle, \quad (11.80)$$

yielding the minimum error probability [61]

$$P_E = 1 - \frac{1}{N} \left( \sum_{k=1}^D |c_k| \right)^2. \quad (11.81)$$

When  $N > D$  the given symmetric states are linearly dependent and form an overcomplete set. In this case the detection states  $|\mu_j\rangle$  are non-orthogonal and non-normalized, with  $\langle \mu_j | \mu_j \rangle = D/N$ . When  $N = D$ , however, the states  $|\psi_j\rangle$  are linearly independent and therefore can be discriminated also unambiguously, as has been pointed out in the corresponding section of this review. Assuming equal a priori probabilities, the minimum failure probability,  $Q_F$ , for unambiguous discrimination of symmetric states has been derived to be  $Q_F = 1 - N \min |c_k|^2$  [24]. Comparing this with the expression for  $P_E$ , the minimum error probability is found to be smaller than  $Q_F$ . It is worth mentioning that the minimum error probability,  $P_E$ , on the one hand, and the

failure probability,  $Q_F$ , on the other hand, have been considered as distinguishability measures for ordering different ensembles of  $N$  equally probable symmetric pure states, and it has been found that these two measures impose different orderings [62].

We used the preceding general results for studying minimum-error discrimination between  $m$ -photon polarization states, referring to a fixed number  $m$  of indistinguishable photons ( $m = 1, 2, \dots$ ). These states are superposition states of the  $m + 1$  orthonormal polarization basis states

$$|\gamma_k^{(m)}\rangle = \frac{1}{\sqrt{(m-k)!k!}} (a_1^\dagger)^{m-k} (a_2^\dagger)^k |0\rangle \equiv |m-k, k\rangle, \quad (11.82)$$

where  $k = 0, 1, \dots, m$ . The basis states correspond to the  $m + 1$  different possibilities of distributing  $m$  indistinguishable photons among two orthogonal polarization modes, characterized by the photon creation operators  $a_1^\dagger$  and  $a_2^\dagger$ , respectively, and  $|0\rangle$  is the vacuum state of the field. A specific set of symmetric  $m$ -photon polarization states is defined by

$$|\psi_j^{(m)}\rangle = \frac{1}{\sqrt{m!}} \left( \cos \theta a_1^\dagger + \sin \theta e^{i\frac{2\pi}{N}j} a_2^\dagger \right)^m |0\rangle, \quad (11.83)$$

where  $j = 1, \dots, N$ . Interestingly, for  $N = 4$  and  $\theta = \pi/4$  these symmetric states are identical with the states that result when the standard protocol for quantum key distribution [65] is applied to  $m$ -photon pulses. Assuming again equal a priori probabilities, the minimum-error probability for discriminating the states reads for  $m \leq N - 1$

$$P_E^{(m)}(\theta) = 1 - \frac{1}{N} \left[ \sum_{k=0}^m \sqrt{\binom{m}{k}} \cos^{m-k} \theta \sin^k \theta \right]^2. \quad (11.84)$$

For the case of two-photon-polarization states,  $m = 2$ , we found [61] that for  $\theta = \pi/4$  the minimum error probability is given by  $P_E = 1 - (3 + 2\sqrt{2})/(2N)$  which is smaller than the value that would result for the corresponding single-photon polarization states, being  $P_E = 1 - 2/N$ . In general, the states  $|\psi_j^{(m)}\rangle$  can be considered to consist of  $m$  identical copies of indistinguishable qubits, or photons, respectively, being each in the state  $\cos \theta |1, 0\rangle + \sin \theta e^{i\frac{2\pi}{N}j} |0, 1\rangle$ . Therefore our results show that by performing a joint measurement on  $m$  copies, instead of a measurement on a single copy only, the probability of making a correct guess for the actual state can be enhanced. However, in order to physically realize a joint measurement of this kind, in many cases an  $m$ -photon interaction process would be necessary, while a measurement on a single copy can always be achieved much more simply with the help of linear optics. We also gave a recipe for a linear optical multiport performing the generalized measurement that discriminates with minimum error between  $N$  equiprobable symmetric single-photon polarization states and

we discussed how the corresponding two-photon-polarization states could be discriminated, at least in principle, with the help of polarization-dependent two-photon absorption [61].

### Subset discrimination and quantum filtering in a two-dimensional Hilbert space

While in the previous section we dealt with distinguishing between  $N$  individual pure states, we now turn to the error-minimizing discrimination between two subsets of a given set of  $N$  pure states. In our work [63] we studied this task for two sets of linearly dependent pure states that collectively span only a two-dimensional Hilbert space. Let us formulate the problem more precisely. We want to devise a measurement that allows us to decide, with the smallest possible error and without inconclusive answers, whether the actual state of the system belongs to the subset of states  $\{|\psi_1\rangle, \dots, |\psi_M\rangle\}$ , or to the complementary subset of the remaining states  $\{|\psi_{M+1}\rangle, \dots, |\psi_N\rangle\}$  with  $M < N$ . To avoid confusion, in this section we denote the a priori probabilities of the individual pure states by  $\eta_j^i$ , where  $j = 1, \dots, N$ . Note that for  $M = 1$  our subset-discrimination problem is also called minimum-error *quantum state filtering*, in correspondence to the problem of unambiguous quantum state filtering that has been treated in the respective section of this review. The detection operators  $\Pi_1$  and  $\Pi_2$ , referring to the two possible measurement outcomes for minimum-error subset-discrimination, are defined in such a way that the quantity  $\langle\psi_j|\Pi_1|\psi_j\rangle$  accounts for the probability to infer, from performing the measurement, the state of the system to belong to the first subset, if it has been prepared in the state  $|\psi_j\rangle$ . Obviously, this inference is correct if  $j \leq M$ . Similarly, the quantity  $\langle\psi_j|\Pi_2|\psi_j\rangle$  is defined as the probability for inferring the state to belong to the second subset. The overall error probability reads

$$P_{\text{err}}^{M(N)} = 1 - \left( \sum_{j=1}^M \eta_j^i \langle\psi_j|\Pi_1|\psi_j\rangle + \sum_{j=M+1}^N \eta_j^i \langle\psi_j|\Pi_2|\psi_j\rangle \right), \quad (11.85)$$

where  $\sum_{j=1}^N \eta_j^i = 1$  and  $\Pi_1 + \Pi_2 = I$ . In general, with respect to the optimum measurement strategies, the problem of subset-discrimination is equivalent to the problem of distinguishing between the two mixed states

$$\rho_1 = \frac{1}{\eta_1} \sum_{j=1}^M \eta_j^i |\psi_j\rangle\langle\psi_j| \quad \text{with} \quad \eta_1 = \sum_{j=1}^M \eta_j^i, \quad (11.86)$$

$$\rho_2 = \frac{1}{\eta_2} \sum_{j=M+1}^N \eta_j^i |\psi_j\rangle\langle\psi_j| \quad \text{with} \quad \eta_2 = \sum_{j=M+1}^N \eta_j^i, \quad (11.87)$$

provided that the mixed states occur just with the a priori probabilities given by  $\eta_1$  and  $\eta_2$ , respectively [63]. This equivalence also becomes immediately

obvious from comparing Eqs. (11.48) and (11.85). Thus the minimum error probability for subset-discrimination can be obtained by applying the Helstrom solution, Eq. (11.58), to the problem of discriminating between  $\rho_1$  and  $\rho_2$ .

In our work we adopted another approach that does not entail any increase in the overall calculation effort, but has the advantage of yielding direct information about the method that realizes the optimum measurement. We first proved that from the restriction to a two-dimensional Hilbert space it follows that the optimum detection operator  $\Pi_1$  (or  $\Pi_2$ , respectively) can be expressed as the projector onto a particular optimum pure state. By solving the extremal problem that minimizes the expression (11.85) we then determined this optimum pure state, as well as the resulting minimum value  $P_E^{M(N)}$ . Thus we found that the minimum probability of making an error in distinguishing to which of the two subsets  $\{|\psi_1\rangle, \dots, |\psi_M\rangle\}$  or  $\{|\psi_{M+1}\rangle, \dots, |\psi_N\rangle\}$  a given quantum state belongs is given by [63]

$$P_E^{M(N)} = \frac{1}{2} - \sqrt{R^2 + |S|^2}, \quad (11.88)$$

where  $R$  and  $S$  can be expressed as

$$R = \sum_{j=1}^M \eta_j^i \left( |\langle \psi_1 | \psi_j \rangle|^2 - \frac{1}{2} \right) - \sum_{j=M+1}^N \eta_j^i \left( |\langle \psi_1 | \psi_j \rangle|^2 - \frac{1}{2} \right), \quad (11.89)$$

$$\begin{aligned} S = & \sum_{j=1}^M \eta_j^i \frac{\langle \psi_2 | \psi_j \rangle \langle \psi_j | \psi_1 \rangle - \langle \psi_2 | \psi_1 \rangle \langle \psi_1 | \psi_j \rangle|^2}{\sqrt{1 - |\langle \psi_1 | \psi_2 \rangle|^2}} \\ & - \sum_{j=M+1}^N \eta_j^i \frac{\langle \psi_2 | \psi_j \rangle \langle \psi_j | \psi_1 \rangle - \langle \psi_2 | \psi_1 \rangle \langle \psi_1 | \psi_j \rangle|^2}{\sqrt{1 - |\langle \psi_1 | \psi_2 \rangle|^2}}. \end{aligned} \quad (11.90)$$

In particular, we applied this result to the case of quantum state filtering for three arbitrary, but linearly dependent states, spanning a two-dimensional Hilbert space. Let us introduce the matrix  $C$  by the definition  $C_{ij} = \langle \psi_i | \psi_j \rangle$  for  $i, j = 1, 2, 3$ . Linear dependence then implies  $\det(C) = 0$ , or

$$|\langle \psi_1 | \psi_2 \rangle|^2 + |\langle \psi_1 | \psi_3 \rangle|^2 + |\langle \psi_2 | \psi_3 \rangle|^2 = 1 + 2 \operatorname{Re} \langle \psi_1 | \psi_2 \rangle \langle \psi_1 | \psi_3 \rangle \langle \psi_2 | \psi_3 \rangle. \quad (11.91)$$

Therefore the minimum error probability obtained in [63] can be written as

$$P_E^{1(3)} = \frac{1}{2} - \sqrt{\frac{1}{4} - \eta_2^i \eta_3^i (1 - |\langle \psi_2 | \psi_3 \rangle|^2) - \eta_1^i \sum_{j=2}^3 \eta_j^i |\langle \psi_1 | \psi_j \rangle|^2}. \quad (11.92)$$

For  $\eta_3^i = 0$  the expression reduces to the Helstrom formula (11.59) for discriminating two pure states.

As an example, we considered three equally probable symmetric states,

$$|\psi_k\rangle = \cos\theta |\gamma_1\rangle + e^{i\frac{2\pi}{3}(k-1)} \sin\theta |\gamma_2\rangle, \quad k = 1, 2, 3, \quad (11.93)$$

where  $|\gamma_1\rangle$  and  $|\gamma_2\rangle$  denote two orthonormal basis states. For these states also the minimum error probability  $P_E$  for distinguishing them individually can be analytically expressed, cf. Eq. (11.81). We found that the ratio  $P_E^{1(3)}(\theta)/P_E$  varies between 0.56 (for  $\theta \approx \pi/12$ ) and 0.5 (for  $\theta = 0$  or  $\pi/4$ ). Hence the minimum-error probability for distinguishing one state from the set of the two others is only about half as large as the minimum-error probability for distinguishing all three states separately. For the special example of the equally probable trine states, given by Eq. (11.72), the minimum error probability for quantum state filtering is obtained to be  $P_E^{1(3)} = 1/6$ . The error-minimizing filtering strategy consists in performing a projection onto the state  $|\psi_1\rangle$  and guessing the system to be in this state when the detector clicks, and to be in one of the other states when a projection on the orthogonal state is successful. On the other hand, in the case of the trine states, optimum unambiguous quantum state filtering would yield the minimum failure probability  $Q_F^{1(3)} = 1/3$ , as can be verified from the formula given in the corresponding chapter of the review. In this case the measurement corresponds to projecting onto the direction orthogonal to  $|\psi_1\rangle$ , which unambiguously identifies the set of the other states. Obviously the minimized error probability and the optimized failure probability for unambiguous discrimination differ just by the factor one half, in agreement with the limit set by Eq. (11.66).

### Distinguishing a pure state from a uniformly mixed state in arbitrary dimensions

The solution of the minimum-error discrimination problem for two arbitrary quantum states, either pure or mixed, is well known and results in the compact Helstrom formula (11.58) for the minimum error probability,  $P_E$ . However, the explicit analytical evaluation of  $P_E$  poses severe difficulties when the dimensionality  $D$  of the relevant Hilbert space is larger than two. This is due to the fact that applying the Helstrom formula amounts to calculating the eigenvalues of a  $D$ -dimensional matrix. In the following we consider a simple yet non-trivial example of an error-minimizing state discrimination problem in an arbitrary dimensional Hilbert space that can be solved analytically, and, in addition, might be related to potential applications. Our problem consists in deciding with minimum error whether a quantum system is prepared either in a given pure state or in a given uniformly mixed state [42], i. e. we have to discriminate between the two quantum states described by

$$\rho_1 = |\psi\rangle\langle\psi|, \quad \rho_2 = \frac{1}{d} \sum_{j=1}^d |u_j\rangle\langle u_j|. \quad (11.94)$$

Here the states  $|u_j\rangle$  are supposed to be mutually orthonormal, i. e.  $\langle u_i|u_j\rangle = \delta_{ij}$ . With  $D_S$  denoting the dimensionality of the physical state space of the quantum system, the relation  $d \leq D_S$  has to be fulfilled. We note that in the special case  $d = D_S$  the state  $\rho_2$  is the maximally mixed state that describes a completely random state of the quantum system, containing no information at all. Discriminating between the density operators  $|\psi\rangle\langle\psi|$  and  $\rho_2$  then amounts to deciding whether the state  $|\psi\rangle$  has been reliably prepared, or whether the preparation has totally failed [44].

To simplify the representation, in the following we restrict ourselves to the special case that the a priori probabilities of the two states are  $\eta_1 = 1/(d+1)$  for the pure state, and  $\eta_2 = d/(d+1)$  for the mixed state, respectively [42]. This means that in the corresponding quantum state filtering scenario all possible pure states would have equal a priori probabilities. In order to calculate the minimum error probability, using the Helstrom formula (11.58), it is necessary to determine the eigenvalues  $\lambda$  of the operator

$$\Lambda = \frac{1}{d+1} \left( \sum_{j=1}^d |u_j\rangle\langle u_j| - |\psi\rangle\langle\psi| \right). \quad (11.95)$$

We found that the eigenvalues are given by [42]

$$\begin{aligned} \lambda_1 &= -\frac{1}{d+1} \sqrt{1 - \|\psi^\parallel\|^2}, \\ \lambda_2 &= -\lambda_1, \quad \lambda_k = \frac{1}{d+1} \quad (k = 3, \dots, d+1), \end{aligned} \quad (11.96)$$

where we introduced the notation  $|\psi^\parallel\rangle$  for the component of  $|\psi\rangle$  that lies in the subspace spanned by the states  $|u_1\rangle, \dots, |u_d\rangle$ ,

$$\|\psi^\parallel\|^2 = \langle\psi^\parallel|\psi^\parallel\rangle = \sum_{j=1}^d |\langle u_j|\psi\rangle|^2. \quad (11.97)$$

When the quantum states to be discriminated are linearly independent, i. e. when  $\|\psi^\parallel\| \neq 1$ , there exists exactly one eigenvalue that is negative, given by  $\lambda_1$ . Therefore according to Eq. (11.55) the detection operators for performing the minimum-error measurement are given by  $\Pi_1 = |\phi_1\rangle\langle\phi_1|$  and  $\Pi_2 = I_{D_S} - \Pi_1$ , where  $|\phi_1\rangle$  is the eigenstate belonging to the negative eigenvalue,  $\lambda_1$ . On the other hand, when  $\rho_1$  and  $\rho_2$  are linearly dependent, i. e. when  $\|\psi^\parallel\| = 1$ , a negative eigenvalue does not exist. In this case the optimum measurement strategy is described by the detection operator  $\Pi_2 = I_{D_S}$  which means that the resulting minimum error probability,  $P_E = 1/(d+1)$ , is achievable by guessing the system always to be in the state  $\rho_2$ , without performing any measurement at all. The minimum error probability resulting from the above eigenvalues reads [42]

$$P_E = \frac{1}{d+1} \left( 1 - \sqrt{1 - \|\psi^{\parallel}\|^2} \right). \quad (11.98)$$

We still mention that the previous considerations can be easily extended to the case that the pure state and the uniformly mixed state given in Eq. (11.94) occur with arbitrary a priori-probabilities  $\eta_1$  and  $\eta_2 = 1 - \eta_1$ , respectively. The minimum error probability for distinguishing between them is then given by [64]

$$P_E = \frac{1}{2} \left[ \eta_1 + \frac{\eta_2}{d} - \sqrt{\left( \eta_1 + \frac{\eta_2}{d} \right)^2 - 4\eta_1 \frac{\eta_2}{d} \|\psi^{\parallel}\|^2} \right]. \quad (11.99)$$

Let us now compare the minimum probability of errors,  $P_E$ , with the smallest possible failure probability,  $Q_F$ , that can be obtained in a strategy optimized for unambiguously discriminating between the pure state and the uniformly mixed state. The solution of the latter problem coincides with the solution to the problem of optimum unambiguous quantum state filtering. Assuming again that  $\eta_1 = 1/(d+1)$ , the minimum failure probability is  $Q_F = 2\|\psi^{\parallel}\|/(d+1)$  [42]. Supposing nonorthogonality of the two states, characterized by  $0 < \|\psi^{\parallel}\| \leq 1$ , we observe that  $P_E/Q_F \leq 1/2$ , in accordance with Eq. (11.66). When the two states are linearly dependent, i. e. when  $\|\psi^{\parallel}\| = 1$ , it follows that  $P_E/Q_F = 1/2$ . On the other hand, for nearly orthogonal states, where  $\|\psi^{\parallel}\| \ll 1$ , we find that  $P_E/Q_F \approx \|\psi^{\parallel}\|/4$ . Obviously in this case the minimum error probability is drastically smaller than the optimum failure probability for unambiguous discrimination.

As an application of the minimum-error strategy described above, we discussed the problem of discriminating between a pure and a mixed two-qubit state [42]. An arbitrary bipartite qubit state, shared among two parties A (Alice) and B (Bob), can be expressed with the help of the four orthonormal basis states  $|00\rangle, |01\rangle, |10\rangle$  and  $|11\rangle$ , where  $|mn\rangle$  stands for  $|m\rangle_A \otimes |n\rangle_B$ , with  $|0\rangle$  and  $|1\rangle$  denoting any two orthonormal basis states of a single qubit. As an interesting special case we considered the problem that Alice and Bob want to decide whether the two-qubit system is either in a given pure state  $|\psi\rangle$ , occurring with the a priori probability  $\eta_1 = 1/4$ , or in a uniform mixture of the three symmetric states

$$|u_1\rangle = |00\rangle, \quad |u_2\rangle = |11\rangle, \quad |u_3\rangle = \frac{1}{2}(|01\rangle + |10\rangle). \quad (11.100)$$

We found that in this case the minimum error probability is given by

$$P_E = \frac{1}{4} \left( 1 - \frac{1}{\sqrt{2}} |\langle 01|\psi\rangle - \langle 10|\psi\rangle| \right). \quad (11.101)$$

The same result would hold true if  $|u_1\rangle$  and  $|u_2\rangle$  were replaced by the two symmetric Bell states  $(|00\rangle \pm |11\rangle)/\sqrt{2}$ . Minimum-error discrimination is achieved by performing a projection measurement onto the eigenstate  $|\phi_1\rangle$  that belongs

to the negative eigenvalue  $\lambda_1$  of the operator  $\Lambda$ , cf. Eq. (11.95). In general, this eigenstate will be a superposition of the four two-qubit basis states. The optimum measurement strategy therefore requires a correlated measurement that has to be carried out collectively on the two qubits. On the other hand, if in our specific example the discrimination would have to be performed by local measurements on the qubits only, without any communication between Alice and Bob, the smallest achievable error probability would be always 1/4, independent of the choice of the state  $|\psi\rangle$  [42]. We note that the problem we considered is of particular interest in the context of quantum state comparison [30], where one wants to determine whether the states of quantum systems are identical or not. It has been shown [34] that for comparing two unknown single-particle states it is crucial to discriminate the anti-symmetric state of the combined two-particle system from the uniform mixture of the mutually orthogonal symmetric states. In this context it is interesting that recently an application of particle statistics to the problem of minimum-error discrimination has been presented [66]. Discrimination between bipartite states will be discussed in more detail in a separate chapter.

## 11.4 Discriminating multiparticle states

So far we have been considering the situation in which the person discriminating between states is in possession of the entire system that is guaranteed to be in one of the allowed states. However, if our system consists of subsystems, these can be distributed among different parties, and then these parties have to determine which state they are sharing by measuring their subsystems and communicating among themselves. This adds another layer of complexity to the problem.

The simplest example is to suppose that we have a two-qubit state, and we give one of the qubits to Alice and the other to Bob. Alice and Bob know that the state is either  $|\psi_0\rangle$  or  $|\psi_1\rangle$ , and by performing local operations and communicating classically (this is abbreviated as LOCC), they want to determine which state they have. We shall consider both the case of minimum-error and unambiguous discrimination. The object is to develop a procedure that Alice and Bob can use to discriminate between the states.

It is possible to immediately obtain some bounds on how successful these procedures can be. If both states are equally likely, and both qubits are measured together, then we know that the states can be successfully unambiguously discriminated with a probability of  $P_{IDP} = 1 - |\langle\psi_0|\psi_1\rangle|$  [7]- [9]. This clearly represents an upper bound on what can be accomplished using LOCC. In the case of minimum-error discrimination, the best probability of correctly identifying the state that is obtainable if both states are measured together is

$$P = \frac{1}{2} + \frac{1}{2}\text{Tr}|(\eta_0|\psi_0\rangle\langle\psi_0| - \eta_1|\psi_1\rangle\langle\psi_1|)|, \quad (11.102)$$

where  $\eta_j$  is the a priori probability for  $|\psi_j\rangle$ , for  $j = 1, 2$ . This is again an upper bound to what can be achieved using LOCC. A natural question is whether these bounds are, in fact, achievable. It was recently shown that they are.

In order to see how, let us start with an extreme case; we shall assume that  $|\psi_0\rangle$  and  $|\psi_1\rangle$  are orthogonal. Walgate, *et al.* proved that in this case the states can be distinguished perfectly using only LOCC [67]. They did, in fact, much more than this, they showed that two orthogonal states, of any dimension, shared by any number of parties can be perfectly distinguished by LOCC. Their proof rests on the fact that any two bipartite states can be expressed in the form

$$\begin{aligned} |\psi_0\rangle &= \sum_{j=0}^n |j\rangle_A |\xi_j\rangle_B \\ |\psi_1\rangle &= \sum_{j=0}^n |j\rangle_A |\xi_j^\perp\rangle_B, \end{aligned} \quad (11.103)$$

where  $\{|j\rangle_A | j = 1, \dots, n\}$  is an orthonormal basis for Alice's space and the states  $|\xi_j\rangle_B$  and  $|\xi_j^\perp\rangle_B$ , which are not normalized, are orthogonal in Bob's. Alice measures her state in the  $|j\rangle_A$  basis and communicates her result to Bob. If her result was  $|j_0\rangle$ , then Bob measures his particle in order to determine whether it is in the state  $|\xi_{j_0}\rangle_B$  or  $|\xi_{j_0}^\perp\rangle_B$ , which he can do perfectly since the states are orthogonal. Note that the measurement that Bob makes depends on the result of Alice's measurement. If the states are split among more than two parties, this procedure can be applied several times. For example, if there are three parties, Alice, Bob and Charlie, then we initially group Bob and Charlie together so that the state can be considered bipartite. Alice performs her measurement and tells Bob and Charlie the result. They now share one of two known, orthogonal states, and they can apply the above procedure again to find out which. The answer will tell them what the original state was.

The case when  $|\psi_0\rangle$  and  $|\psi_1\rangle$  are not orthogonal (and of arbitrary dimension) was investigated by Virmani, *et al.* [68], and they were able to apply the above decomposition to the problem of minimum-error discrimination. They found a strategy, for arbitrary a priori probabilities, using only LOCC that achieves the optimal success probability given in Eq. (11.102). In addition, they found strong numerical evidence that, when the two states are equally likely, unambiguous discrimination is possible with a probability of  $P_{IDP}$  using LOCC, and they found a class of states for which they could prove that this was true. A proof that this is true for all bipartite states was provided by Chen and Yang [69].

The procedure that makes LOCC unambiguous discrimination with a success probability of  $P_{IDP}$  possible is closely related to the one for discriminating orthogonal states. Alice makes a projective measurement on her particle that gives her no information about whether the state is  $|\psi_0\rangle$  or  $|\psi_1\rangle$ , and she

then communicates her result to Bob. Based on what Alice has told him, Bob chooses a measurement to make on his particle. In particular, he applies the procedure for the optimal unambiguous discrimination of single qubit states to his particle. However, in this procedure one must know the two states that one is discriminating between, and it is this information that is provided by the result of Alice's measurement.

Together with Mimih we took a somewhat different approach to unambiguous discrimination of two-qubit states [70]. The motivation was to study bipartite state discrimination schemes that could be used in quantum communication protocols, quantum secret sharing, in particular [71]- [74]. A quantum cryptography protocol based on two-state unambiguous discrimination already existed [38], and this suggested that the discrimination of bipartite states might find application as well. In secret sharing, Alice and Bob are both sent information that allows them to decode a message if they act together, but neither party can decode it by themselves. The schemes discussed above are not well suited for this type of application, because the information gained by the two parties is not the same. In particular, after the measurements have been made (note that Alice must communicate her result to Bob so that he can make his), Alice knows nothing and Bob knows which state has been sent. We were interested in schemes that are more symmetric.

Our approach was to examine situations in which the classical communication between the parties was limited. One possibility is to allow no classical communication. In that case each party has three possible measurement results, 0 corresponding to  $|\psi_0\rangle$ , 1 corresponding to  $|\psi_1\rangle$ , and  $f$  for failure to distinguish. If  $|\psi_0\rangle$  is sent, then Alice and Bob both measure 0 or both measure  $f$ , so that they both know, without communicating, that  $|\psi_0\rangle$  was sent or that the measurement failed. If  $|\psi_1\rangle$  is sent, then they both measure either 1 or  $f$ . In the case of qubits, we found that the best that can be done is to identify one of the states and fail the rest of the time, i.e. we never get a positive identification for the second state. The situation improves if we go to qutrits. In that case there are examples of states that can be distinguished with the success probability equal to  $P_{IDP}$ . One is given by the two states (the states  $|0\rangle$ ,  $|1\rangle$ , and  $|2\rangle$  are an orthonormal basis for the qutrits)

$$\begin{aligned} |\psi_0\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |22\rangle) \\ |\psi_1\rangle &= \frac{1}{\sqrt{2}}(|11\rangle + |22\rangle). \end{aligned} \quad (11.104)$$

If Alice and Bob each measure 0, they know they were sent  $|\psi_0\rangle$ , if they each measure 1, they know they were sent  $|\psi_1\rangle$ , and if they both measure 2 they failed.

We also considered the situation in which Alice and Bob make measurements, and later pool their results to determine which state was sent. However, conditional measurements were banned, i.e. situations in which the

measurement made by one party depends on the measurement result of the other were not allowed. Conditional measurements seem to have the problem that, as has been noted, the information the parties receive is not the same, and, that if there is a delay between the time the qubits are received and the time they are measured, then the qubits must be protected against decoherence until the measurement is made. In the procedures we studied, we found that it was optimal for each party to make projective measurements. If the two states shared the same Schmidt basis, then these procedures could successfully distinguish them with a probability of  $P_{IDP}$ , but if they did not, the probability of success was, in general, smaller than  $P_{IDP}$ .

The discrimination of more than two states shared by two parties has begun to be investigated, and presently results only exist for the case in which all of the states are orthogonal to each other. Ghosh, *et al.* have shown that it is not possible, in general, to deterministically distinguish either three or four orthogonal two-qubit states using only local operations and classical communication [75]. A general condition on when orthogonal, bipartite  $2 \times d$  states (one of the particles is a qubit and the other a qudit), can be distinguished by LOCC was found recently by Walgate and Hardy [76]. In the case of  $2 \times 2$  states, they found that for three of them to be perfectly distinguishable by LOCC, at least two of them must be product states, and for four, all of them must be product states. As far as we are aware, the problem of distinguishing three or more nonorthogonal states shared between two or more parties using LOCC has not yet been studied although, very recently, Chefles [77] derived a necessary and sufficient condition for a finite set of states in a finite dimensional multiparticle quantum system for LOCC unambiguous discrimination. This suggests that there is much still to be learned about distinguishing multipartite states using local operations and classical communication and, in general, separable quantum operations have to be considered.

## 11.5 Outlook

Quantum state discrimination is a very rapidly evolving field just like many other areas of quantum information and quantum computing. We have reviewed here the two most important - and simplest - state discrimination strategies, unambiguous discrimination and minimum-error discrimination. The minimum-error strategy for the discrimination of two mixed states has been one of the first problems that was solved exactly. With the recent progress toward the unambiguous discrimination of mixed states we expect that in the near future the problem of unambiguous discrimination of two mixed states will be solved completely. Then attention will quite naturally turn toward the discrimination of more than two states where so far only special cases were solved completely and partial results were obtained in the general case. A rapidly emerging field with a lot of room for quick progress is the discrimination of multiparticle states using LOCC only. Applications in

the area of quantum cryptography and probabilistic quantum algorithms will surely follow but in a field with such a rapidly changing landscape it would not be responsible to predict more than the immediately foreseeable future.

## References

1. M. A. Nielsen and I. L. Chuang: *Quantum Computation and Information* (Cambridge University Press, 2000)
2. A. Peres: *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1995)
3. C. W. Helstrom: *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976)
4. A. Chefles: Contemp. Phys. **41**, 401 (2000); arXiv:quant-ph/0010114 (2000).
5. M. Sasaki, S. M. Barnett, R. Jozsa, M. Osaki and O. Hirota: Phys. Rev. A **59**, 3325 (1999)
6. S. M. Barnett, C. R. Gilson, and M. Sasaki, J. Phys. A: Math. Gen. **34**, 6755 (2001); arXiv:quant-ph/0107024 (2001)
7. I. D. Ivanovic: Phys. Lett. A **128**, 257 (1987)
8. D. Dieks: Phys. Lett. A **126**, 303 (1988)
9. A. Peres: Phys. Lett. A **128**, 19 (1988)
10. G. Jaeger and A. Shimony: Phys. Lett. A **197**, 83 (1995)
11. M. Ban: Phys. Lett. A **213**, 235 (1996)
12. K. Kraus: *States, Effects and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983)
13. M. A. Neumark: Izv. Akad. SSSR Ser. Mat. **4**, (1940)
14. J. Bergou, M. Hillery, and Y. Sun: J. Mod. Opt. **47**, 487 (2000)
15. B. Huttner, A. Muller, J. D. Gautier, H. Zbinden, and N. Gisin: Phys. Rev A **54**, 3783 (1996)
16. R. B. M. Clarke, A. Chefles, S. M. Barnett, and E. Riis: Phys. Rev. A **63**, 040305(R) (2001)
17. A. Chefles: Phys. Lett. A **239**, 339 (1998)
18. S. Zhang, Y. Feng, X. Sun, and M. Ying: Phys. Rev. A **64**, 062103 (2001)
19. L. M. Duan and G. C. Guo: Phys. Lett. A **261**, 25 (1999)
20. X. Sun, S. Zhang, Y. Feng, and M. Ying: Phys. Rev. A **65**, 044306 (2002)
21. A. Peres and D. R. Terno: J. Phys. A **31**, 7105 (1998)
22. Y. Sun, M. Hillery, and J. A. Bergou: Phys. Rev. A **64**, 022311 (2001)
23. M. Masoud, A. Steinberg, and J. A. Bergou: in preparation
24. A. Chefles and S. Barnett: Phys. Lett. A **250**, 223 (1998)
25. J. Fiurášek and M. Ježek: Phys. Rev. A **67**, 012321 (2003); arXiv:quant-ph/0208126 (2002)
26. Y. Sun, J. A. Bergou, and M. Hillery: Phys. Rev. A **66**, 032315 (2002)
27. J. A. Bergou, U. Herzog, and M. Hillery: Phys. Rev. Lett. **90**, 257901 (2003)
28. S. Zhang and M. Ying: Phys. Rev. A **65**, 062322 (2002)
29. M. Takeoka, M. Ban, and M. Sasaki: Phys. Rev. A **68**, 012307 (2003)
30. S. M. Barnett, A. Chefles, and I. Jex: Phys. Lett. A **307**, 189 (2003)
31. M. Ježek: arXiv:quant-ph/0202085 (2002)
32. Y. Sun and J. Bergou: in preparation
33. J. Bergou, U. Herzog, and M. Hillery: in preparation

34. I. Jex, E. Andersson, and A. Chefles: quant-ph/0305120 (2003)
35. T. Rudolph, R. W. Spekkens, and P. S. Turner: Phys. Rev. A **68**, 010301(R) (2003)
36. Ph. Raynal, N. Lütkenhaus, and S. van Enk: Phys. Rev. A **68**, 022308 (2003)
37. Y. Eldar: Phys. Rev. A **67**, 042309 (2003)
38. C. H. Bennett: Phys. Rev. Lett. **68**, 3121 (1992)
39. D. Deutsch and R. Jozsa: Proc. R. Soc. London A **439**, 553 (1992)
40. J. A. Bergou, U. Herzog, and M. Hillery: in preparation
41. C. A. Fuchs: PhD thesis, Univ. of New Mexico (1995); arXiv:quant-ph/9601020 (1996)
42. U. Herzog: accepted for J. Phys. B; arXiv:quant-ph/0307038 (2003)
43. S. M. Barnett and E. Rijs: J. Mod. Opt. **44**, 1061 (1997)
44. K. Hunter: Phys. Rev. A **68**, 012306 (2003); arXiv:quant-ph/0211148 (2002)
45. A. S. Holevo: J. Multivar. Anal. **3**, 337 (1973)
46. H. P. Yuen, R. S. Kennedy, and M. Lax: IEEE Trans. Inform. Theory **IT-21**, 125 (1975)
47. M. Ban, K. Kurokawa, R. Momose, and O. Hirota: Int. J. Theor. Phys. **55**, 22 (1997)
48. Y. C. Eldar, A. Megretski, and G. C. Verghese: IEEE Trans. Inform. Theory **IT-49**, 1007 (2003); arXiv:quant-ph/0211111 (2002)
49. C.-L. Chou and L. Y. Hsu: Phys. Rev. A **68**, 042305 (2003); arXiv:quant-ph/03044117 (2003)
50. R. S. Kennedy, M. I. T. Res. Lab: Electron. Quart. Progr. Rep. **110**, 142 (1973)
51. S. M. Barnett: Phys. Rev. A **64**, 030303(R) (2001)
52. E. Andersson, S. M. Barnett, C. B. Gilson, and K. Hunter: Phys. Rev. A **65**, 052308 (2002); arXiv:quant-ph/0201074 (2002)
53. M. Ježek, J. Řeháček, and J. Fiurášek: Phys. Rev. A **65**, 060301 (2002); arXiv:quant-ph/0201109 (2002)
54. Y. Eldar: Phys. Rev. A **68**, 052303 (2003); arXiv:quant-ph/0304077 (2003)
55. R. B. M. Clarke, V. M. Kendon, A. Chefles, S. M. Barnett, E. Riis, and M. Sasaki: Phys. Rev. A **64**, 012303 (2001)
56. S. Franke-Arnold, E. Andersson, S. M. Barnett, and S. Stenholm: Phys. Rev. A **63**, 052301 (2001)
57. J. Preskill. In *Lecture Notes for Physics 229: Quantum Information and Computation* (Cambridge University Press, 1998)
58. R. Jozsa, M. Koashi, N. Linden, S. Popescu, S. Presnell, D. Shepherd, and A. Winter: arXiv:quant-ph/0303167 (2003)
59. A. Chefles and S. M. Barnett: J. Mod. Opt. **45**, 1295 (1998)
60. C.-W. Zhang, C.-F. Li, and G.-C. Guo: Phys. Lett. A, **261**, 25 (1999)
61. U. Herzog: Fortschr. Phys. **49**, 981 (2001); arXiv:quant-ph/0105139 (2001)
62. A. Chefles: Phys. Rev. A **66**, 042325 (2002)
63. U. Herzog and J. A. Bergou: Phys. Rev. A **65**, 050305(R) (2002)
64. U. Herzog and J. A. Bergou: in preparation
65. C. H. Bennett: Phys. Rev. Lett. **68**, 3121 (1992)
66. S. Bose, A. Ekert, Y. Omar, N. Paunković and V. Vedral: arXiv:quant-ph/0309090 (2003)
67. J. Walgate, A. Short, L. Hardy, and V. Vedral: Phys. Rev. Lett. **85**, 4972 (2000)

68. S. Virmani, M. F. Sacchi, M. B. Plenio, and D. Markham: Phys. Lett. A **288**, 62 (2001)
69. Yi-Xin Chen and Dong Yang: Phys. Rev. A **65**, 022320 (2002)
70. M. Hillery and J. Mimih: Phys. Rev. A **67**, 042304 (2003)
71. M. Hillery, V. Bužek, and A. Berthiaume: Phys. Rev. A **59**, 1829 (1999)
72. A. Karlsson, M. Koashi, and N. Imoto: Phys. Rev. A **59**, 162 (1999)
73. W. Tittel, H. Zbinden, and N. Gisin: Phys. Rev. A **63**, 042301 (2001)
74. R. Cleve, D. Gottesman, and H. -K. Lo: Phys. Rev. Lett. **83**, 648 (1999)
75. S. Ghosh, G. Kar, A. Roy, D. Sarkar, A. Sen(De), and U. Sen: Phys. Rev. A **65**, 062307 (2002)
76. J. Walgate and L. Hardy: Phys. Rev. Lett. **89**, 147901 (2002)
77. A. Chefles: arXiv:quant-ph/0302066 (2003)



# **12 Quantum States: Discrimination and Classical Information Transmission. A Review of Experimental Progress**

Anthony Chefles

Department of Physical Sciences, University of Hertfordshire, Hatfield AL10 9AB,  
Hertfordshire, UK  
Email:[A.Chefles@herts.ac.uk](mailto:A.Chefles@herts.ac.uk)

## **12.1 Introduction**

In recent years, there has been something of a renaissance in the study of quantum state discrimination. The reasons for this are manifold. Among them are the fact that quantum state discrimination is of enormous importance in quantum communications. The recent widespread interest in all things related to the information-theoretic properties of quantum systems [1] has led to renewed attention being given to the problem of distinguishing between quantum states.

The theoretical limitations on our ability to distinguish among a set of quantum states are described in this volume in the chapter by Hillery, Bergou and Herzog. See also [2]. These can be interpreted as limitations on our ability to distinguish among classical messages stored in these states. These limitations have direct impact on important practical quantum technologies such as quantum key distribution [3, 4].

A further reason why quantum state discrimination has been the focus of much recent attention is the fact that, due to advances in experimental techniques and improvements in apparatus, it is now possible to perform many of the high precision quantum measurements which are required to distinguish well among a set of quantum states. The purpose of this chapter is to review the achievements made in this field to date.

One topic which is closely related to quantum state discrimination is classical information transmission using quantum states. These are not quite, however, the same thing. In quantum state discrimination, a receiver must perform a measurement and use the result of this together with any prior information they may have about the set of possible states to determine, as accurately as possible, which state was actually sent. An ideal case would be that where the receiver has complete information about the actual state prior to transmission. However, in this situation, where the receiver already knows the state of a system prior to its transmission, the system cannot be used to send any information to the receiver [5].

Nevertheless, in spite of such scenarios, some measurements which perform optimal discrimination among a set of quantum states are closely related

to those required to optimally extract classical information from another set of quantum states. In fact, some of the key advances in experimental quantum state discrimination and classical information transmission were made because of a recognition of such relationships. For this reason, the two fields of experimental quantum state discrimination and classical information transmission with quantum states have become historically intertwined to the extent that a full presentation of one is impossible without a detailed consideration of the other. Consequently, this chapter also reviews the progress made to date in classical information transmission using quantum states.

Our focus will be on scenarios where we aim to send classical information in quantum states at the maximum possible rate. We will thus not consider quantum key distribution, where the rate of information transmission is compromised by the conflicting desire to ensure security. We will also not consider entanglement-assisted classical information transmission protocols, such as dense coding [6], where the transmitter and receiver share entangled quantum systems.

We begin with the simplest, non-trivial scenario in state discrimination: how well can one distinguish between two, non-orthogonal pure states? In section 12.2 we describe the experiment carried out by Barnett and Riis [7]. Here, the aim was to determine, with the minimum probability of error, which of these states was prepared. This experiment had two possible outcomes, with one corresponding to each of the possible states. Due to their non-orthogonality, each of the states could not be guaranteed, with unit probability, to give rise only to its own corresponding measurement outcome. It would, with some non-zero probability, lead to the outcome which corresponded to the other state. When this occurred, the result of the state discrimination attempt was erroneous. This aim of the Barnett-Riis experiment was to attain the theoretical minimum probability of error for two non-orthogonal states, which is given by the Helstrom bound [8]. In this experiment, the Helstrom bound was attained to an impressive degree of precision.

In the Barnett-Riis experiment, like in all the experiments we shall discuss, the actual quantum states used were quantum states of light. In this experiment, and in all others we shall discuss bar one, the quantum states were non-orthogonal optical polarisation states. However, in the final experiment which we shall discuss, which is a demonstration of a novel quantum effect known as classical capacity superadditivity, the quantum states were states of both polarisation and position.

We will describe this experiment at the end of this chapter. Following our discussion of the Barnett-Riis experiment, we shall discuss experiments which discriminate between two non-orthogonal states using a different strategy, known as unambiguous state discrimination. In minimum error discrimination, every result is conclusive, in the sense that the measurement always returns one of the possible input states as the answer. However, if we allow for inconclusive results, then something quite remarkable happens: we can

sometimes reduce the error probability of the conclusive results to zero! This can be done for any two non-orthogonal, but not identical, pure states. For a measurement of this kind, there are three potential outcomes: two of these correspond to each of the possible states and there is a further, inconclusive result.

A pair of pure quantum states spans a two dimensional Hilbert space, and so a conventional von Neumann measurement, based on orthogonal projections, will not be suitable here. We must consider more general measurement strategies which are known, appropriately, as generalised measurements. In section 12.3, we give a brief discussion of such measurements. Following this, we describe the two unambiguous state discrimination experiments which have been carried out to date. The first was carried out by Huttner et al [9], who ingeniously used a fibre loss mechanism to transform, with some probability, a pair of non-orthogonal optical polarisation states into an orthogonal pair. These could then be perfectly discriminated using standard polarisation filtering and photodetection. However, in this scheme, signals corresponding to inconclusive results could not be directly accessed as the photons corresponding to such results were absorbed by the system. In a subsequent experiment by Clarke et al [10], which did not rely on loss mechanisms, all three possible results were made directly accessible. In this experiment, a suitable arrangement of polarising beamsplitters was used to either transform the possible states into orthogonal states, or to make them completely indistinguishable, giving an inconclusive result. The maximum probability of implementing the former transformation was given by the Ivanovic-Dieks-Peres bound [11–13], and when it was implemented, the possible states could be perfectly distinguished by a von Neumann measurement. However, when the latter transformation took place, it was impossible to subsequently distinguish between the states and the result of any attempt to do so would be inconclusive.

We then turn our attention to discriminating among more than two possible states. Section 12.4 of this chapter is devoted to the discussion of discriminating, optimally, among sets of 3 and 4 optical polarisation states. As it happens, of the state discrimination strategies we have described so far, the only one which is possible for such sets of states is minimum error discrimination. The reason for this has to do with fundamental limitations on our ability to unambiguously discriminate among a set of quantum states. If the states are pure, then they can only be unambiguously discriminated if they form a linearly independent set [14]. The polarisation states of a single photon form a two-dimensional space. This space may be spanned, for example, by two linear polarisation states (e.g. horizontal and vertical) or two counter-rotating circular polarisation states. There can be at most two linearly independent states in such a space, and so, if we wish to discriminate among more than two optical polarisation states, unambiguous discrimination is not possible.

For sets of more than two states, the theoretical minimum error probability is often difficult to determine. This is clearly an impediment to optimal experimental discrimination among multiple states. However, there do exist certain, highly symmetrical sets of states for which the minimum error probability can be calculated. Among the sets for which this is so are overcomplete sets. In section 12.5, we describe experiments carried out by Clarke et al [15] on two such sets, the trine and tetrad sets, which have three and four states respectively. The appropriate measurements are closely related to that for unambiguous discrimination between two states. They are also related to the measurements required for optimal classical information extraction from such sets. These measurements were also implemented and are described.

In the theory of classical information transmission using quantum states, an important quantity is the accessible information. In section 12.6, we describe the experimental implementations of accessible information measurements carried out by Mizuno et al [16]. In addition to obtaining results for the trine states which improve upon those of Clarke et al, these authors carried out experiments involving sets of five (quinary) and seven (septenary) states.

The final topic we shall discuss, in section 12.7, is a novel quantum effect known as classical capacity superadditivity. This is an effect which enables one to extract more classical information from strings of quantum signals by performing collective measurements on multiple signal carriers than by measuring them individually. We describe the first experimental demonstration of this effect, which was recently carried out by Fujiwara et al [17]. This experiment, which followed an earlier theoretical suggestion by Peres and Wootters [18], shows that one can extract more classical information from quantum signals prepared in trine states by performing collective measurements on certain length-two strings of such states than by measurements on the individual signal carriers. We conclude in section 12.8 with a general discussion where we also describe some future goals.

## 12.2 Minimum error discrimination between two pure states

### 12.2.1 Background

Consider the following scenario: we have a two dimensional quantum system (a qubit) which is prepared in one of two pure states, which we shall write as  $|\psi_+\rangle$  and  $|\psi_-\rangle$ . We would like to determine which one it is in. It is well-known that when the states are non-orthogonal, it is impossible to perfectly distinguish between them: see Nielsen and Chuang [1] for a simple proof of this fact.

If the states are non-orthogonal, then we must settle for an imperfect strategy. Any strategy will involve a measurement, whose result gives us information about which state was prepared. The nature of this measurement

will depend on the nature of the information we wish to obtain. Here, we will consider a measurement whose result is always one of the possible states. That is, the measurement has two possible outcomes, ‘+’ and ‘−’, corresponding to the two possible states  $|\psi_+\rangle$  and  $|\psi_-\rangle$ .

Whenever the measurement result is ‘+’, we shall interpret this result to indicate that the state was  $|\psi_+\rangle$ , and likewise with ‘−’ and  $|\psi_-\rangle$ . However, the result will not always be correct if the states are non-orthogonal. In general, there will be a non-zero probability that, if the state was  $|\psi_+\rangle$ , then we will obtain the result ‘−’, and likewise with  $|\psi_-\rangle$  and ‘+’. Whenever either of these two situations arises, the result of the measurement is erroneous.

It is clearly of interest to calculate the maximum probability of discriminating between the states correctly, or equivalently, the minimum probability of obtaining an erroneous result. In minimum error discrimination, to calculate the overall error probability we must know the a priori probabilities of both states, that is, the probability that the system was prepared in each of these states. Let us denote the a priori probabilities of the states  $|\psi_+\rangle$  and  $|\psi_-\rangle$  by  $q_+$  and  $q_-$  respectively, where  $q_+ + q_- = 1$ . We must also know the error probabilities for both of the states. Let  $p(-|\psi_+)$  be the probability, given that the state was  $|\psi_+\rangle$ , that we obtain the result ‘−’. Similarly, let  $p(+|\psi_-)$  be the probability, given that the state was  $|\psi_-\rangle$ , we obtain the result ‘+’. The total probability of error is then given by the expression:

$$P_e = q_+ p(-|\psi_+) + q_- p(+|\psi_-). \quad (12.1)$$

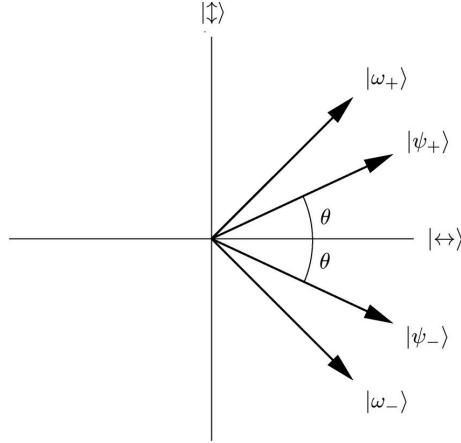
In this expression, the a priori probabilities  $q_+$  and  $q_-$  are fixed, whereas the values of  $p(-|\psi_+)$  and  $p(+|\psi_-)$  depend on the particular measurement. We are interested in varying these quantities over all possible, two-outcome measurements in order to obtain the global minimum value of the total error probability  $P_e$ . This optimisation was first carried out by Helstrom [8] and the resulting minimum error probability is given by the Helstrom bound

$$P_e(\min) = \frac{1}{2} \left( 1 - \sqrt{1 - 4q_+q_- |\langle\psi_+|\psi_-\rangle|^2} \right). \quad (12.2)$$

The measurement which attains this bound is a simple von Neumann measurement. By this, we mean that its statistical properties are based on orthogonal projection operators, in the following way. Corresponding to the results ‘+’ and ‘−’, there are orthogonal projection operators  $P_+$  and  $P_-$  such that  $P_+ + P_- = 1$ , where ‘1’ here is the identity operator. These operators are positive (that is, their expectation values are non-negative for all states) and they also satisfy  $P_\pm^2 = P_\pm$ . For any initial state represented by a density operator  $\rho$ , the probability of obtaining either outcome is simply equal to the expectation value of the corresponding projection operator for this state,

$$p(\pm|\rho) = \text{Tr}(P_\pm \rho). \quad (12.3)$$

In particular, for a pure initial state  $\rho = |\psi\rangle\langle\psi|$  we have  $p(\pm|\psi) = \langle\psi|P_\pm|\psi\rangle$ . In the experiment we shall describe, the actual states used were optical polar-



**Fig. 12.1.** States involved in the Helstrom measurement with equal a priori probabilities. The non-orthogonal states to be distinguished are the optical polarisation states  $|\psi_{\pm}\rangle = \cos(\theta)|\leftrightarrow\rangle \pm \sin(\theta)|\updownarrow\rangle$ , where the states  $|\leftrightarrow\rangle$  and  $|\updownarrow\rangle$  are orthogonal. The optimal measurement is a von Neumann measurement in the basis  $|\omega_{\pm}\rangle = (|\leftrightarrow\rangle \pm |\updownarrow\rangle)/\sqrt{2}$ , which gives the minimum error probability shown in (12.5).

isation states. Any pure state of this kind can be expressed as a superposition of two orthogonal polarisation states. Here, as throughout this chapter, the orthogonal set that we will use to represent more general polarisation states will be the linear horizontally and vertically polarised states. We shall denote these by  $|\leftrightarrow\rangle$  and  $|\updownarrow\rangle$  respectively. The superpositions we shall consider here are states of the form

$$|\psi_{\pm}\rangle = \cos(\theta)|\leftrightarrow\rangle \pm \sin(\theta)|\updownarrow\rangle, \quad (12.4)$$

where  $0 \leq \theta \leq \pi/4$ . For these states, in the case of equal a priori probabilities  $q_+ = q_- = 1/2$ , the Helstrom bound in Eq. (12.2) takes the simple form

$$P_e(\min) = \frac{1}{2} [1 - \sin(2\theta)]. \quad (12.5)$$

The von Neumann measurement which attains this bound has the associated orthogonal projection operators

$$P_{\pm} = |\omega_{\pm}\rangle\langle\omega_{\pm}|, \quad (12.6)$$

where the orthogonal states  $|\omega_{\pm}\rangle$  are defined by

$$|\omega_{\pm}\rangle = \frac{|\leftrightarrow\rangle \pm |\updownarrow\rangle}{\sqrt{2}}. \quad (12.7)$$

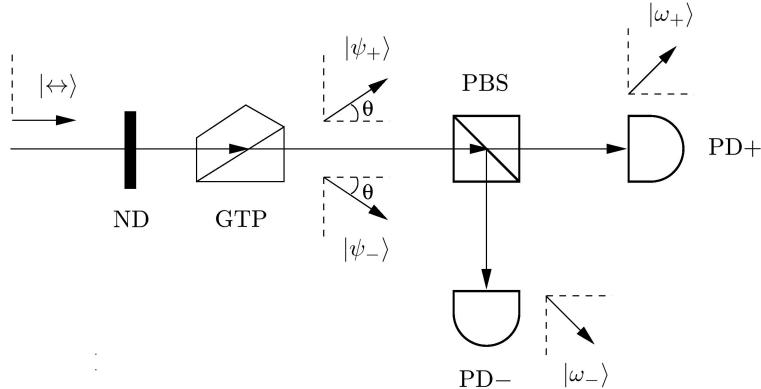
The states we have introduced in (12.4) and (12.7) are depicted in Fig. (12.1). We see that they have simple geometrical properties. The  $|\omega_{\pm}\rangle$  are as close

as they can be to the  $|\psi_{\pm}\rangle$  whilst maintaining orthogonality. The reflection symmetry about the  $|\leftrightarrow\rangle$ -axis is due to the equality of the a priori probabilities  $q_{\pm}$ . We can also see from the figure that errors are unavoidable, since  $|\omega_{\pm}\rangle$  is not orthogonal to  $|\psi_{\mp}\rangle$ .

Von Neumann measurements are discussed in most introductory quantum mechanics texts, e.g. [19]. The Helstrom bound can always be attained by such a measurement. However, as we shall see in subsequent sections, not all optimum measurements for state discrimination are of this form.

### 12.2.2 Discrimination at the Helstrom bound between optical polarisation states

Let us now describe an experiment performed by Barnett and Riis [7] which demonstrated state discrimination at the level of the Helstrom bound. The states prepared by these authors were of the form shown in (12.4) with equal a priori probabilities. The experimental arrangement that was used is shown

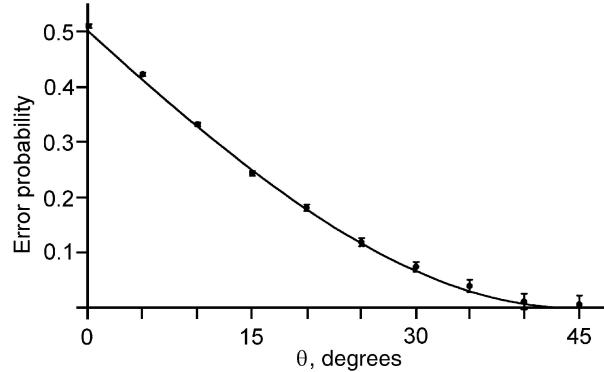


**Fig. 12.2.** Depiction of the apparatus used in the Barnett-Riis experiment. Optical pulses were prepared in a state of horizontal polarisation  $|\leftrightarrow\rangle$ . These were then heavily attenuated ( $\sim 0.1$  photons/pulse) by a neutral density filter (ND). The attenuated pulses then encountered a Glan-Thompson polariser. This is a polarisation filter, and its transmission axis was set to be in either the direction of  $|\psi_+\rangle$  or  $|\psi_-\rangle$ . Each pulse either emerged in the desired state, or it was absorbed. The output pulses then encountered a polarising beamsplitter PBS. This was oriented at  $\pi/4$  radians to the horizontal, so as to transmit the component in the direction  $|\omega_+\rangle$  to detector PD+, and reflect the component in the orthogonal direction  $|\omega_-\rangle$ , towards the detector PD-. If a detection at PD+ is taken to indicate that the initial state was  $|\psi_+\rangle$ , and likewise with PD- and  $|\psi_-\rangle$ , then this measurement gives the minimum probability of error.

in Fig. (12.2). A mode-locked Ti:Sapphire laser, operating at 780nm and with

a repetition rate of 80.3 MHz, generated pulses of light in the horizontally polarised state  $|\leftrightarrow\rangle$ . These were then heavily attenuated by a neutral density filter (ND) to the point where, on average, only 1 in 10 pulses contained a photon. This was done to make the probability of there being 2 or more photons per pulse negligible (1 in 200.)

In state discrimination experiments based on optical polarisation, it is highly desirable to have only one photon per pulse. The reason for this (and this point applies to all of the experiments to be described in this chapter) is that the measurements which we are considering, in general, optimally discriminate among a set of quantum states provided that there is only one copy of the state available. If more than one copy is available, then it is possible that more accurate discrimination will be possible using a collective measurement on all available copies of the state. This is a consequence of the remarkable properties of collective measurements on multiple quantum systems, and we shall return to this matter in section 12.7.



**Fig. 12.3.** Results of the Barnett-Riis experiment to discriminate between two optical polarisation states,  $|\psi_{\pm}\rangle = \cos(\theta)|\leftrightarrow\rangle \pm \sin(\theta)|\uparrow\rangle$ , with minimum error probability. The error probability is shown versus the angle  $\theta$ . Experimental data for specific values of  $\theta$  are shown in comparison with the continuous Helstrom limit in (12.5).

Returning to the Barnett-Riis experiment, following attenuation, the pulses propagated toward a Glan-Thompson polariser (GTP). This was randomly flipped between two settings which changed the initial horizontal polarisation state into either of the states  $|\psi_{\pm}\rangle$  in (12.4), with equal probability. The beam was then analysed at a polarising beam splitter (PBS) oriented at an angle of  $\pi/4$  to the horizontal. We should state here that, throughout this chapter, unless a specific orientation is stated, all polarising beamsplitters will transmit horizontal polarisation and reflect vertical.

To understand the action of this beamsplitter, we refer to the states  $|\omega_{\pm}\rangle$  in (12.4). A photon in the state  $|\omega_+\rangle$  would be transmitted by the beam splitter, while  $|\omega_-\rangle$  would be reflected. The transmitted and reflected states were fed to photodetectors PD+ and PD- respectively. Correct results were obtained when a photon prepared in the state  $|\psi_j\rangle$  was detected at  $D_j$ , where  $j = \pm$ . If the photon was detected at the other ‘wrong’ detector, an error ensued.

Figure (12.3) shows the experimental error probability for several values of  $\theta$ , compared with the theoretical minimum given by the Helstrom bound in (12.5). The agreement is clearly excellent.

### 12.3 Generalised measurements and POVMs

In the preceding section, we discussed a simple state discrimination experiment with two possible outcomes on a system prepared in one of two possible states. Whenever the outcome matched the corresponding state, the result of the state discrimination attempt was correct. Otherwise, it was incorrect.

As we saw, this measurement was a von Neumann measurement, of the kind commonly discussed in elementary quantum mechanics texts [19]. However, with a little thought, we can see that such measurements may be too restrictive for some more general problems in state discrimination. In a von Neumann measurement, the set of outcomes corresponds to a set of orthogonal subspaces of the system’s Hilbert space. If the dimensionality of this space is  $D$ , and  $N$  is the number of measurement outcomes, then it follows that we must have  $N \leq D$ .

In minimum error state discrimination, each measurement outcome corresponds to one of the possible states. So, if the number of states exceeds the Hilbert space dimension, then we have  $N > D$ . From the above argument, we see that this cannot be true for a von Neumann measurement and so we have a problem.

Fortunately, the quantum formalism does not limit us to such measurements. There exists a much more general class of measurements, known as generalised measurements [1], which do not present this difficulty. To understand these measurements, and why they can be regarded as generalisations of von Neumann measurements, it is helpful to have a full understanding of these latter, simpler measurements.

Consider a quantum system with Hilbert space  $\mathcal{H}$ . This space is, for the sake of simplicity, taken to have finite dimensionality  $D$ . The statistical properties of a von Neumann measurement are described in terms of a complete set of  $N$  orthogonal projection operators  $P_k$ , where  $k = 1, \dots, N$ . The completeness condition can be expressed as

$$\sum_{k=1}^N P_k = 1, \quad (12.8)$$

where ‘1’ is the identity operator on  $\mathcal{H}$ . It is helpful to recall here that projection operators are positive operators, that is, that their expectation values for any state are non-negative.

Each of these operators projects onto a different orthogonal subspace of  $\mathcal{H}$ . It clearly follows that  $N \leq D$ . Furthermore, each of these operators corresponds to a distinct outcome of the measurement. These are also indexed by  $k$ . The probability of obtaining outcome  $k$  when the initial state of the system is represented by the density operator  $\varrho$  is

$$p(k|\varrho) = \text{Tr}(P_k \varrho). \quad (12.9)$$

For a pure initial state  $\varrho = |\psi\rangle\langle\psi|$ , this gives  $p(k|\psi) = \langle\psi|P_k|\psi\rangle$ . (12.8) and (12.9) are the general conditions that specify the statistical properties of a von Neumann measurement.

In a generalised measurement, the statistical properties are described in terms of a set of positive operators which are less constrained than those which occur in a von Neumann measurement. These operators are known as positive, operator-valued measure (POVM) elements. If we have a generalised measurement with  $N$  possible outcomes, indexed by  $k = 1, \dots, N$ , then associated with the  $k$ th outcome is a POVM element  $\Pi_k$ . These operators satisfy the following condition similar to (12.8):

$$\sum_{k=1}^N \Pi_k = 1, \quad (12.10)$$

where again ‘1’ is the identity operator on  $\mathcal{H}$ .

The way in which the probabilities of the outcomes of a generalised measurement are calculated is also similar to that in the simpler, von Neumann case. When the initial state of the system is described by a density operator  $\varrho$ , the probability of obtaining outcome  $k$  for this initial state is

$$p(k|\varrho) = \text{Tr}(\Pi_k \varrho) \quad (12.11)$$

which clearly resembles (12.9). Again, for a pure initial state  $\varrho = |\psi\rangle\langle\psi|$ , we have  $p(k|\psi) = \langle\psi|\Pi_k|\psi\rangle$ . These are the only two conditions that the POVM elements  $\Pi_k$  must satisfy in order to represent a physically possible generalised measurement. In particular, these operators do not need to be orthogonal projectors, and so we do not require that  $N \leq D$ . In fact, there are no restrictions whatsoever on  $N$ , the number of possible outcomes of the measurement. In section 12.5, we shall see how this feature can be used for minimum error discrimination among multiple states of a qubit.

## 12.4 Unambiguous discrimination between two pure states

### 12.4.1 Background

Being able to perform measurements with an arbitrary number of outcomes, we might ask ourselves if there is anything to be gained if not every outcome corresponds to the detection one of the states. In other words, do we gain anything if we allow our measurement to have inconclusive results? This issue was first examined by Ivanovic [11], who made the startling discovery that the possibility of occasionally obtaining inconclusive results permits error free discrimination between non-orthogonal states. Ivanovic devised a measurement with the property that, whenever the result is not inconclusive, it is always correct.

To see how this may be done, consider again the two states  $|\psi_{\pm}\rangle$  defined in (12.4). Let us now introduce the additional states

$$|\tilde{\psi}_{\pm}\rangle = \sin(\theta)|\leftrightarrow\rangle \mp \cos(\theta)|\uparrow\downarrow\rangle. \quad (12.12)$$

Notice that  $|\tilde{\psi}_+\rangle$  is orthogonal to  $|\psi_-\rangle$ , and likewise with  $|\tilde{\psi}_-\rangle$  and  $|\psi_+\rangle$ . Consider now a generalised measurement described by the following POVM elements:

$$\Pi_{\pm} = \frac{p(\pm|\psi_{\pm})}{|\langle\tilde{\psi}_{\pm}|\psi_{\pm}\rangle|^2} |\tilde{\psi}_{\pm}\rangle\langle\tilde{\psi}_{\pm}|, \quad \Pi_{?} = 1 - \Pi_{+} - \Pi_{-}. \quad (12.13)$$

The meaning of the coefficients  $p(\pm|\psi_{\pm})$  will become apparent shortly. Since  $|\tilde{\psi}_+\rangle$  is orthogonal to  $|\psi_-\rangle$ , we see, from (12.11), that  $\langle\psi_-|\Pi_+|\psi_-\rangle = 0$ , and so the probability of obtaining the result ‘+’ for this state must be zero. Likewise,  $\langle\psi_+|\Pi_-|\psi_+\rangle = 0$ , so we will never obtain the result ‘-’ for the initial state  $|\psi_+\rangle$ . So, whenever we obtain one of these two results, we can determine exactly what the initial state was. We can easily see from (12.11) that

$$\langle\psi_{\pm}|\Pi_{\pm}|\psi_{\pm}\rangle = p(\pm|\psi_{\pm}). \quad (12.14)$$

From this we infer that  $p(\pm|\psi_{\pm})$  is the probability, given that the system was prepared in the state  $|\psi_{\pm}\rangle$ , that this state will be identified unambiguously.

Unless the states are orthogonal, these probabilities cannot attain the value of 1. There is a third result, the inconclusive result ‘?’ , the probability of which for each state is equal to the expectation value of the operator  $\Pi_{?}$ . As with the minimum error probability strategy we discussed in section 12.2, it is important to optimise this measurement. This means obtaining the maximum unambiguous discrimination probability, or equivalently, the minimum probability of inconclusive results. To obtain this, we must know the a priori probabilities  $q_{\pm}$  of the two states. Given these, the total probability  $P_{?}$  of obtaining an inconclusive result is

$$P_? = 1 - \sum_{j=+,-} q_j p(j|\psi_j). \quad (12.15)$$

The variational problem whose solution is  $P_?(\min)$  essentially consists of determining the values of  $p(\pm|\psi_\pm)$  which minimise (12.15) subject to the constraint that the operator  $\Pi_?$  is positive. For two states with equal a priori probabilities  $q_\pm = 1/2$ , it was established through the work of Ivanovic [11], Dieks [12] and Peres [13] that the minimum attainable value of the inconclusive result probability is given by

$$P_?(\min) = |\langle\psi_+|\psi_-\rangle|. \quad (12.16)$$

This Ivanovic-Dieks-Peres (IDP) bound is obtained when the  $p(\pm|\psi_\pm)$  are both equal to  $1 - |\langle\psi_+|\psi_-\rangle|$ , which is also the total probability of obtaining a conclusive, correct result. Subsequently, Jaeger and Shimony [20] obtained the minimum value of  $P_?$  for two pure states with arbitrary a priori probabilities.

#### 12.4.2 The experiment of Huttner et al

As with the minimum error state discrimination experiment described in section 12.2, optical polarisation states are well-suited to the experimental demonstration of unambiguous state discrimination. To date, two such experiments have been performed. We shall begin with the first, which was carried out by Huttner et al [9]. The setup for this experiment is depicted in Fig. (12.4). Weak pulses of laser light were generated at a wavelength of 1300nm and with duration of approximately 1ns. The pulses were strongly attenuated to the level of 0.1 photons/pulse on average. The light was subsequently transformed into a state of circular polarisation by a  $\lambda/4$  plate. The beam then encountered a rotating polariser, which could be used to produce any state of linear polarisation. The states  $|\psi_\pm\rangle$ <sup>1</sup> in (12.4) were prepared with equal a priori probability and numerous values of the angle  $\theta$ .

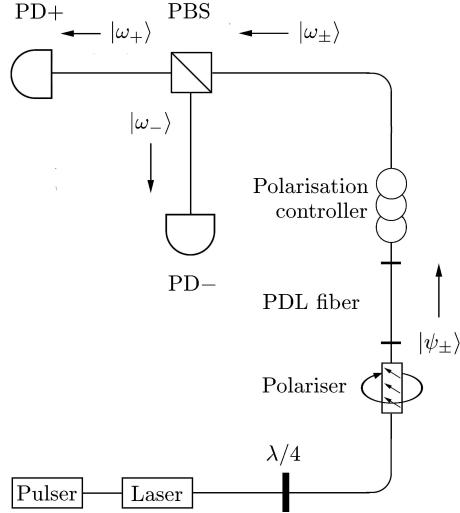
The pulses then travelled through a polarisation-dependent loss (PDL) fibre. This fibre left the amplitude of the initially vertical polarisation component unchanged but attenuated that of the initially horizontal component. The actual polarisations of these components were, however, changed by the fibre and we shall return to this matter shortly. For the horizontal polarisation, the transmission and absorption coefficients were

$$t = \tan(\theta), \quad a = \sqrt{1 - \tan^2(\theta)}. \quad (12.17)$$

If a photon was absorbed here, then the information it carried about the polarisation state was irrevocably lost. The probability of this occurring is

---

<sup>1</sup> Given two orthogonal spatial directions, the convention by which one is labelled horizontal and the other as vertical is, in this context, arbitrary. In our description of this experiment, we have exchanged these labels with respect to the original publication in order to homogenise the conventions used in this chapter.



**Fig. 12.4.** Experimental setup used to unambiguously discriminate between the non-orthogonal optical polarisation states  $|\psi_{\pm}\rangle$  in (12.4) at the IDP limit. The pulser, laser and  $\lambda/4$  plate produced states of circular polarisation. The rotating polariser was then used to transform these into one of the states  $|\psi_{\pm}\rangle$ , with equal a priori probabilities. The horizontal component was attenuated by the polarisation-dependent-loss (PDL) fibre. Whenever the states were transmitted, they were orthogonalised but elliptically polarised. The polarisation controller converted these into the orthogonal linear polarisation states  $|\omega_{\pm}\rangle$ . These were finally discriminated using a polarising beamsplitter, which sent the states  $|\omega_{+}\rangle$  and  $|\omega_{-}\rangle$  to photodetectors PD+ and PD- respectively.

simply the product of  $t^2$  and the probability that the photon is horizontally polarised. For the states  $|\psi_{\pm}\rangle$ , the latter factor is simply  $\cos^2\theta$  and so the absorption probability for both states is  $\cos(2\theta) = |\langle\psi_+|\psi_-\rangle|$ . Notice that this is equal to (12.16), the IDP bound on the probability of inconclusive results. Given that after absorption, no attempt to discriminate between the states is possible, this implies that absorption events do indeed correspond to inconclusive results. It also follows that if, whenever absorption does not take place, the two states can be perfectly discriminated, then such a measurement will be optimal.

To see that this is indeed true, we must understand what happens if the pulses are transmitted. When this happens, the coefficient of the horizontally polarised components of  $|\psi_{\pm}\rangle$  are multiplied by the transmission coefficient  $\tan(\theta)$ . The states are then renormalised. However, this fibre also had the effect of transforming linear polarisation states into elliptical states. To rectify this, a polarisation controller was inserted to convert the states back into linear polarisation states. The total transformation undergone by the states

was

$$|\psi_{\pm}\rangle \rightarrow |\omega_{\pm}\rangle, \quad (12.18)$$

where the states  $|\omega_{\pm}\rangle$  are those defined in (12.7). These states, being orthogonal, can be perfectly distinguished, using a polarising beam splitter as was done in the Barnett-Riis experiment. As can be seen in the figure, the pulses will propagate to polarising beam splitter PBS. Like in the Barnett-Riis experiment, this was oriented at  $\pi/4$  to the horizontal. It deflected the  $|\omega_+\rangle$  state into the path leading to detector PD+. The  $|\omega_+\rangle$  state was deflected toward detector PD-. So, whenever a photon was transmitted through the fibre, which, for non-empty pulses, occurred with probability  $\cos(2\theta)$ , the original non-orthogonal polarisation states were transformed into an orthogonal set which could, ideally, be perfectly distinguished at the detectors. The wrong path would not be taken, so that when successful, the discrimination attempt would always give the correct answer.

For  $\theta = \pi/6$ , an error rate of 1.7% was obtained. Comparing this with the minimum error probability in the Helstrom measurement, which is approximately 6.7%, we see a clear improvement.

One disadvantage of using a PDL fibre is the fact that photons lost did not go to a detector. They simply did not register anywhere. Thus, the occasions on which photons failed to result in a click at either PD+ or PD-, which we interpret to be inconclusive results, could not be distinguished from those null events due to the weakness of the pulse. However, this does not detract from the fact that the error probability obtained for the detected photons was significantly lower than the Helstrom bound.

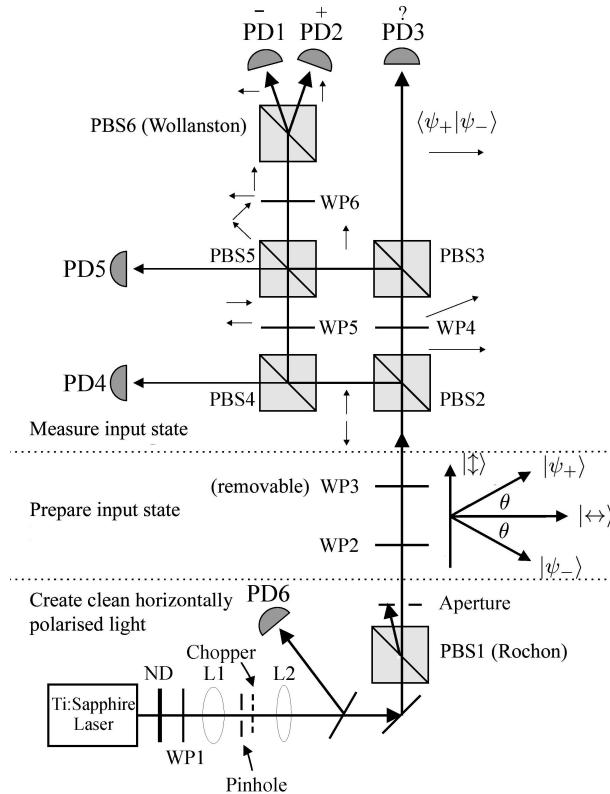
#### 12.4.3 The experiment of Clarke et al

Here, we shall describe an implementation of unambiguous polarisation discrimination at the IDP limit, performed by Clarke et al [10], which used free-space interferometry and overcame the limitations of the fibre-based implementation. The absorption in the interferometer was negligible implying that, up to small experimental errors, either a conclusive discrimination or an measurable inconclusive result was obtained, with the latter being in accordance with the IDP bound (12.16). Consequently, this experiment was the first full demonstration of the IDP measurement.

The states chosen to be discriminated were again the optical polarisation states  $|\psi_{\pm}\rangle$  in (12.4) where  $0 \leq \theta \leq 45^\circ$ . The apparatus used in this experiment is depicted in Fig. (12.5). This figure is vertically partitioned into three regions which, from the bottom up, illustrate the elements used to create clean, horizontally polarised light, transform this into one of the states  $|\psi_{\pm}\rangle$  and finally perform the state discrimination measurement. We shall describe these stages in turn.

The light source was a mode-locked Ti:Sapphire laser operating at 780 nm with a repetition rate of 80.3 MHz. The pulse duration was 1 ps, corresponding to a pulse length of 300  $\mu\text{m}$ . This ensured that there was only one

pulse in the optical system at any one time and that the length of the pulse was much shorter than the path length of the interferometer. The output was focused with lens L1 through a  $60\ \mu\text{m}$  pinhole to produce a clean spherical wavefront. The light was then passed through lens L2 to produce a shallow focus on the centres of the  $1\ \text{mm}^2$  photodiodes (Centronix, BPX65) beyond the interferometer. The photodiodes were arranged to be of equal optical path lengths away. Almost 100% of the light input into the interferometer reached the detectors.



**Fig. 12.5.** Experimental setup used to prepare, process and discriminate between the states  $|\psi_+\rangle$  and  $|\psi_-\rangle$ . Here, L = lens, ND = neutral density filter to attenuate the light, PD = photodiode, WP = half-wave plate and PBS = polarising beam splitter.

Neutral density filters were inserted to attenuate the light to an average of 0.2 photons per pulse (4 pW) at the interferometer input. Fine adjustment of the intensity was possible by rotating the half-wave plate WP1, placed before the Rochon polarising beamsplitter, PBS1. This type of beamsplitter

was chosen for its high extinction ratio, measured to be greater than 1 part in 5000. A Wollaston type beamsplitter, PBS6, was used in the analyser part of the experiment for the same reason.

Due to the very low light levels, phase sensitive detection, using a chopper wheel, was required to recover the analogue signals from the photodetectors PD1-5. The photodiodes had a nominal quantum efficiency of 83% and light levels equivalent to less than 0.01 photons per pulse were detectable.

To normalise out the amplitude noise of the Ti:Sapphire laser, a pick-off beam was measured on PD6 when any measurements of photodiodes PD1-5 were taken. Again, phase-sensitive detection was performed, but using a separate lock-in amplifier. Photodiodes 1-5 were calibrated relative to each other to better than 1% by changing the distribution of light around the interferometer with half-wave plates.

The input to the interferometer was obtained from the horizontally polarised straight through beam of PBS1. This polarisation state was then changed using two half-wave plates, WP2 and WP3 (which was removable). At this point, it is convenient to review the transformation of polarisation states effected by half-wave plates, as we shall frequently encounter these devices throughout this chapter. A half-wave plate oriented at an angle  $\alpha$  measured anticlockwise to the horizontal will transform the polarisation in the following way

$$|\leftrightarrow\rangle \rightarrow \cos(2\alpha)|\leftrightarrow\rangle + \sin(2\alpha)|\uparrow\rangle, \quad (12.19)$$

$$|\uparrow\rangle \rightarrow \sin(2\alpha)|\leftrightarrow\rangle - \cos(2\alpha)|\uparrow\rangle, \quad (12.20)$$

that is, it rotates the polarisation state anticlockwise by the angle  $2\alpha$ . In WP2,  $\alpha$  had the value of  $\theta/2$ , preparing the polarisation state  $|\psi_+\rangle$ . In WP3, the value of  $\alpha$  was  $\pi - \theta$ . When inserted, this half-wave plate had the effect of transforming  $|\psi_+\rangle$  into  $|\psi_-\rangle$ . In this way, the input states were easily exchanged.

The interferometer itself was constructed from four AR coated polarising beamsplitters PBS2-5 mounted on a machined monolithic aluminium block. The optical pathlength difference in the two arms was inferred to be less than 4  $\mu\text{m}$  over the 80 mm total pathlength from the extinction ratio obtained when used in a Mach-Zehnder operation. PBS5 was capable of being rotated around and translated along all axes with piezo-electric transducers. The AR coated  $\lambda/2$  half-wave plates used were measured to maintain the linearity of polarisation to 1 part in 2000.

The role of the interferometer in this experiment was similar to that of the PDL fibre in the Huttner experiment. This section of the apparatus is that which, apart from PDL5 and PDL6, lies vertically between half-wave plates WP3 and WP6 in Fig. (12.5). This purpose of the interferometer was to remove a fraction of the amplitude of horizontal polarisation of the states  $|\psi_{\pm}\rangle$ . This subtracted amplitude propagated to the inconclusive result detector PD3, while the reduction of this amplitude transformed the non-orthogonal

states  $|\psi_{\pm}\rangle$  into a pair of orthogonal states, which could then be perfectly discriminated.

To describe the action of the interferometer, it is convenient to consider separately the parts of the states  $|\psi_{\pm}\rangle$  which propagated clockwise and counterclockwise from PBS2 to PBS5. The counterclockwise component is the horizontal component (since PBS5 transmits only the state  $|\leftrightarrow\rangle$ ). This component immediately encountered the half-wave plate WP4. If this is oriented at an angle  $\alpha = \frac{1}{2}\cos^{-1}(\sqrt{1 - \tan^2(\theta)})$  above the horizontal then we see from (12.19) that it will effect the transformation

$$|\leftrightarrow\rangle \rightarrow \sqrt{1 - \tan^2(\theta)}|\leftrightarrow\rangle + \tan(\theta)|\uparrow\rangle. \quad (12.21)$$

In practice, the orientation of WP4 was varied and we shall discuss this shortly. Subsequently, the light travelling along this arm encountered a further polarising beamsplitter PBS3. Here the horizontal and vertical polarisation states in the r.h.s. of (12.21) were separated. The horizontal component continued upward to photodetector PD3. The probability of a photon in the state shown in the r.h.s. of (12.21) propagating toward PD3 is simply the square of the probability amplitude for transmission at PBS3, i.e.  $1 - \tan^2(\theta)$ . To obtain the probability of detection at PD3 for one of the states  $|\psi_{\pm}\rangle$ , we simply multiply this factor by the probability for each of these states that the photon is transmitted by PBS2, which is  $\cos^2(\theta)$ . The result is  $\cos(2\theta) = \langle\psi_+|\psi_-\rangle$ , which is precisely the IDP limit in (12.16). A detection at PD3 is an inconclusive result.

If an inconclusive result did not occur, then the vertical component of the state on the r.h.s. of (12.21) propagated toward PBS5 with amplitude  $\tan(\theta)$ , where it was reflected upwards toward WP6. Under these circumstances, the horizontal component of  $|\psi_{\pm}\rangle$  was converted to a state of vertical polarisation  $|\updownarrow\rangle$ . Its amplitude was  $\tan(\theta)\cos(\theta) = \sin(\theta)$ .

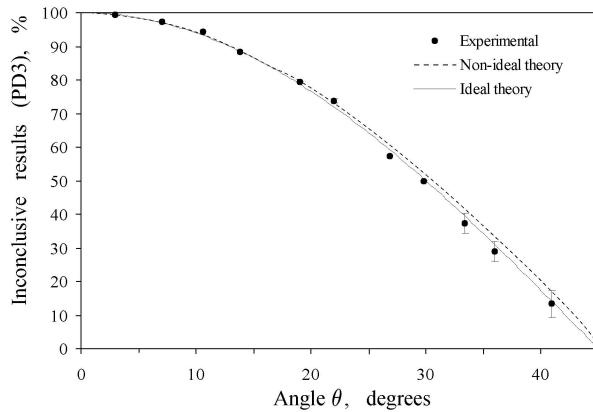
Meanwhile, the vertical component of the state  $|\psi_{\pm}\rangle$  propagated clockwise though the interferometer, being reflected by both PBS2 and PBS4. It then encountered half-wave plate WP5. This half-wave plate transformed  $|\updownarrow\rangle$  into  $|\leftrightarrow\rangle$ . Thus, the vertically polarised component of  $|\psi_{\pm}\rangle$ ,  $\pm\sin(\theta)|\updownarrow\rangle$ , travelled clockwise through the interferometer and was transmitted by PBS5, propagating towards WP6 as the state  $\pm\sin(\theta)|\leftrightarrow\rangle$ . This was superposed with the counterclockwise component  $\sin(\theta)|\updownarrow\rangle$ . Thus, when an inconclusive result is not obtained, the interferometer transforms the states  $|\psi_{\pm}\rangle$ , up to a normalisation factor, in the following way:

$$|\psi_{\pm}\rangle \rightarrow -|\omega_{\pm}\rangle, \quad (12.22)$$

where  $|\omega_{\pm}\rangle$  are given by (12.7) and the minus sign is irrelevant and may be neglected. These states are, of course, orthogonal, and as in the Huttner experiment can be perfectly discriminated. Again, this was accomplished with a polarising beamsplitter, here WP6. The normalisation factor is  $\sqrt{2}\sin(\theta)$

which, when squared, gives the probability for this transformation to occur. The square is simply  $1 - \cos(2\theta) = 1 - \langle \psi_+ | \psi_- \rangle$ , which is the maximum probability of unambiguous discrimination according to the IDP limit in (12.16).

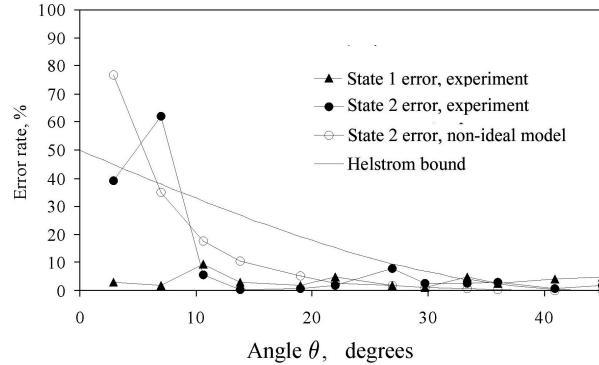
The experiment was performed by preparing the input state  $|\psi_+\rangle$  in approximately 4 degrees steps from  $0 \leq \theta \leq 45$  degrees. The exact angle was inferred by measuring the light in each arm immediately after PBS2 using all of the detectors. High input powers were used to obtain a good signal to noise ratio. The light transmitted through PBS3 was then varied by rotating WP4 such that the signal from PD2 was at a minimum. This occurred at the theoretically expected angle  $\alpha = \frac{1}{2}\cos^{-1}(\sqrt{1 - \tan^2(\theta)})$ . The measured signal at PD2 is defined as the measured error rate since in theory 100% of the light should reach PD1. In practice, it was non-zero due to experimental imperfections.



**Fig. 12.6.** The dark points show the experimental signal detected on PD3, corresponding to inconclusive results, required to orthogonalise the input states. The errors are derived from the maximum possible deviation required to observe a significant increase in the discrimination error rate. The continuous curve shows the ideal theoretical values for the inconclusive loss rates,  $\cos 2\theta$ . A model using the characteristics of the non-ideal beam splitters was used to generate the dashed curve. The estimated error of this line is approximately twice the experimental error, and was derived by optimising the extinguishment of the signal on PD2 to the 0.1% level.

Figure (12.6) shows the rate of inconclusive results, at the orientation of WP4 which minimises error, against  $\theta$ , for  $0 \leq \theta \leq 45$  degrees. The RMS deviation from the ideal theoretical curve is 1.3%, clearly validating the Ivanovic-Dieks-Peres measurement scheme. The error in PD3 was estimated by rotating WP4 until the signal on PD1 increased noticeably. For angles less than 30 degrees the sensitivity became so great that this error could not be estimated quantitatively and the uncertainty is less than the size of the points

in the figure. A model based upon the measured characteristics of the PBS's, described in [10], reproduces the experimental results within the estimated error values.



**Fig. 12.7.** The experimentally observed error rates (PD2 and PD1) obtained with input states  $|\psi_+\rangle$  and  $|\psi_-\rangle$  respectively. Also shown is the theoretical model using non-ideal beamsplitters with the same characteristics as in the experiment, and the Helstrom bound (12.5).

The error rates obtained with incident angle are plotted in Fig. (12.7). The states  $|\psi_+\rangle$  and  $|\psi_-\rangle$  are shown at the same angle position. Also plotted is error rate (12.5) associated with the best possible von Neumann measurement. The data clearly shows error probabilities that are below this level. For  $14^\circ < \theta < 45^\circ$  the average experimental error rate for the two input states is 2.8%. For smaller angles the error rate rises significantly. The behaviour of the interferometer using non-ideal PBS's was also modelled, based on the calibration data obtained previously. The experimental procedure was followed, optimising the loss for state  $|\psi_+\rangle$  and then flipping to state  $|\psi_-\rangle$  to obtain the error rates. These results are also shown in Fig. (12.7) and are in good qualitative agreement with the experimental results. For small  $\theta$ , the leakages of the PBS's were such that the errors present are of comparable size to the ideal signals.

## 12.5 Discrimination among multiple states with minimum error probability and the accessible information

### 12.5.1 Background

#### 5.1.1 Minimum error probability

In this section, we shall describe experimental implementations of the measurements which discriminate among certain sets of more than two states with minimum probability of error [15]. To do so, it is convenient to describe the problem of minimum error discrimination among multiple states in full generality. Consider a set of  $N$  possible states with density operators  $\varrho_j$ . Let these be prepared with a priori probabilities  $q_j$ . From this point onwards, we shall often consider a set of quantum states with specific a priori probabilities. This is an ensemble of quantum states.

We consider a measurement with  $N$  outcomes. The  $j$ th outcome, which has corresponding POVM element  $\Pi_j$ , is associated with the  $j$ th state,  $\varrho_j$ . If we obtain the result  $j$ , then we take this to signify that the initial state was  $\varrho_j$ . If this is true, then our result is correct, if not, then we have an error. From (12.9) we see that the error probability is

$$P_e = 1 - \sum_{k=1}^N q_k \text{Tr}(\varrho_k \Pi_j), \quad (12.23)$$

where the sum gives the probability that the state will be correctly identified. The necessary and sufficient conditions on the POVM elements for this probability to attain its minimum value were discovered independently by Holevo [21] and Yuen, Kennedy and Lax [22] and these are

$$\Pi_k (q_k \varrho_k - q_j \varrho_j) \Pi_j = 0 \quad \forall j, k, \quad (12.24)$$

$$\left( \sum_{k=1}^N q_k \varrho_k \Pi_k \right) - q_j \varrho_j \geq 0 \quad \forall j. \quad (12.25)$$

The general minimum error state discrimination problem is as follows: given  $N$  states  $\varrho_j$  with associated a priori probabilities  $q_j$ , find a set of  $N$  positive operators  $\Pi_k$  satisfying (12.24), (12.25) and also (12.10). The general solution to this problem is unknown. However, the solution is known for some special cases. The sets of states we shall be concerned with in this section are overcomplete sets. These are sets of pure states which satisfy the following condition:

$$\frac{D}{N} \sum_{j=1}^N |\psi_j\rangle\langle\psi_j| = 1, \quad (12.26)$$

where  $D$  is the dimensionality of the Hilbert space  $\mathcal{H}$  and ‘1’ is the identity operator on  $\mathcal{H}$ . For ensembles of such states with equal a priori probabilities, the POVM elements that minimise the error probability are [21, 22]

$$\Pi_j = \frac{D}{N} |\psi_j\rangle\langle\psi_j|. \quad (12.27)$$

This gives for the minimum error probability the value

$$P_e(\min) = 1 - \frac{D}{N}. \quad (12.28)$$

This measurement is actually an example of a very important type of measurement frequently encountered in quantum state discrimination theory, known as the square-root measurement [23]. It can be constructed for an arbitrary set of  $N$  pure states  $|\psi_j\rangle$  in the following way. Define the operator

$$\Phi = \sum_{k=1}^N |\psi_k\rangle\langle\psi_k|. \quad (12.29)$$

The square-root measurement then has the POVM elements  $\Pi_j = |\omega_j\rangle\langle\omega_j|$  where the (typically unnormalised) states  $|\omega_j\rangle$  are defined by

$$|\omega_j\rangle = \Phi^{-1/2} |\psi_j\rangle. \quad (12.30)$$

We shall encounter a further application of this measurement in section 12.7.

In this section, we shall consider four specific overcomplete sets. These are known as the trine, antitrine, tetrad and antitetrad sets. In all cases, the ensembles will have equal a priori probabilities, and we shall simply refer to these as the trine/antitrine/tetrad/antitetrad ensembles.

### 5.1.2 Accessible information

The minimum error probability is used as a measure of how precisely one can identify the signal states on average. Instead of identifying each individual state as precisely as possible, one might prefer to extract as much classical information as possible from a sequence of quantum states. A quantity which expresses how well this can be done is the Shannon mutual information [24].

Let us suppose that one party, Alice, prepares the ensemble of quantum states  $\{|\psi_j^N\rangle, q_j\}$ . A classical message  $x_j$  is encoded in the quantum state  $|\psi_j\rangle$ . The a priori probability of this state is  $q_j$ . Let us call the ensemble of possible classical messages ensemble  $X = \{x_j, q_j\}$ . The Shannon entropy of  $X$  is

$$H(X) = - \sum_{j=1}^N q_j \log q_j, \quad (12.31)$$

where, throughout this chapter, all logarithms have base 2.  $H(X)$  quantifies how much we know about  $X$  just from the a priori probabilities  $q_j$ . Actually, the minimum value of  $H(X)$  is zero, corresponding to knowing everything about  $X$ , and the maximum value of  $H(X)$  corresponds to knowing nothing better than a random guess for each element of  $X$ , so it is more natural to say that  $H(X)$  measures what we don't know about  $X$ . She transmits the states to her colleague, Bob, who applies a measurement with  $M$  possible outcomes  $\{y_k\}$ , characterised by the POVM elements  $\{\Pi_k\}$ . The outcome, say,  $y_k$ , gives Bob more information about  $X$ . The new probability distribution conditioned by  $y_k$  is

$$P(\psi_j^N | y_k) = \frac{P(y_k | \psi_j^N) q_j}{P(y_k)}, \quad (12.32)$$

where  $P(y_k) \equiv \sum_j P(y_k | \psi_j^N) q_j$  is the probability of having  $y_k$ , and  $P(y_k | \psi_j^N)$  is given by (12.11). One can then define the average conditional entropy as

$$H(X|Y) = - \sum_{k=1}^M P(y_k) \sum_{j=1}^N P(\psi_j^N | y_k) \log [P(\psi_j^N | y_k)]. \quad (12.33)$$

This quantifies the remaining uncertainty about  $X$  after having the knowledge of the conditioning variable  $Y = \{y_k, P(y_k)\}$ . The information gained as a result of making the measurement  $Y$  is naturally defined by

$$\begin{aligned} I(X : Y) &= H(X) - H(X|Y), \\ &= \sum_{j=1}^M \sum_{k=1}^N \left[ q_j P(y_k | \psi_j^N) \log \left( \frac{P(y_k | \psi_j^N)}{\sum_{k'=1}^N q_{k'} P(y_k | \psi_{k'}^N)} \right) \right]. \end{aligned} \quad (12.34)$$

This is the Shannon mutual information between  $X$  and  $Y$ . In general, minimizing the average error probability and maximizing the mutual information are different problems. In fact, one often finds the optimal POVMs are different in each case. To extract as much information as possible, Bob has to maximise the mutual information with respect to the POVM. The maximum value

$$I_{\text{Acc}} = \max_{\{\Pi_k\}} I(X : Y), \quad (12.35)$$

is called the accessible information of the ensemble  $X$ . To obtain the accessible information, it is not necessary for the number of measurement outcomes to be equal to the number of signal states. As it happens, in the experiments described in this section, they are equal. However, this will not be the case for some ensembles we shall consider in the sequel.

For future reference (section 12.7), we shall note that the accessible information can be further maximised with respect to the a priori probabilities  $q_j$ . The resulting quantity is known as the first-order classical capacity

$$C_1 = \max_{\{P(x)\}} \max_{\{\Pi_y\}} I(X : Y). \quad (12.36)$$

This quantity is the maximum rate at which one can send classical information, with arbitrarily low error probability, using a set of pure quantum states  $|\psi_j\rangle$  for any a priori probabilities and any measurement at the receiving end which acts on the signal states individually. In practice, to make the error probability negligible, one must use redundant channel coding. It is this quantity that limits the performance of all modern communications systems. However, this is not the ultimate classical capacity allowed by quantum mechanics, as we shall see in section 12.7.

In this section, we shall describe experiments, performed by Clarke et al [15], on certain overcomplete ensembles which minimise the error probability and attain the accessible information. We shall find that the measurements required to achieve these limits are closely related to each other, and indeed, also to the measurement described in the preceding section for optimal unambiguous discrimination between two states.

### 12.5.2 The trine/antitrine experiment

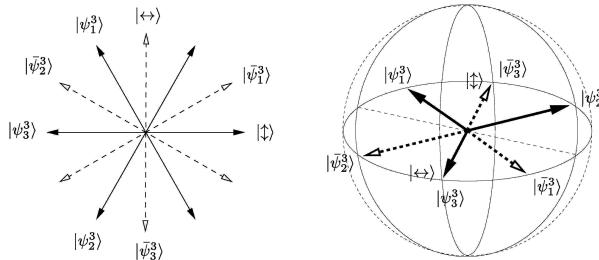
The optical polarisation states making up the trine ensemble are

$$|\psi_1^3\rangle = -\frac{1}{2} (| \leftrightarrow \rangle + \sqrt{3} | \downarrow \rangle), \quad (12.37)$$

$$|\psi_2^3\rangle = -\frac{1}{2} (| \leftrightarrow \rangle - \sqrt{3} | \downarrow \rangle), \quad (12.38)$$

$$|\psi_3^3\rangle = | \leftrightarrow \rangle. \quad (12.39)$$

These states have equal a priori probability of 1/3. They lie in the equatorial plane of the Poincaré sphere separated by 60°, as is shown in Fig. (12.8). The superscript label denotes the fact that there are three members of the trine ensemble.



**Fig. 12.8.** The trine states shown as polarisation directions (left) and on the Poincaré sphere (right). The antitrine (orthogonal) states are also shown (dashed).

Closely related to these are the states in the antitrine ensemble. The antitrine ensemble consists of the set of states where each is orthogonal to

one of the trine states. We shall write these states as  $|\bar{\psi}_j^3\rangle$ , where  $j = 1, 2, 3$ . They are

$$|\bar{\psi}_1^3\rangle = \frac{1}{2} (\sqrt{3}|\leftrightarrow\rangle - |\uparrow\downarrow\rangle), \quad (12.40)$$

$$|\bar{\psi}_2^3\rangle = -\frac{1}{2} (\sqrt{3}|\leftrightarrow\rangle + |\uparrow\downarrow\rangle), \quad (12.41)$$

$$|\bar{\psi}_3^3\rangle = |\uparrow\downarrow\rangle. \quad (12.42)$$

It is a simple matter to verify that  $\langle\psi_j^3|\bar{\psi}_j^3\rangle = 0$ . Both the trine and antitrine ensembles are overcomplete. Their minimum error probabilities are then given by (12.28). These are states of a qubit and so  $D = 2$ . Since there are 3 states in each set, we have  $N = 3$ . Substituting these values into (12.28), we find that the minimum error probabilities for these sets are

$$P_e^3(\min) = \bar{P}_e^3(\min) = \frac{1}{3}, \quad (12.43)$$

where the overbar distinguishes the trine and antitrines.

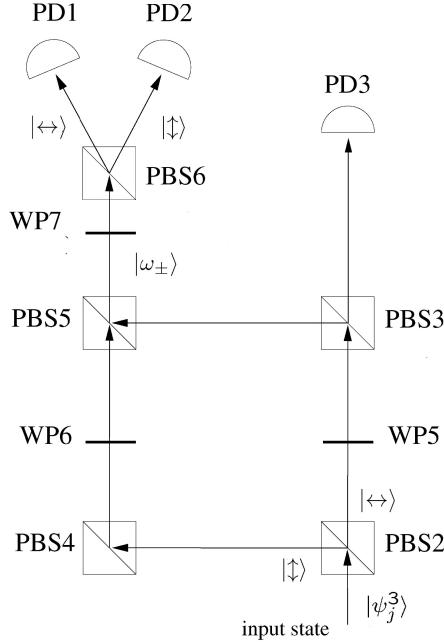
Figure (12.9) depicts the experimental apparatus used to carry out minimum error discrimination for the trine ensemble. The states entered the apparatus at the bottom of the figure and propagated through an interferometer to three photodetectors. Counts at photodetectors PD1, PD2 and PD3 corresponded to detection of the states  $|\psi_1^3\rangle$ ,  $|\psi_2^3\rangle$  and  $|\psi_3^3\rangle$  respectively. When one of these states resulted in a click at its corresponding photodetector, then this gave rise to a correct identification. If, on the other hand, one of the other photodetectors clicked, then the result was erroneous.

Rather than give a detailed explanation of how the three states propagate through the apparatus, we shall exploit a connection between this experiment and the experiment by Clarke et al on unambiguous discrimination between two states we discussed in section 12.4. Comparing Figs. (12.9) and (12.5), we see that they are highly similar. In fact, the only differences are that we have omitted common elements relating to state preparation, and relabellings. In particular, in the unambiguous discrimination experiment, the rightmost detector corresponded to inconclusive results while in the trine experiment it corresponded to correct detection of the state  $|\psi_3^3\rangle$ . We shall see why below.

We shall see here that the measurement which optimally unambiguously discriminates between the states  $|\psi_{\pm}\rangle$  for a particular value of the angle  $\theta$  is the trine measurement! To demonstrate this, we note, using (12.12), that if  $\theta = \pi/6$ , then the states  $|\tilde{\psi}_{\pm}\rangle$  are

$$|\tilde{\psi}_{\pm}\rangle = \frac{1}{2} (|\leftrightarrow\rangle \mp \sqrt{3}|\uparrow\downarrow\rangle). \quad (12.44)$$

Comparison of this with (12.37) and (12.38) shows that, up to irrelevant phases,  $|\tilde{\psi}_+\rangle = |\psi_2^3\rangle$  and  $|\tilde{\psi}_-\rangle = |\psi_1^3\rangle$ . Furthermore, for these states, we can calculate the factors  $p(\pm|\psi_{\pm}\rangle)/|\langle\psi_{\pm}|\psi_{\pm}\rangle|^2$  which occur in (12.13) and find that they are both equal to  $2/3$ .



**Fig. 12.9.** Optical network for the trine generalised measurements. PBS = polarising beam splitter, WP = half-wave plate and PD = photodetector.

To state this connection formally, we see from (12.27) that the POVM elements for the trine measurement are

$$\Pi_j^3 = \frac{2}{3} |\psi_j^3\rangle\langle\psi_j^3|. \quad (12.45)$$

In view of the the above considerations, if we make use of (12.12) we find, for  $\theta = \pi/6$ , that  $\Pi_1^3$  and  $\Pi_2^3$  are equal to  $\Pi_-$  and  $\Pi_+$  respectively. Furthermore, it necessarily follows that  $\Pi_3^3$  is equal to the inconclusive result POVM element  $\Pi_?$  in (12.13). This is a consequence of the resolution of the identity in (12.10) and the overcompleteness of the trine states in (12.26). This, in particular, explains the identification of the same detector with inconclusive results in the IDP measurement and detection of the state  $|\psi_3^3\rangle$  in the trine measurement. To summarise, this argument shows that the trine measurement is a special case of the IDP measurement, specifically when the angle  $\theta$  in the IDP measurement has the value  $\pi/6$ . This explains why the measurement apparata are for both experiments are identical.

So, performing the trine experiment amounts to nothing more than preparing and feeding the trine states into the main interferometer of the IDP measurement for the states given by (12.4). Theoretically, the probabilities of obtaining clicks at the three photodetectors in Fig. (12.9) with initial

state  $|\psi_j^3\rangle$  is

$$p(k|\psi_j^3) = \langle\psi_j^3|\bar{\Pi}_k^3|\psi_j^3\rangle = \begin{cases} 2/3 : j = k \\ 1/6 : j \neq k \end{cases} \quad (12.46)$$

This set of probabilities is sometimes known as the channel matrix.

For the trine ensemble the accessible information is attained for a measurement with POVM elements based on the antitrine states,

$$\bar{\Pi}_j^3 = \frac{2}{3}|\bar{\psi}_j^3\rangle\langle\bar{\psi}_j^3|. \quad (12.47)$$

The optimality of this measurement strategy was conjectured in [18, 21] and finally proven by Sasaki et al [25]. The corresponding accessible information is  $\log \frac{3}{2} = 0.585$  bits. The channel matrix for the measurement which attains the accessible information has the elements

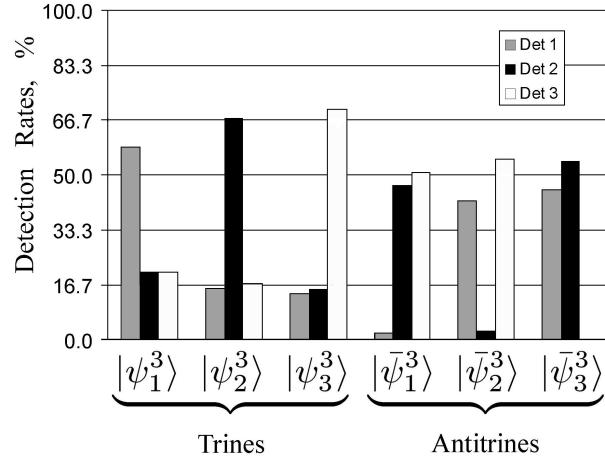
$$p(k|\psi_j^3) = \langle\psi_j^3|\bar{\Pi}_k^3|\psi_j^3\rangle = \begin{cases} 0 : j = k \\ 1/2 : j \neq k \end{cases} \quad (12.48)$$

The maximum mutual information that can be attained with a von Neumann measurement is  $(-\frac{1}{3} + \frac{1}{2}\log 3) = 0.459$  bits which is clearly less than the value attainable using the antitrine measurement.

In order to implement the POVM corresponding to the maximum mutual information, instead of constructing the optical network that corresponds to POVM elements  $|\bar{\psi}_j^3\rangle\langle\bar{\psi}_j^3|$ , and applying it to the trine states, the optical network used in the minimum error discrimination for the trine ensemble was applied to the antitrine ensemble. This measurement, called the antitrine measurement, is obviously completely equivalent theoretically and more practical experimentally. We find that the state  $|\bar{\psi}_j^3\rangle$  should never trigger photodetector  $j$ . It will, however, lead to detection in either of the remaining two photodetectors with equal probability of  $\frac{1}{2}$ .

Before discussing the results it is useful to review the sources of error. The largest is the noise of the detector signal and from the phase sensitive detection processing. The observed uncertainty in the measurements of the output ports where a zero count rate is expected in the antitrine experiment is approximately 2.5%. This level of error is often similar to the average signal present. We therefore conclude that the zero count measurements are limited by the detector noise if the measured signal is of order 2%. This has important implications for the errors in the derived mutual information. The noise is mainly attributable to the weakness of these signals in comparison to the magnitude of ground loop and stray light noise, and also variations in the chopping wheel frequency leading to varying offsets. The error due to the amplitude normalisation procedure involving PD6 was estimated to be less than 0.5%, which is much smaller than the Ti:Sapphire amplitude variations of up to 4% over a few minutes.

A second large source of error is the non-ideal nature of the beamsplitters. The calibrated birefringence properties are given in. It is sufficient here to



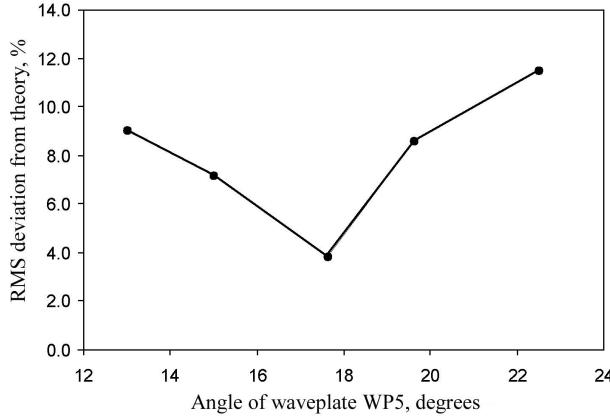
**Fig. 12.10.** Histogram showing the trine and antitrine results. Normalisation ensures the sum of counts in detectors 1, 2 and 3 is 100% for each input state. The theoretical ratios are  $\frac{2}{3} : \frac{1}{6} : \frac{1}{6}$  for the trine and  $\frac{1}{2} : \frac{1}{2} : 0$  for the antitrine states.

state that when purely horizontal light is input into PBS3, a power leakage of approximately 0.9% towards PBS4 was measured. Even this small amount of reflected light can have large effects on the ratio of light reaching detectors PD1 and PD2 due to interference effects.

The drift of the interferometer during measurements was negligible compared to the above errors. This was evaluated by monitoring the level of destructive interference that could be observed on detectors PD1 and PD2 over several hours using high light intensities.

The results in Fig. (12.10) demonstrate that the trine and antitrine measurements are in close agreement with theoretical predictions. The ratios  $\frac{2}{3} : \frac{1}{6} : \frac{1}{6}$  and  $\frac{1}{2} : \frac{1}{2} : 0$  are clearly visible for the respective measurements. The RMS deviation of the trine and antitrine channel matrix elements was 3.8% from the theoretically expected values. For the antitrine experiment, the measured value of the mutual information was 0.491 bits with a lower limit, taking into account experimental imperfections, of 0.464 bits. While this does fall short of the theoretical accessible information 0.415 bits, it is a clear improvement on the theoretical maximum that can be obtained with a von Neumann measurement, which is 0.59 bits. Figure (12.11) also shows the effect on the RMS error of rotating wave plate WP5 around its theoretical value.

Close inspection of Fig. (12.10) demonstrates the effect of the leakage of PBS3. For antitrine input state  $|\psi_3^3\rangle$  the light on PD1 and PD2 is split into the ratio 46:54. In an ideal experiment there would be no light travelling towards PBS4 from PBS2 and the split would be 50:50. Indeed, this ratio is



**Fig. 12.11.** The angle of WP5 was varied around the theoretical value of  $\theta/2 = 17.63^\circ$ . The graph shows the RMS deviation of the results from the optimum theoretical results.

observed to better than  $\pm 0.2\%$  if an opaque card is placed between PBS2 and PBS4. The leakage from PBS2, when interfering with the light from PBS3, is enough to skew the results by 8%. This demonstrates the sensitivity of the apparatus to practical sources of error.

### 12.5.3 The tetrad/antitetrad experiment

The states making up the tetrad set are

$$|\psi_1^4\rangle = \frac{1}{\sqrt{3}} \left( -| \leftrightarrow \rangle + \sqrt{2} e^{-2\pi i/3} | \uparrow \downarrow \rangle \right), \quad (12.49)$$

$$|\psi_2^4\rangle = \frac{1}{\sqrt{3}} \left( -| \leftrightarrow \rangle + \sqrt{2} e^{+2\pi i/3} | \uparrow \downarrow \rangle \right), \quad (12.50)$$

$$|\psi_3^4\rangle = \frac{1}{\sqrt{3}} \left( -| \leftrightarrow \rangle + \sqrt{2} | \uparrow \downarrow \rangle \right), \quad (12.51)$$

$$|\psi_4^4\rangle = | \leftrightarrow \rangle, \quad (12.52)$$

where the superscript label again denotes the number of states. Again, all states have equal a priori probability, in this case  $1/4$ . The first two states correspond to elliptical polarisations and the second two represent linear polarisations.

By analogy with the antitrine set, we can consider the antitetrad set. This is a set of four states, each of which is orthogonal to one of the tetrad states. Denoting these states by  $|\bar{\psi}_j^4\rangle$ , where  $j = 1, 2, 3, 4$ , we find that, up to physically unimportant phase factors, these states are

$$|\bar{\psi}_1^4\rangle = -\frac{1}{\sqrt{3}} \left( \sqrt{2}e^{2\pi i/3} |\leftrightarrow\rangle + |\uparrow\downarrow\rangle \right), \quad (12.53)$$

$$|\bar{\psi}_2^4\rangle = -\frac{1}{\sqrt{3}} \left( \sqrt{2}e^{-2\pi i/3} |\leftrightarrow\rangle + |\uparrow\downarrow\rangle \right), \quad (12.54)$$

$$|\bar{\psi}_3^4\rangle = -\frac{1}{\sqrt{3}} \left( \sqrt{2} |\leftrightarrow\rangle + |\uparrow\downarrow\rangle \right), \quad (12.55)$$

$$|\bar{\psi}_4^4\rangle = |\uparrow\downarrow\rangle. \quad (12.56)$$

One can easily verify that  $\langle \psi_j^4 | \bar{\psi}_j^4 \rangle = 0$ . The tetrad set is overcomplete. The measurement which discriminates among the states in the tetrad ensemble with minimum error probability has, from (12.27), the POVM elements

$$\Pi_j^4 = \frac{1}{2} |\psi_j^4\rangle \langle \psi_j^4|. \quad (12.57)$$

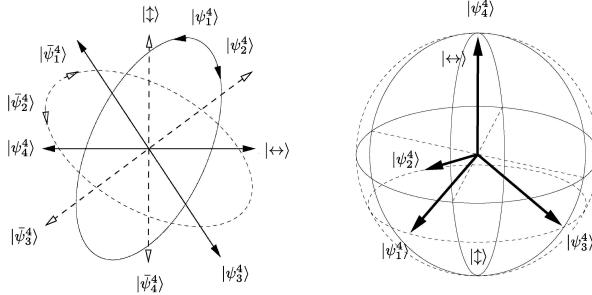
From this, we can calculate the elements of the channel matrix. These are

$$p(k|\psi_j^4) = \langle \psi_j^4 | \Pi_k^4 | \psi_j^4 \rangle = \begin{cases} 1/2 : j = k \\ 1/6 : j \neq k \end{cases} \quad (12.58)$$

From these we obtain the minimum error probability for this ensemble

$$P_e^4(\min) = \frac{1}{2}, \quad (12.59)$$

in accordance with (12.28). The minimum error probability for the antitetrad ensemble is the same.



**Fig. 12.12.** The tetrad states shown as polarisation directions (left) and on the Poincaré sphere (right). Antitetrad (orthogonal) states are also shown (dashed) in the left diagram.

The accessible information for the tetrad ensemble is not as strongly established as it is for the trine ensemble. Davies [26] conjectured that it can be reached by a measurement with the POVM elements

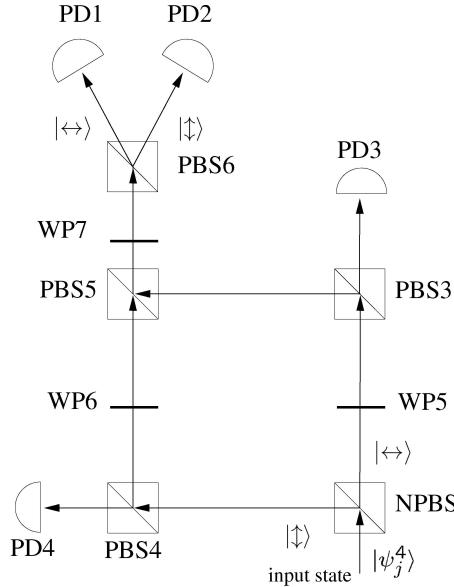
$$\bar{\Pi}_j = \frac{1}{2} |\bar{\psi}_j^4\rangle\langle\bar{\psi}_j^4|. \quad (12.60)$$

The channel matrix for this measurement has the elements

$$p(k|\psi_j^4) = \langle\psi_j^4|\bar{\Pi}_k^4|\psi_j^4\rangle = \begin{cases} 0 & : j = k \\ 1/3 & : j \neq k \end{cases} \quad (12.61)$$

The mutual information associated with this measurement strategy (conjectured to be the accessible information) is  $\log \frac{4}{3} = 0.415$  bits. This again exceeds the value of  $\frac{3}{2}(1 - \frac{1}{2}\log 3) = 0.311$  bits, which is the highest value that can be attained with a von Neumann measurement.

The optical network used to implement the optimal measurements for the tetrad ensemble is shown in Fig. (12.13). It is similar to the trine network in Fig. (12.9), with an extra detector PD4. The key differences are a non-polarising beam splitter (NPBS) in place of PBS2 and the difference in the operation of half-wave plate WP5.



**Fig. 12.13.** Optical network for the tetrad generalised measurement. NPBS = non-polarising beam splitter, PBS = polarising beam splitter, WP = wave plate and PD = photodetector.

For a full description of how the network in the tetrad experiment performs the required measurement, see [15]. We note here the effect of the half-wave plate WP5 on each of the four tetrad states. This half-wave plate is oriented at an angle of  $\alpha = \frac{\pi}{2} + \sin^{-1}(\frac{1}{\sqrt{3}})$  radians above the horizontal. Apart

from phase factors, state  $|\psi_1^4\rangle$  is converted into  $|\psi_2^4\rangle$  and vice versa, similarly for states  $|\psi_3^4\rangle$  and  $|\psi_4^4\rangle$ . Remembering the first beam splitter is 50/50 non-polarising, the network is thus essentially symmetrical in its operation. The full details are given in [15]. The final result is that state  $|\psi_1^4\rangle$  reaches photodetector PD1 with probability  $\frac{1}{2}$  and each of the remaining photodetectors with probability  $\frac{1}{6}$ . Similarly each of the remaining three states will trigger its associated photodetector with probability  $\frac{1}{2}$  and the others with probability  $\frac{1}{6}$ .

The maximum mutual information measurement was realised in the same way as for the trine ensemble, by using the antitetrad states as input. If the antitetrad states are introduced into this network then the state  $|\bar{\psi}_j^4\rangle$  should never trigger photodetector  $j$ . It will, however, lead to detection in any of the remaining three photodetectors with equal probability of  $\frac{1}{3}$ .

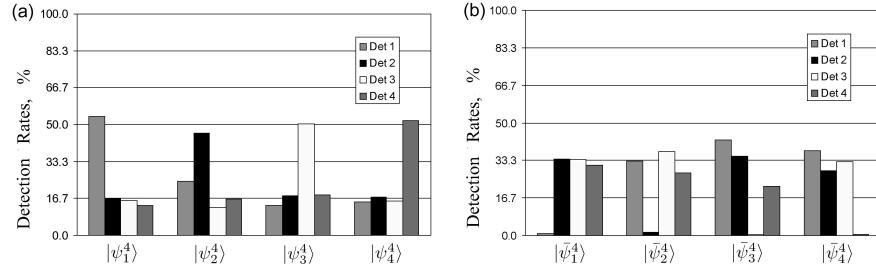
The first beamsplitter PBS1 in the interferometer for the trine experiment was changed to a non-polarising beamsplitter NPBS. Alignment of the interferometer was achieved by simulating a conventional Mach-Zehnder operation with vertically polarised input light. Wave plate WP5 was aligned with the slow axis in the vertical direction and WP7, a quarter-wave plate, was rotated so that there was complete mixing of the output states of the interferometer.

After the alignment, half-wave plate WP5 was rotated clockwise by  $54.74^\circ/2 = 27.37^\circ$ . The angle between two tetrad states is  $125.26^\circ$  on the Poincaré sphere, or  $54.74^\circ$  in polarisation.

The four tetrad and four antitetrad polarisation states required to perform this experiment were constructed using half-wave plates. For full details, see [15]. It was possible to change between the 8 input states in a matter of seconds using discrete predefined angles in the half-wave plate holders.

The experiment was performed by measuring the signal at detectors PD1-4 as the input states were changed between the 8 states. The results of the tetrad and antitetrad measurements are shown in Figs. (12.14a) and (12.14b) after calibration and normalisation.

As with the trine measurements, it is clear that the experimental results are in good agreement with the theoretical predictions. The  $\frac{1}{2} : \frac{1}{6} : \frac{1}{6} : \frac{1}{6}$  and  $\frac{1}{3} : \frac{1}{3} : \frac{1}{3} : 0$  ratios for the tetrad and antitetrad states respectively are clearly evident. For the tetrad experiment, the overall RMS deviation from the theoretical values of the channel matrix elements was 2.9%. For the antitetrad experiment, the measured value of the mutual information was 0.363 bits with a lower limit, taking into account experimental imperfections, of 0.339 bits. While this does fall short of the theoretical accessible information 0.415 bits, it is a significant improvement on the theoretical maximum that can be obtained with a von Neumann measurement, which is 0.311 bits.



**Fig. 12.14.** Histograms showing (a) the tetrad and (b) the antitetrad results, normalised so that the sum of the counts in detectors 1 to 4 is 100% for each input state. The theoretical ratios for the tetrad and antitetrad ensembles are  $\frac{1}{2} : \frac{1}{6} : \frac{1}{6} : \frac{1}{6}$  and  $\frac{1}{3} : \frac{1}{3} : \frac{1}{3} : 0$  respectively

## 12.6 Accessible information experiments with symmetric qubit ensembles

### 12.6.1 Background

In the preceding section, we described experimental investigations of the accessible information for certain ensembles, namely the trine and tetrad ensembles. In this section, we shall describe subsequent accessible information experiments, carried out by Mizuno et al [16], which involved qubit ensembles known as real symmetrical ensembles. To give the general form of such ensembles, consider again the linear horizontal and vertical polarisation states  $|\leftrightarrow\rangle$  and  $|\updownarrow\rangle$ . The real symmetric qubit states are defined as

$$|\psi_j\rangle = \cos\left(\frac{j\pi}{M}\right)|\leftrightarrow\rangle + \sin\left(\frac{j\pi}{M}\right)|\updownarrow\rangle, \quad (12.62)$$

where  $j = 0, \dots, M-1$ , for any positive, non-zero integer  $M$ . The experiments we shall describe involve ensembles where all states have equal a priori probability  $1/M$ . The accessible information is, in general, a difficult quantity to calculate. However, for such ensembles it can be calculated exactly [25]. As it happens, the trine ensemble described in the preceding section is a particular real symmetric ensemble. It corresponds to the case of  $M = 3$ .

It is known [25] that if  $M$  is even then there is actually a von Neumann measurement, i.e. a pair of orthogonal projectors, which attains the accessible information. If  $M$  is odd, however, then a POVM with at least 2 elements is generally required. However, it is also known that a POVM with 3 elements is always sufficient. The three elements can, in fact, be taken to be have rank one, i.e. they are of the form  $\Pi_j = |\omega_j\rangle\langle\omega_j|$ , where the (unnormalised) states  $|\omega_j\rangle$  are given by

$$|\omega_0\rangle = -\sin\left(\frac{\gamma}{2}\right) |\uparrow\rangle, \quad (12.63)$$

$$|\omega_1\rangle = \frac{1}{\sqrt{2}} \left( -|\leftrightarrow\rangle + \cos\left(\frac{\gamma}{2}\right) |\downarrow\rangle \right), \quad (12.64)$$

$$|\omega_2\rangle = \frac{1}{\sqrt{2}} \left( |\leftrightarrow\rangle + \cos\left(\frac{\gamma}{2}\right) |\uparrow\rangle \right), \quad (12.65)$$

and  $\gamma$  is determined by the conditions

$$\cos\left(\frac{\gamma}{2}\right) \equiv \cot\left(\frac{m\pi}{M}\right), \quad \sin\left(\frac{\gamma}{2}\right) \equiv -\sqrt{1 - \cot^2\left(\frac{m\pi}{M}\right)}, \quad (12.66)$$

for an integer parameter  $m$  within the range  $\frac{M}{4} < m < \frac{M}{2}$ . We will refer to the  $|\omega_j\rangle$  as the measurement vectors.

For the trine ensemble, the optimum POVM corresponds to  $m = 1$ . From this, one can verify that  $\cos(\gamma/2) = 1/\sqrt{3}$  and  $\sin(\gamma/2) = -\sqrt{2/3}$ . On substituting these values into (12.63)-(12.65) and normalising the states, we find that these states are the antitrine states given by (12.40)-(12.42). We should, of course, expect this since we know from the preceding section that the antitrine measurement attains the accessible information for the trine ensemble.

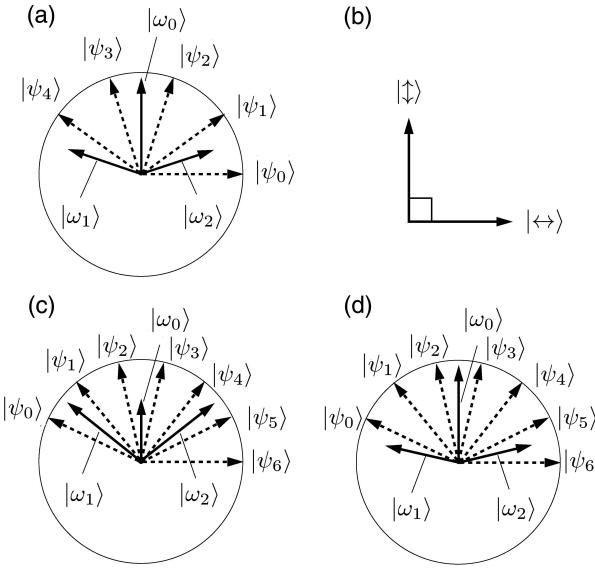
In the  $M = 5$  (quinary) and  $M = 7$  (septenary) cases, (12.63)-(12.65) result in three measurement vectors with two distinct norms. In the case of quinary letter states,  $m = 2$  and the set of measurement vectors defined by (12.63)-(12.65) is unique. The relationship between the quinary letter states and the three measurement vectors is depicted in Fig. (12.15a). Fig. (12.15b) depicts the directions of horizontal and vertical polarisation for reference.

In the septenary case, the measurement vectors in (12.63)-(12.65) are not uniquely defined. There are two different sets of these vectors, and hence two different POVMs, corresponding to the cases of  $m = 2$  and  $m = 3$ . These are depicted in Figs. (12.15c) and (12.15d).

### 12.6.2 Experiments with the trine, quinary and septenary ensembles

We noted above that for even  $M$ , the accessible information can be attained with a simple von Neumann measurement. The experiments which were performed related to the more interesting case of an odd number of states, where three outcomes are required. Here we shall describe experiments for the cases of  $M = 3, 5$  and  $7$ . In particular, we will see that the results for  $M = 3$  improve upon those obtained by Clarke et al discussed in the preceding section.

The experimental setup, which took the form of a Mach-Zehnder interferometer, is shown in Fig. (12.16). In many respects, it resembles the trine setup described in section 12.5, with the main differences being the configuration and settings of half-wave plates. The light source was a He-Ne laser (Spectra-Physics, model 117A) operating at a wavelength of 632.8 nm and

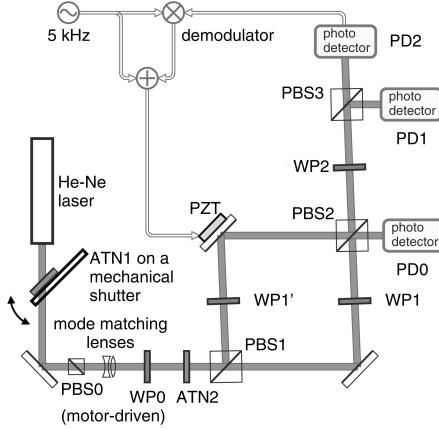


**Fig. 12.15.** The quinary and septenary states along with the measurement vectors associated with their accessible information measurements. (a) depicts the quinary states and measurement vectors. (b) illustrates the directions of horizontal and vertical polarisation states for overall reference. The septenary signals and then measurement vectors for the optimum strategies corresponding to  $m = 2$  and  $m = 3$  are depicted in (c) and (d) respectively.

power of 1 mW. The light was attenuated by attenuator ATN1 by a factor of  $10^{-6}$ . It was then purified to the state  $|\leftrightarrow\rangle$  using polarising beam splitter PBS0. The state set  $\{|\psi_j\rangle\}$  was then produced using the half-wave plate WP0. This was driven by a stepping motor, and this worked as a modulator to produce the state set. The light was subsequently fed into another attenuator, ATN2, which attenuated it by a factor of  $10^{-4}$ . Combining the effect of both attenuators, we see that at the input to the interferometer, the light power was around  $10^{-4}$  fW ( $\approx 3 \times 10^5$  photons/sec). This implies that the beam contained approximately  $10^{-3}$  photons per meter.

The Mach-Zehnder interferometer had two polarising beamsplitters, PBS1 and PBS2. Each was carefully mounted so as to operate with an extinction ratio of 1 : 1000. Each path of the interferometer contained one half-wave plate, WP1 and WP1'. The angle of WP1 to the horizontal was adjusted to  $\gamma/4$ . As a consequence, the polarisation of the light was rotated by  $\gamma/2$ . WP1' was inserted for symmetry and adjusted to have no effect on the polarisation state.

The beams from the two paths were recombined at PBS2, resulting in two output beams from the interferometer. One was detected at photodetector



**Fig. 12.16.** Experimental setup used in the accessible information experiments of Mizuno et al [16]. The states were produced by the components leading up to the attenuator ATN2. The Mach-Zehnder interferometer which terminates at PBS2 processed the states for detection. The measurement had three outcomes, realised by triggering photodetectors PD0-PD2.

PD0. The other propagated toward WP2. This half-wave plate was oriented at an angle of  $\pi/8$  to the horizontal. This beam was finally split at PBS3, sending horizontal and vertical polarisations states to photodetectors PD2 and PD1 respectively. This setup implemented the measurement with measurement vectors  $|\omega_k\rangle$  given by (12.63)-(12.65).

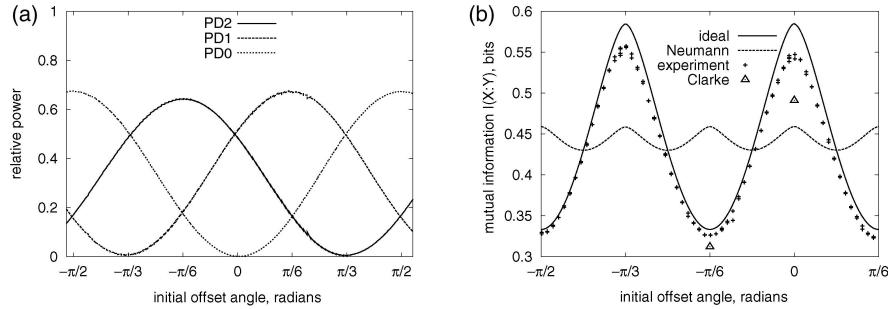
Generally, in a PBS, the separation of the two orthogonal linear polarisations depends on the incident angle of the beam. In this experiment, it was possible to obtain a separation in excess of 1:1000 for both horizontal and vertical polarisation. This was achieved by careful alignment of the interferometer to obtain the optimal angle of incidence. This is represented in Fig. (12.16) as a slight skewing of the interferometer, of approximately 0.02 radians.

In order to analyse the performance of the detector circuit, the mutual information was measured not only for the signal set  $\{|\psi_j\rangle\}$  for which the measurement is optimal, but also for the states  $\{|\psi_j(\theta_0)\rangle\}$  defined by

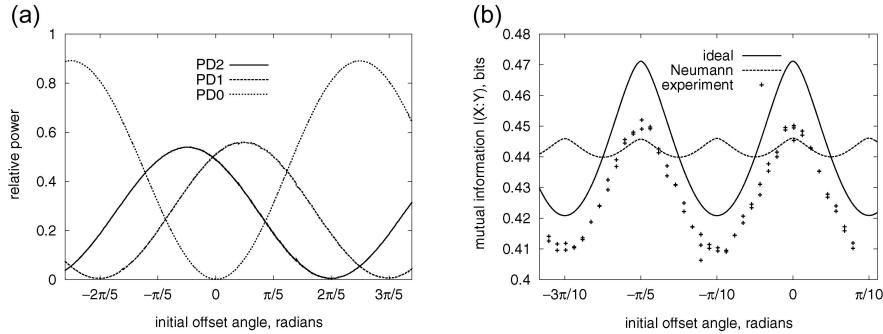
$$|\psi_j(\theta_0)\rangle = \cos\left(\frac{j\pi}{M} + \theta_0\right) |\leftrightarrow\rangle + \sin\left(\frac{j\pi}{M} + \theta_0\right) |\downarrow\rangle. \quad (12.67)$$

We see from this expression and (12.62) that  $|\psi_j(0)\rangle = |\psi_j\rangle$ . The offset angle  $\theta_0$  was increased in steps of  $\pi/90$  radian (two degrees)

The half-wave plate WP1 was adjusted to give the values of the angle  $\gamma$  corresponding to the optimum measurements for the trine, quinary and



**Fig. 12.17.** (a) shows the dependence of the relative outputs at the three PDs on the offset angle of the injected beam in the trine experiment. (b) shows the dependence of the mutual information on the initial offset angle (“experiment”, pluses). The theoretical value (“ideal”, solid curve) and the optimal von Neumann case (“Neumann”, dashed curve) are shown for comparison. The values obtained in the experiment of Clarke et al (“Clarke”, triangles at  $\theta_0 = 0$  and  $-\pi/6$ ) are also shown.



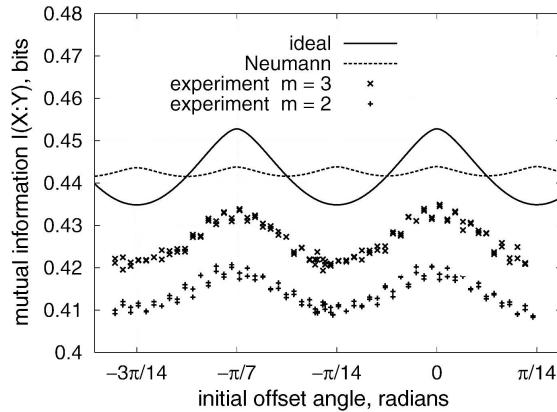
**Fig. 12.18.** (a) depicts the dependence of the relative outputs at the three PDs on the offset angle of the injected beam in the quinary experiment. (b) shows the corresponding dependence of the mutual information on the initial offset angle.

septenary ensembles. In particular, for the septenary ensemble, the optimum measurements corresponding to both  $m = 2$  and  $m = 3$  were carried out.

Figure (12.17a) shows the relative output counts at the three detectors as the polarisation of the input light was varied in the trine case. This relative power corresponds to the probability for the measurement outcome to occur for a single input photon.

For the polarisation angles  $\{-\pi/6, \pi/6, \pi/2\}$  and  $\{-\pi/3, 0, \pi/3\}$  we have the trine and antitrine measurements respectively. The mutual information in Fig. (12.17b) corresponds to the data depicted in Fig. (12.17a). At the optimum operating point, corresponding to the best detection strategy, the mutual information exceeds that attainable with the best von Neumann mea-

surement. Its value also exceeds that obtained by Clarke et al (represented as triangles in the figure). The reason for this is the smaller PBS error. The experimental value is slightly lower than the theoretical maximum and this is due mainly to a residual PBS error of approximately 0.1% and also to the imperfect contrast of interference. It was found that even though the PBS error was not the limiting factor of the interference contrast, it had non-negligible effects on the mutual information.



**Fig. 12.19.** Mutual information as a function the initial offset angle  $\theta_0$  in the septenary experiment. Here, the results of the experiment for  $m = 2$  are depicted by crosses, while the results for  $m = 3$  are pluses.

Figure (12.18) shows the relative output counts at the three detectors for the quinary case. This data also shows a visible improvement over the best possible von Neumann measurement. Again, the discrepancy between the experimental result and the theoretical value is mainly attributable to the PBS error and imperfect contrast.

Figure (12.19) shows the values of the mutual information obtained with the two possible optimum detection schemes for the septenary case. Unfortunately, even the theoretical value gives only a very small improvement over that which can be attained with a von Neumann measurement, and this improvement was impossible to verify experimentally.

## 12.7 Superadditivity of the classical capacity

### 12.7.1 Background

In this section we shall describe a novel quantum effect known as the superadditive quantum coding gain which was recently demonstrated experimentally

by Fujiwara et al [17]. The longer paper by Takeoka et al [27] gives a more extensive discussion of this experimental work. See also refs. [28–31] for further theoretical discussions of this effect. Prior to discussing this experiment, we shall describe the relevant theory to place it in context.

In section 12.5, we defined the first order classical capacity  $C_1$  of a set of states as the maximum value of the mutual information with respect to both the measurement strategy and the a priori probabilities of the states. In the accessible information only the former maximisation is carried out.

In this definition of the capacity, the letter states are measured individually. However, as we shall see here, if we allow collective measurements to be performed on long strings of letter states, a higher value of the capacity may be obtained. This effect is known as the superadditive quantum coding gain (SQCG). It is explained in more detail below.

Instead of considering measurements on single letter states, we can consider a collective measurement on a string of  $n$ -letter states. Following the notation of section 12.5, the set of  $n$ -letter strings, or codewords, will be denoted by  $\{\mathbf{x}\}$ , where each  $\mathbf{x}$  is of the form  $x_{j_1} \dots x_{j_n}$ . The letter  $x_{j_r}$  is conveyed by the letter state  $|\psi_{x_{j_r}}\rangle$ , where  $r = 1, \dots, n$ . So, an  $n$ -letter string will be conveyed by the corresponding tensor product state  $|\Psi_{\mathbf{x}}\rangle = |\psi_{x_{j_1}}\rangle \otimes \dots \otimes |\psi_{x_{j_n}}\rangle$ . The measurements we shall consider are again represented by POVM elements. We shall denote the set of possible results of a measurement by  $\{\mathbf{y}\}$ . In contrast to our discussion in section 12.5, we do not require the measurements to be on individual signal carriers. This implies that the POVM elements can be any positive operators on the entire  $n$ -particle Hilbert space satisfying (12.10), where ‘1’ is the identity operator on this space. This is known as collective decoding and for such a measurement, the channel matrix has the form

$$P(\mathbf{y}|\mathbf{x}) \equiv \text{Tr}(\Pi_{\mathbf{y}}\Psi_{\mathbf{x}}). \quad (12.68)$$

The mutual information for this measurement is

$$I(X^n : Y^n) = \sum_{\mathbf{x}} P(\mathbf{x}) \sum_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}) \log \left[ \frac{P(\mathbf{y}|\mathbf{x})}{\sum_{\mathbf{x}'} P(\mathbf{x}') P(\mathbf{y}|\mathbf{x}')} \right]. \quad (12.69)$$

Using this definition we can define the capacity in the same way as was done in section 12.5 as the maximum of the mutual information with respect to the measurement and the a priori probabilities, that is, as

$$C_n = \max_{\{P(\mathbf{x})\}} \max_{\{\Pi_{\mathbf{y}}\}} I(X^n : Y^n). \quad (12.70)$$

We call this the classical capacity of order  $n$ . The order of the capacity is the number of signal systems involved in the collective decoding. These capacities have the property of being superadditive:

$$C_n \geq n C_1. \quad (12.71)$$

What this means is that the maximum amount of information that can be reliably extracted from  $n$  signals using collective decoding is at least as great as the amount that can be extracted if the  $n$  signals states are decoded (measured) individually.

A set of states which exhibits a SQCG is a set for which there is a strict inequality in (12.71). One such set is the trine set. In this section, we shall denote the trine states by  $|\psi_j\rangle$ , where  $j = 0, 1, 2$  as in (12.62) in the preceding section with  $M = 3$ . The first order capacity  $C_1$  of this set has been carefully studied and evaluated to be 0.6454 bits [32, 33]. It is attained by discarding one of the three states and using only the remaining pair, for example  $\{|\psi_0\rangle, |\psi_1\rangle\}$ , and assigning to these equal a priori probability of  $1/2$ . For the trine states, the first order capacity  $C_1$  can be attained by a von Neumann measurement in the orthonormal basis

$$|\omega_0\rangle = \frac{\sqrt{2+\sqrt{3}}}{2\sqrt{3}}|\psi_0\rangle + \frac{\sqrt{2-\sqrt{3}}}{2\sqrt{3}}|\psi_1\rangle, \quad (12.72)$$

$$|\omega_1\rangle = \frac{\sqrt{2-\sqrt{3}}}{2\sqrt{3}}|\psi_0\rangle + \frac{\sqrt{2+\sqrt{3}}}{2\sqrt{3}}|\psi_1\rangle. \quad (12.73)$$

For the trine states, a SQCG can be achieved with length-two codes. However, the experimental demonstration of this we shall describe does not, unlike the experiments described so far, use only polarisation qubits. So, for time being, we shall represent general states of a qubit in terms of a non system-specific orthonormal basis set  $\{|0\rangle, |1\rangle\}$ .

There are nine possible length-two codewords for the trines. Peres and Wootters [18] showed that, using only the following three:

$$\begin{aligned} |\Psi_{jj}\rangle &= |\psi_j\rangle \otimes |\psi_j\rangle \\ &= \frac{1}{2}(1 + \cos(\phi_j))|0\rangle \otimes |0\rangle + \frac{1}{2}\sin(\phi_j)\left(|0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle\right) \\ &\quad + \frac{1}{2}(1 - \cos(\phi_j))|1\rangle \otimes |1\rangle, \end{aligned} \quad (12.74)$$

with equal probability of  $1/3$ , and where  $\phi_j = 2\pi j/3$ , a SQCG gain can be achieved by decoding with the square-root measurement. From (12.29) and (12.30) we see that this has the POVM elements  $\Pi_{jj} = |\omega_{jj}\rangle\langle\omega_{jj}|$ , where

$$|\omega_{jj}\rangle = \left(\sum_{k=0}^2 |\Psi_{kk}\rangle\langle\Psi_{kk}|\right)^{-\frac{1}{2}}|\Psi_{jj}\rangle. \quad (12.75)$$

With this measurement, one can extract  $I(X^2 : Y^2) = 1.3690$  bits of information. This is greater than double the value of  $C_1 = 0.6454$  bits. The SQCG per letter is then  $I(X^2 : Y^2)/2 - C_1 = 0.0391$  bits.

The measurement vectors  $|\omega_{jj}\rangle$  may be written as

$$|\omega_{00}\rangle = c |\Psi_{00}\rangle - \frac{s}{\sqrt{2}} |\Psi_{11}\rangle - \frac{s}{\sqrt{2}} |\Psi_{22}\rangle, \quad (12.76)$$

$$|\omega_{11}\rangle = -\frac{s}{\sqrt{2}} |\Psi_{00}\rangle + c |\Psi_{11}\rangle - \frac{s}{\sqrt{2}} |\Psi_{22}\rangle, \quad (12.77)$$

$$|\omega_{22}\rangle = -\frac{s}{\sqrt{2}} |\Psi_{00}\rangle - \frac{s}{\sqrt{2}} |\Psi_{11}\rangle + c |\Psi_{22}\rangle, \quad (12.78)$$

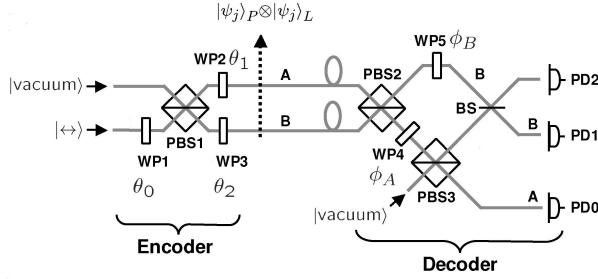
where  $c = \cos(\frac{\gamma}{2}) = (\sqrt{2} + 1)/\sqrt{6}$ , and  $s = \sin(\frac{\gamma}{2}) = (\sqrt{2} - 1)/\sqrt{6}$  ( $\gamma \simeq 19.47^\circ$ ). As it happens, this measurement is the optimal measurement for minimum error discrimination among the states  $|\Psi_{jj}\rangle$  with equal a priori probabilities [23].

### 12.7.2 Experiment with the trine states

The SQCG demonstration carried out by Fujiwara et al [17] made use of two physically different kinds of qubit carried by a single photon. In each of the three length-two codeword states, the first letter was encoded in the polarisation state of the photon, while the second letter was encoded in the same photon's location. The reason why both qubits were not encoded in the polarisation states of different photons is that performing the square-root measurement on a pair of such systems would involve photon-photon interactions. The non-linear processes needed to carry this out are not, at the time of writing, sufficiently reliable as to make the small SQCG visible above experimental error.

Like the other experiments we have described so far, the experiment which demonstrated the SQCG involved both state preparation and measurement. The first part of the experiment involved the preparation of the length-two codeword states in (12.74). The second part involved performing the square-root measurement on these states. The entire optical circuit used is shown in Fig. (12.20). The angles  $(\theta_0, \theta_1, \theta_2)$  of the three half-wave plates (WPs) had the values  $(0^\circ, 0^\circ, 0^\circ)$ ,  $(30^\circ, -30^\circ, -15^\circ)$  and  $(30^\circ, 30^\circ, 15^\circ)$  for  $|\Psi_{00}\rangle$ ,  $|\Psi_{11}\rangle$ , and  $|\Psi_{22}\rangle$ , respectively.

The decoding part of the apparatus realised the square-root measurement, which has three outcomes corresponding to the three possible codeword states. Each of the three possible outcomes was signalled by a count at one of three avalanche photodetectors (PD0-PD2). When photodetector PD $j$  was triggered, the received codeword state was decided to be  $|\Psi_{jj}\rangle$ , for  $j = 0, 1, 2$ . The states of the polarisation qubit for the first letter will, as in the other experiments we have described which involve optical polarisation states  $|\leftrightarrow\rangle$  and  $|\downarrow\downarrow\rangle$ . The location qubit for the second letter corresponded to the two possible exits A and B from polarising beamsplitter PBS1. The length-two codeword states lie in the space spanned by the orthonormal states



**Fig. 12.20.** Quantum circuit to realise the encoding and collective decoding of the states in (12.74) using the square-root measurement. The angles of the WPs,  $\theta_0$ ,  $\theta_1$ , and  $\theta_2$  were chosen as described in the text. Also,  $\phi_A = -\gamma/2 = -9.74^\circ$  and  $\phi_B = -45^\circ$ .

$$|0\rangle_P \otimes |0\rangle_L = |\leftrightarrow\rangle_A \otimes |\text{vacuum}\rangle_B, \quad (12.79)$$

$$|0\rangle_P \otimes |1\rangle_L = |\text{vacuum}\rangle_A \otimes |\leftrightarrow\rangle_B, \quad (12.80)$$

$$|1\rangle_P \otimes |0\rangle_L = |\uparrow\rangle_A \otimes |\text{vacuum}\rangle_B, \quad (12.81)$$

$$|1\rangle_P \otimes |1\rangle_L = |\text{vacuum}\rangle_A \otimes |\uparrow\rangle_B. \quad (12.82)$$

The light source used in this experiment was a He-Ne laser, emitting CW light and operating at a wavelength of 632.8 nm with 1 mW power. The light was strongly attenuated to the level of  $10^{-2}$  photons present at any one time in the entire circuit. The signal photons propagated through the apparatus toward the three Si PDs. The quantum efficiency of each of the PDs was 70% and the dark count rate was 100 [count/sec]. The photons were guided through a multimode optical fibre with coupling efficiency of approximately 80%.

Evaluation of the mutual information,  $I(X^2 : Y^2)$ , which was necessary for verification of the SQCG, was achieved by experimentally determining the elements of the 3-by-3 channel matrix  $P(kk|jj) \equiv |\langle \omega_{kk}|\Psi_{jj} \rangle|^2$ . These were obtained by collecting the statistical data of counts at the three photodetectors and the states which caused them. The error performance was determined solely by the non-orthogonality of the signal states, imperfect alignment of the whole interferometer and the dark count of the PDs.

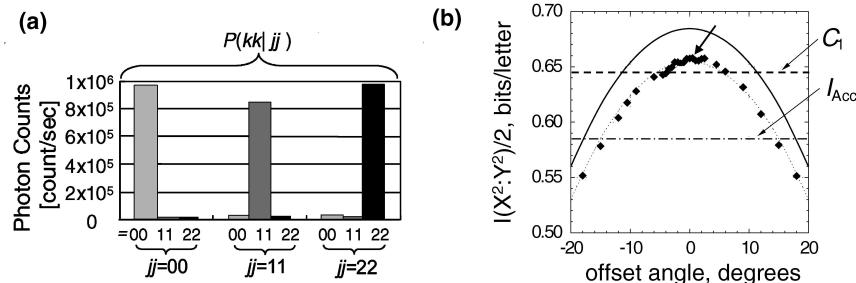
Inspection of Fig. (12.20) shows that the entire circuit contained two linked polarisation Mach-Zehnder interferometers. These had to be adjusted simultaneously to a proper operating point. This was done using a bright reference beam and piezo transducers, having low noise voltage sources. The visibility of the entire interferometer was typically 98%. The reference beam was shut off following this adjustment. The signal light was then guided into the encoder. The three codeword states were prepared, and for each one, photon counts were measured for a duration of five seconds. The temporal stability of this procedure depended on there being a relative path length

change no greater than 3 nm for at least than 200 sec. This bound gives an error in the mutual information of  $\pm 0.005$  bits.

Figure (12.21a) depicts the elements of the experimentally determined channel matrix. Theoretically, the diagonal elements should be equal to  $c^2 = 0.9714$  and the off-diagonal elements should have the value  $s^2/2 = 0.0143$ . The total number of events counted during an interval of 1 sec was typically of the order of  $10^6$ . The average count for the off-diagonal elements was  $1.9 \times 10^4$ . Some of the counts were due to background photons. These gave rise to approximately 300 counts per second. Incorporating the effect of dark counts, the total background photon count was 2% of the average off-diagonal element photocount.

For the experimentally determined channel matrix, the mutual information was evaluated to be  $I(X^2 : Y^2) = 1.312 \pm 0.005$  bits. For the sake of experimental clarity, the mutual information was measured for the codeword state set  $\{|\Psi_{jj}\rangle\}$  rotated with respect to the measurement vector set  $\{|\omega_{kk}\rangle\}$  around the axis of 3-fold symmetry. The results of this are shown in Fig. (12.21b). Here, the theoretical mutual information is the solid curve and the optimum value corresponds to the histogram in Fig. (12.21a). At this point, the departure of the experimental results from theoretically obtainable value is mainly a consequence of the imperfection of the PBSs. Furthermore, fluctuation of the data points was mainly due to thermal drift. The corresponding errors were approximately 0.005 bits.

The experimentally determined mutual information per letter was  $0.656 \pm 0.003$  bits per letter, which is clearly an improvement over the first order capacity  $C_1 = 0.6454$  bits per letter. The latter is depicted by the dashed lines. These results clearly demonstrate the SQCG for length-two codewords constructed using the trine states.



**Fig. 12.21.** (a) Histogram of photon counts for the channel matrix elements  $P(kk|jj)$  corresponding to the maximum mutual information, which corresponds to the result indicated by the leftmost arrow in (b). (b) depicts the measured (diamonds) and theoretical (solid curve) mutual information as a function of the offset angle. The dotted curve is a guide for eyes. The theoretical  $C_1$  and accessible information  $I_{\text{Acc}}$  are shown by the dashed and one-dotted lines, respectively.

## 12.8 Discussion

The purpose of this chapter has been to review the experimental progress that has been made to date in the area of quantum state discrimination. The theory of state discrimination dates back three decades to the pioneering works of Helstrom [8], Holevo [21], Yuen, Kennedy, Lax [22] and others. However, it is only relatively recently that the measurement schemes devised by these authors have been implemented experimentally. As we pointed out in the introduction, discrimination among a set of non-orthogonal quantum states at the limits imposed by quantum theory requires highly sophisticated, high quality apparatus. It is only recently that these have come to meet the standards required to carry out the experimental procedures. This area of research, both theory and experiment, has also benefited considerably from its close association with the emerging subject of quantum information technology, which is becoming one of the most important and fascinating areas of science as a whole.

We saw that quantum state discrimination is closely related to the transmission of classical information using quantum systems, though sometimes in very unexpected ways. Tasks which are seemingly unrelated (such as unambiguous state discrimination and the trine/antitrine measurements) may nevertheless be theoretically equivalent. This is highly advantageous, as it allows us to perform a number of investigations of quantum measurement limits using the same experimental apparatus, by changing only the input states.

Measurements of the kind we have described will find numerous applications as quantum communications technology takes hold and moves from the laboratory to the marketplace. We also expect to see, in due course, further refinements of, and variations on the experiments we have described. One major goal of optical state discrimination is to move from signal carriers which are weak, almost empty optical pulses, to single photons. These could be produced, for example, using parametric down-conversion. Single-photon sources would be highly desirable for many practical uses of quantum communication, such as in quantum key distribution, where optimal state discrimination could be useful to the receiver (and indeed also an eavesdropper!). There are also interesting applications of quantum state discrimination to purely scientific problems, such as quantum state comparison [34–36], which would also benefit from such improvements. Also, from the point of view of improving proof-of-principle demonstrations of state discrimination, the stochastic nature of the experiments we have described is, to a large extent, due to the weakness of the pulses, very few of which ever contain a photon. This issue would be resolved by deterministic, single-photon sources.

A further future goal is the implementation of state discrimination measurements using non-optical systems. Some interesting proposals have been put forward wherein the states to be discriminated are those of alternative quantum systems, such as cold ions stored in a linear trap [37]. Here, the ac-

tual state discrimination experiment is a quantum computation. Such proposals are promising, in view of the fact that elementary quantum computations can be carried out using such systems [38] .

## Acknowledgements

There are several people to whom I would like to express my gratitude. Firstly, I would like to thank Matteo G. Paris and Jarda Rehacek for inviting me to make this contribution. I would also like to thank Masahide Sasaki and Masahiro Takeoka for their helpful suggestions and clarifications, and also for permission to use their materials. Over the years, I have had several enjoyable collaborations and conversations on problems relating to the subject of state discrimination, both theory and experiment. Were it not the skill, wisdom and insight of my colleagues involved in the experimental aspects of this subject, most of the achievements described in this review would not have been made. In this regard, I would like to thank, in addition to the people I have mentioned above, Makoto Akiba, Stephen M. Barnett, Roger B. M. Clarke, Mikio Fujiwara, Tetsuya Kawanishi, Vivien M. Kendon, Jun Mizuno and Erling Riis. I would also like to thank Richard Jozsa for introducing me to TexPoint, which greatly simplified the production of many of the figures. This work was supported by a University of Hertfordshire Postdoctoral Fellowship.

## References

1. M. A. Nielsen and I. L. Chuang: *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge 2000).
2. A. Chefles: Contemp. Phys. **41**, 401 (2000) and references therein.
3. M. Dusek, M. Jahma and N. Lütkenhaus: Phys. Rev. A **62**, 022306 (2000).
4. S. J. D. Phoenix, S. M. Barnett and A. Chefles: J. Mod. Opt **47**, 507 (2000).
5. A. Chefles: Phys. Rev. A **66**, 042325 (2002).
6. C. H. Bennett and S. Wiesner: Phys. Rev. Lett. **69**, 2881 (1992).
7. S. M. Barnett and E. Riis: J. Mod. Opt. **44**, 1061 (1997).
8. C. W. Helstrom: *Quantum Detection and Estimation Theory* (Academic Press, New York 1976).
9. B. Huttner, A. Muller J. D. Gautier, H. Zbinden and N. Gisin: Phys. Rev. A **54**, 3783 (1996).
10. R. B. M. Clarke, A. Chefles, S. M. Barnett and E. Riis: Phys. Rev. A **63**, 040305(R) (2001).
11. I. D. Ivanovic: Phys. Lett. A **123**, 257 (1987).
12. D. Dieks: Phys. Lett. A **126**, 303 (1988).
13. A. Peres: Phys. Lett. A **128**, 19 (1988).
14. A. Chefles: Phys. Lett. A **239**, 339 (1998).
15. R. B. M. Clarke, V. M. Kendon, A. Chefles, S. M. Barnett, E. Riis and M. Sasaki: Phys. Rev. A **64**, 012303 (2001).

16. J. Mizuno, M. Fujiwara, M. Akiba, T. Kawanishi, S. M. Barnett and M. Sasaki: Phys. Rev. A **65**, 012315 (2001).
17. M. Fujiwara, M. Takeoka, J. Mizuno and M. Sasaki: Phys. Rev. Lett. **90**, 167906 (2003).
18. A. Peres and W. K. Wootters: Phys Rev. Lett. **66**, 1119 (1991).
19. B. H. Bransden and C. J. Joachain: *Quantum Mechanics*, 2nd edn (Addison-Wesley, Reading, Mass., 2002).
20. G. Jaeger and A. Shimony: Phys. Lett. A **197**, 83 (1995).
21. A. S. Holevo: J. Multivar. Anal. **3**, 337 (1973).
22. H. P. Yuen, R. S. Kennedy and M. Lax: IEEE Trans. Inform. Theory **IT-21**, 125 (1975).
23. M. Ban, K. Kurokawa, R. Momose, and O. Hirota: Int. J. Theor. Phys. **36**, 1269 (1997).
24. T. Cover and J. Thomas: *Elements of Information Theory* (John Wiley and Sons, New York, 1991).
25. M. Sasaki, S. M. Barnett, R. Jozsa, M. Osaki and O. Hirota: Phys. Rev. A **59**, 3325 (1999).
26. E. B. Davies: IEEE Trans. Inform. Theory **IT-24**, 596 (1978).
27. M. Takeoka, M. Fujiwara, J. Mizuno and M. Sasaki: Experimental Superadditive Quantum Coding Gain, arXiv:quant-ph/0306034 (2003).
28. M. Sasaki, K. Kato, M. Izutsu and O. Hirota: Phys. Lett. A **236**, 1 (1997).
29. M. Sasaki, K. Kato, M. Izutsu and O. Hirota: Phys. Rev. A **58**, 146 (1998).
30. J. R. Buck, S. J. van Enk and C. A. Fuchs: Phys. Rev. A **61**, 032309 (2000).
31. S. Usami, T. S. Usuda, I. Takumi, R. Nakano and M. Hata: *Quantum Communication, Computing, and Measurement* 3 ed by P. Tombesi and O. Hirota (Kluwer Academic/Plenum, New York, 2001) p 35; T. S. Usuda, S. Usami, I. Takumi and M. Hata: Phys Lett. A **305**, 125 (2002).
32. M. Osaki, M. Ban, and O. Hirota: *Quantum Communication, Computing, and Measurement* 2 ed by P. Kumar, G. M. D'Ariano, and O. Hirota (Kluwer academic/Plenum publishers, New York, 2000) p 17.
33. P. W. Shor: The Adaptive Classical Capacity of a Quantum Channel, or Information Capacities of Three Symmetric Pure States in Three Dimensions, arXiv:quant-ph/0206058 (2002).
34. S. M. Barnett, A.Chefles and I. Jex: Phys. Lett. A **307**, 189 (2003).
35. I. Jex, E. Andersson and A. Chefles: Comparing the States of Many Quantum Systems, arXiv:quant-ph/0305120 (2003). To appear in Journal of Modern Optics.
36. A. Chefles, E. Andersson and I. Jex: Unambiguous Comparison of Multiple Pure Quantum States. In preparation.
37. S. Franke-Arnold, E. Andersson, S.M. Barnett and S. Stenholm: Phys. Rev. A **63**, 052301 (2001).
38. J. I. Cirac and P. Zoller: Phys. Rev. Lett. **74**, 4091 (1995).



# Index

- absorption
  - absorption index, 385
  - polarization selective, 421, 430
  - strong, 391
- accessible information, 489, 493, 497, 499–501, 503, 505
- ADC, 248
- algebra
  - Lie, 377
  - SU(2), 377
- algorithm
  - Deutsch-Jozsa, 440, 442
  - EM, 75, 387
    - convergence, 387
  - EMU, 77, 96
  - expectation-maximization, 75, 387
  - FBP, 389, 395, 410
  - iterative, 387, 388, 408, 411
- amplifier, 248, 249, 264, 275–278, 281
  - quantum-injected parametric amplifier, 336
- amplitude damping channel, 93
- ancilla, 429
  - ancilla space, 430
  - ancilla state, 451
  - ancilla system, 452
- anticorrelation, 274
- antitetrad states, 496, 499
- antitrine states, 491, 494, 501, 503, 505, 511
- array detection, 282
- balanced homodyne detection, 241
- ballistic expansion, 214
- Bayes rule, 157, 246
- Bayesian view, 155
- BCH formula, 13, 17
- BE, 214
- beamsplitter, 434, 498, 499
  - polarizing, 132, 133, 430, 475, 481, 492, 498, 501, 509, 510
- Bell inequality, 166
- Bell states, 461
- BHD, 241
- bipartite states notation, 304
- Bloch
  - sphere, 336, 378
  - vector, 378
- Bloch sphere, 336
- BMC, 359
- Born rule, 151
- Bose mode coalescence, 359
- calculus of variations, 387
- Cauchy-Schwarz inequality, 78
- cavity, 193, 226
- CCD, 291, 385, 394
- central limit theorem, 26
- channel matrix, 493, 494, 497, 509, 510
- classical capacity, 491, 505, 506, 510
  - superadditivity, 472, 505, 506, 510
- cloning machine, 336, 346, 348
  - fidelity, 362
  - optimal, 336, 348
- coherence
  - second-order, 271
- coherent state
  - coherent, 12
- collective decoding, 506, 508
- commutation relation, 193
- complete positivity, 90
- completely positive map, 89
- computing time, 412
- conclusive result, 442, 453
- concurrence, 126
- contextuality, 353, 362

- convexity, 68, 77
- correlation
  - between quadratures, 290
  - matrix, 287
  - polarization, 273
  - two-mode, 271
- cost
  - quadratic, 380
  - rectangular, 380, 382
- CP map, 303
  - inversion, 304
  - Kraus decomposition, 305
  - representations, 303
- Cramér-Rao lower bound, 50, 84, 325, 380
  - constrained, 326
- CRLB, 84, 380, 381
- cross section, 385, 391, 397
- CT, 384
- data
  - incomplete, 234
  - phase sensitive, 377
  - simulated, 388
- DC, 243
- de Finetti theorem, 151
  - classical, 156, 159, 162
  - for quantum operations, 177, 179
    - proof, 180
  - quantum, 157, 162, 166
    - proof, 167, 173
- depolarizing channel, 93
- detection state, 451, 453, 454
- detector
  - array, 282, 285, 290
  - balanced homodyne, 251, 292
  - BHD-balanced-homodyne, 272
  - DC balanced-homodyne, 271
  - DC pulsed homodyne, 252
  - efficiency, 254, 255
  - homodyne, 410
  - losses, 256
  - noise, 280
  - photodiode, 248
  - photoemissive, 243
  - photon-number-discriminating, 261
  - position-sensitive, 385
  - RF, 279
  - single photon, 275
- deterministic inversion, 197
- direct sampling method, 198
- disentangling theorem, 339
- distribution
  - exchangeable, 161, 170, 171
  - finitely exchangeable, 161
  - marginal, 13, 194, 197, 207, 211
  - multinomial, 85
  - phase-averaged, 261
  - quadrature, 244, 260, 261
  - symmetric, 161, 162
  - Wigner, 243, 244, 253–256
- downconversion, 404, 511
- EM, 387
- EMU, 76
- ensemble, 488
- entangled probes, 94
- entropy, 126, 145, 201
  - von Neumann, 199, 206, 208, 210
- EPR, 239
- equation
  - extremal, 388, 396, 398
    - for CP maps, 92
    - for optimal POVM, 105, 109
  - Liouville, 199
  - Schrödinger, 198, 408
- error volume, 86
- estimation
  - biased, 381
  - constrained, 382
  - ML, 380
  - NFM, 380
  - single-parameter, 381
  - unconstrained, 382
  - unphysical, 381
- estimator, 377
  - point, 377
  - uncertainty, 381
- exchangeable sequence
  - of density operators, 157
  - of quantum operations, 178, 179
  - of quantum states, 156
- exponential attenuation, 385
- faithfulness, 309
  - and separability, 318
  - ensemble of states, 315
  - entangled state, 310

- generic bipartite state, 312
- in infinite dimensions, 319
- measure of, 314
- patching of unfaithful states, 316
- set of generators, 310
- FBP, 384, 409
- fidelity, 82, 126, 206, 207, 209, 210, 212, 218, 336, 437, 438, 447
- filtered back projection, 254
- Fisher information, 50, 84, 325, 412
  - matrix, 85
- function
  - Bessel, 400
  - Boolean, 440
  - characteristic, 194
  - Gaussian-Laguerre, 244
  - Hermite-Gaussian, 244
  - Husimi, 288
  - Laguerre-Gaussian, 401
  - likelihood, 377, 378
  - partition function, 200
  - pattern function, 197, 198, 210–212, 258, 259, 261, 267, 271
  - Q-function, 288
  - sampling function, 198
  - Wigner, 11, 24, 193–196, 243, 409, 411
  - Wigner, unphysical, 207
- GHZ, 166
- GRIPS, 270
- half-wave plate, 132, 133, 484, 485, 492, 498, 499, 501, 503, 508
- Helstrom bound, 443, 450, 470, 473–475, 482, 488
- Helstrom formula, 445, 457–459
- Hermite polynomial, 408
- heterodyne detection, 19
- Hilbert-Schmidt distance, 87
- Holevo-Yuen-Kennedy-Lax conditions, 488
- hologram, 402, 404, 405
- homodyne detection, 16, 241, 247, 270, 378
- 8-port scheme, 378
- IDL, 389
- IDP, 471
- inconclusive result, 108, 422–424, 429, 431
- inference, 84, 155
- information, 84
- interferometer, 482, 484, 491, 498, 499, 501, 503, 509
  - Mach-Zehnder, 377
  - MZ, 377, 378
  - perfect-neutron, 394
- ion trap, 511
- Ivanovic-Dieks-Peres bound, 471, 479, 480, 482, 485
- Jamiolkowski isomorphism, 90
- Kraus decomposition, 90
- Kullback-Leibler distance, 386
- Lagrange multiplier, 200
- Laplace's indifference principle, 204
- laser, 475, 480, 482, 501, 509
  - pulsed, 382
- least squares inversion, 103
- LG, 401
- light
  - circularly polarized, 272
  - linearly polarized, 272
  - nonclassical, 275
  - Poissonian, 264
  - polarization squeezed, 270
  - squeezed, 260, 279
  - thermal, 264
- likelihood, 377, 408
  - Poissonian, 379, 396
- likelihood functional, 77, 91, 99
- linear and positive problem, 74, 386
- linear entropy, 126, 145, 147
- linear map, 309
- linear optical sampling, 248, 264
- linearly dependent states, 433, 436, 459, 460
- linearly independent states, 430, 433, 459
- LinPos, 74, 101, 323, 386
- LO, 247
- local oscillator, 16, 247, 408
- LOCC, 461, 462, 464
- MaxEnt, 192

- MaxEnt principle, 204
- maximum likelihood estimation
  - of quantum processes, 88
- maximum-likelihood estimation, 48, 64, 67, 133, 141, 143–145, 147, 323
- absorption tomography, 384
- bootstrap, 331
- Gaussian state, 55
- homodyne tomography, 50, 406
- of quantum measurements, 98
- of quantum phase, 376
- of quantum processes, 324
- of quantum states, 70
- phase tomography, 393
- photonic qutrits, 401
- spin tomography, 50
- measurement
  - Bell measurement, 356
  - collective, 476, 505, 506, 508
  - conditional, 463
  - correlation, 267
  - generalized, 471, 477
  - incomplete, 205
  - informationally complete, 168, 169
  - joint, 455
  - minimal, 169, 170
  - of modes, 247
  - simultaneous, 289
  - square-root, 453, 488, 507, 508
  - uninformative, 175
  - von Neumann, 471, 473, 486, 494, 501, 505
- MEMS, 126
- microwells, 214
- minimum error state discrimination, 104, 470, 473, 488, 491–494, 497
- mirror-symmetric states, 450
- ML, 64, 323, 377
- moments, 262
- Monte Carlo simulations, 92
- Moore-Penrose pseudo inverse, 314
- mutual information, 489, 491, 494, 497, 499, 503, 506, 509, 510
- MZ, 377
- Neumark’s theorem, 428, 451
- neutral density filter, 475, 484
- NFM, 379, 380
- noise
  - Gaussian, 218
  - phase-insensitive, 378
- nPCT, 393, 397
- OAM, 401
- observation level, 192, 200, 201
  - complete, 192
  - extension, 202
  - incomplete, 192
  - reduction, 203
- OHT, 241
- OPA, 337
- operator
  - detection, 423, 424, 456, 459
  - Lagrange, 449
  - positive, 473, 477–479
  - projection, 473, 477
  - pseudo-spin, 339
- optical fibre, 401
- optical homodyne tomography, 254
- orbital angular momentum, 401
- overcomplete states, 451, 454, 488, 491
- parametric downconversion, 41
- parametric oscillator
  - gain, 338
  - injected, 338
- parity operator, 225–227
  - displaced, 227, 230, 232
- pattern function
  - averaging, 321
- Pauli problem, 240
- PDL, 480
- phase
  - accumulated, 395
  - ambiguous, 395
  - asymptotic dispersion, 380
  - auxiliary shifter, 378, 395, 396
  - dispersion, 380
  - estimate, 377, 378, 380
  - estimation
    - Gaussian, 380
    - ML, 382
    - Poissonian, 380
    - semi-classical, 379, 382
  - estimator
    - asymptotical dispersion, 382
    - ML, 382
    - NFM, 382

- unconstrained, 382
- fluctuations, 383
- ideal measurement, 376
- measurement, 376
  - standard limit, 381
- ML tomography, 398
- operational, 378, 379
- operator, 376
- Pegg-Barnett, 256
- reference, 244, 397, 399
- resolution, 377
- shifter, 452
- Susskind-Glogower, 376
- true, 377, 380, 382
- von Mises distribution, 400
- phase space, 193, 194, 203, 210, 212
  - rotation, 223
- photocurrent, 248–250
- photodetection, 14
- photodetector, 475, 481, 482, 484, 492, 498, 499, 503, 508, 509
- physical constraints, 382
- Poincaré sphere, 122–124, 130–133, 139, 141, 270, 491, 497
- polarization, 122
- polarization controller, 481
- polarization-dependent-loss fibre, 480, 481
- polarizer, 475, 481
- positive, operator-valued measure, 477–479, 488, 489, 491, 493, 494, 497, 500, 506, 507
- positivity, 67
- POVM, 308, 376, 422, 477
- pretty good measurement, 105
- prior information, 390
- probabilistic operations, 95
- QE, 280
- QIOPA, 336
- QO, 301
- QST, 240, 337
- quadrature, 13, 408
  - bins, 205
  - histogram, 198
  - measurement, 408
  - operator, 195
  - phase-dependent, 267
  - phase-independent, 261
- quadrature amplitude, 243, 248, 251, 261, 271, 285, 290
- quadrature fluctuations, 241
- quantum Bayesian, 152
- quantum channel, 91
- quantum cloning machine, 336
- quantum computation, 511
- quantum computer, 89
- quantum dense coding, 470
- quantum efficiency, 15, 18, 249
- quantum filtering, 437, 440, 441, 456–459
- quantum gates, 89
- quantum harmonic oscillator, 193
- quantum homodyne tomography, 31, 195, 322, 408
  - errors, 412
  - FBP, 410
  - ML, 409, 412
  - of a field displacement, 323
  - of an On/Off photo-detector, 327
- quantum hypothesis testing, 442
- quantum information, 469, 511
- quantum key distribution, 439, 469, 511
- quantum operation, 89, 301, 362
  - antiunitary, 358
  - exchangeable, 178
  - pure, 306
  - purification, 306
  - separable, 464
  - spin-flip, 335
  - universal-NOT, 335
  - universality, 340, 341, 343, 348, 362
- quantum state
  - $m$ -photon-polarization, 455
  - Bayesian view of, 151
  - Bell, 125, 126
  - coherent, 194, 251, 252, 256, 257, 283–285, 288
  - even coherent, 194, 207, 211
  - exchangeable, 165
  - Fock, 222, 261
  - Gaussian, 195
  - Greenberger-Horne-Zeilinger state, 166
  - maximally entangled, 126
  - meaning of, 151, 239
  - mixed, 118

- motional, 214, 219, 220
- multimode coherent, 274
- multiparticle, 339, 461, 464
- odd Fock, 207, 211
- of angular momentum, 269
- of polarization, 269
- one-photon, 227
- positivity, 409, 412
- prior, 245
- pure, 118
- quadrature squeezed, 257
- sampling, 409
- Schrödinger-cat, 339, 342
- squeezed, 214, 222, 255, 260
- squeezed vacuum, 197, 223, 260
- teleportation, 337
- thermal, 264
- two-mode, 267, 269
- two-qubit, 460, 461, 463, 464
- uniformly mixed, 458, 460
- vacuum, 280
- Werner, 125
- quantum state comparison, 438, 447, 461, 511
- quantum state discrimination, 82, 104, 469
- quarter-wave plate, 499
- quasi-probability, 194
- qubit, 472
  - massive, 339
- quinary states, 472, 501, 503, 505
- quorum, 192, 248
- qutrit, 404, 406
- Ramsey interferometry, 226
- reconstruction
  - deterministic, 245
  - fidelity, 210
  - mesh, 386, 397
  - ML compared to FBP, 392
  - nondeterministic, 245
  - two-mode, 270
  - Wigner function, 409
- refraction
  - complex index, 393
  - index, 397
- relative entropy, 82
- representation
  - quadrature, 255
- Schwinger, 377
- RF, 241
- scattering, 385, 397
- SDP, 106
- second-order coherence, 263, 268
- secret sharing, 463
- semidefinite programming, 106
- septenary states, 472, 501, 505
- Shannon entropy, 489
- signal
  - discrete, 378
  - Gaussian, 378
  - Poissonian, 379, 398
- SLD, 264
- SO(3), 344
- source
  - coherent, 378
  - Poissonian, 387
  - single-photon, 511
  - thermal neutrons, 391
- SPDC, 337
- spontaneous parametric down conversion, 337
- SQCG, 505
- Stern-Gerlach apparatus, 102
- Stokes parameters, 120, 122–124, 127–130, 132–136, 142, 145
- SU(1,1), 339
- SU(2), 270, 340, 341
- subset discrimination, 436, 440, 456
- success rate
  - relative, 109
  - renormalized, 109
- superadditive quantum coding gain, 505–508, 510
- superselection rules, 101
- symmetric states, 435, 450, 451, 453–455, 457, 460, 500, 501
- system
  - higher dimensions, 401
  - two-level, 378
- tangle, 126, 147
- Tele-UNOT, 358
- teleportation of gates, 358
- temporal-spatial mode, 250
- tetrad states, 452, 472, 496, 498, 499
- tomography, 240, 254

- absorption, 384
- ML, 387, 390
- Poissonian, 388
- classical imaging, 23
- homodyne tomography, 31
  - adaptive, 39
  - multimode, 40
- kernel function, 30
- noise deconvolution, 32, 37
- null estimator, 30, 39
- Pauli tomography, 363, 364, 366, 368
- phase contrast, 393
- quantum, 25
- quorum for, 28
- spin system, 36
- transformation
  - Bernoulli, 411
  - deterministic, 299
  - discrete Fourier, 379
  - Hilbert, 198
  - input-output, 377
- inverse Radon, 197, 389, 395, 409
- inverse Radon, problems, 197
- of polarization, 270
- probabilistic, 299
- Radon, 24
- trine states, 451, 452, 458, 472, 491, 492, 494, 503, 505–507, 511
- twin-beam, 41
- U-NOT, 336
- unambiguous state discrimination, 108, 445, 447, 448, 452, 455, 458, 471, 478, 479, 493, 511
- uncertainty product, 257
- uninformative prior, 155
- universal NOT gate, 336
- universality, 336
- unknown
  - quantum operation, 176
  - quantum state, 152
- UOQCM, 336
- WF, 194
- Wootters statistical distance, 82