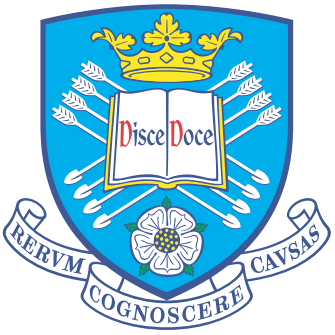

The University Of Sheffield.

Thesis submitted for the degree of

Doctor of Philosophy

Department of Physics and Astromony

Classical and Quantum Imaging and Metrology Using Far Field Radiation

---

Mark Edward Pearce

November 16, 2015

Abstract

In this thesis I examine how the statistical properties of radiation limit our ability to perform imaging and metrological procedures. In particular I focus on radiation in the far field zone of the source. The classical and quantum theories of parameter estimation are introduced and subsequently utilised throughout, along with the theory of optical coherence. Classical and quantum imaging protocols are examined with the aid of a resolution criterion and the criterion is shown to reproduce the results of previous works. This method is also extended to previously un-investigated situations and the effect of imperfect measurements is explored. Intensity correlation measurements are investigated in great detail and for the first time a rigorous comparison is made between higher-order intensity correlation measurements of the type introduced by Hanbury Brown and Twiss. The importance of considering covariances in intensity correlation data is demonstrated and I give a full, detailed account of how to include this in the formulation. I also show how the optimal arrangement for an intensity correlation measurement can be found, therefore allowing the best precision in parameter estimation to be achieved. A quantum mechanical description of blackbody radiation is used to examine the state arriving at a detector in the far field. By using the quantum Fisher information an interesting connection is found between the statistical independence of photons in the source plane and the acquisition of information in the far field.

To William Alfred Pearce

Declaration of Authorship

I, Mark Edward Pearce, declare that the work presented in this thesis is my own research except where otherwise stated and has not been previously submitted for a degree in this or another university. Parts of the work submitted in this thesis have been published as follows:

Publications

1. Carlos Pérez-Delgado, Mark E. Pearce and Pieter Kok, *Fundamental Limits of Classical and Quantum Imaging*, Physical Review Letters **109**(12), 123601, 21 September 2012.

2. Mark E. Pearce, Thomas Mehringer, Joachim von Zanthier and Pieter Kok, *Precision Estimation of Source Dimensions from Higher-Order Intensity Correlations*, Physical Review A **92**, 043831, 21 October 2015.

Signed: _____

Dated: _____

Acknowledgements

Firstly, I would like to thank my supervisor, Pieter Kok for continued support and encouragement throughout my university education and research career. He has been a fantastic teacher and a truly inspirational influence. I certainly could not have completed this research without his guidance.

I would like to thank my friends and colleagues with whom I have shared an office whilst working towards my PhD; Carl Whitfield, Emiliano Cancellieri, Michael Woodhouse, Ian Estabrook, Jasminder Sidhu, Dominic Hostler, Earl Campbell, Giuseppe Buonaiuto, and Tom Bullock. They have been a great source of inspiration and it has been a pleasure to work alongside so many talented physicists. A special thank you is reserved for Samuel Coveney for teaching me the CUDA programming language, which proved essential in my research and I am therefore hugely grateful. Thanks is also due to Nigel Clarke for allowing me to use his computing equipment, without which I could not have performed my research.

I would also like to extend a huge thank you to my family and friends who have supported and encouraged me throughout my time at university. I am grateful to my parents, Angela and Ed, and my sister Kate, who have been both loving and supportive throughout my studies. My partner Helen is also due a special thank you, her patience and understanding have been crucial in me finishing this work.

I am thankful to Prof. Joachim von Zanthier, Steffen Oppel, and Thomas Mehringer for their ongoing collaboration. Their work has been instrumental in the development of my ideas and their hospitality during our visit to Erlangen was greatly appreciated.

Contents

Chapter 1

Introduction

Imaging plays a fundamental role in science and technology. Historically, every time a new imaging technique was introduced, science has leaped forward. For example, recent imaging applications include exoplanet detection [Macintosh et al., 2014] and the velocity measurement of molecular markers along DNA [Heller et al., 2013]. When examining imaging instruments it is of particular importance that we can characterise their resolution. The resolution of an imaging procedure quantifies the finest details we can distinguish on the object we are imaging. The wave nature of light dictates that there are physical limits to the resolution of optical microscopes and telescopes. In order to see smaller details in microscopy we could illuminate with light of shorter and shorter wavelengths, but this is not always practical: increasingly energetic light may destroy biological samples, and in astronomy the accessible wavelengths are beyond our control. We therefore need to find alternative techniques to improve the resolution of our imaging methods that overcome the diffraction limit. The first attempts to determine the resolution of an imaging procedure were given by Rayleigh [Rayleigh, 1879] and Abbe [Abbe, 1873]. Both Abbe and Rayleigh defined resolution phenomenologically as the ability for observers to distinguish two overlapping intensity distributions. By defining resolution in this way, taking into account only the expectation value of the intensity across the image plane, the statistical nature of the measured intensity is ignored. Since statistical fluctuations are present in almost all physical systems, a full characterisation of the resolution of an imaging system must take these effects into account. Imaging techniques that yield finer detail than that dictated by the Rayleigh and Abbe limits are referred to as *super-resolving* techniques. In microscopy techniques such as photo-activated localised microscopy (PALM) [Hess et al., 2006], stochastic optical reconstruction microscopy (STORM) [Rust et al., 2006] and stimulated-emission depletion microscopy (STED) [Hell and Wichmann, 1994] achieve super-resolved images using fluorescent markers. These methods combine standard intensity measurements with prior information about the sample preparation and post-processing to achieve super-resolution. In the year 2000 Pendry developed a new theoretical method of super-resolution that involved using materials of negative refractive index as lenses [Pendry, 2000]. In this paper, Pendry shows that a lens made of a material with a refractive index of -1 is theoretically capable of producing perfect images that are not limited by diffraction. In practice, materials that exhibit refractive indexes of -1 are difficult to manufacture. They also tend to suffer from high dispersion, which limits there resolving power in practice. In this thesis we

use the rigorous method of parameter estimation theory to define the resolution of the systems we investigate. The success of classical theories in describing optical phenomena meant that the introduction of quantum mechanical methods was slow to be adopted by the optical community. Eventually though, the quantum nature of light was subtly revealed and a new branch of quantum physics was established.

The development of quantum optics began in the late 1950's, sparked by the discovery of a new type of interferometer by Hanbury Brown and Twiss [Brown and Twiss, 1956a, Brown and Twiss, 1956b, Brown and Twiss, 1957, Brown and Twiss, 1958a, Brown and Twiss, 1958b, Brown and Twiss, 1958c]. They called this new type of interferometer an *intensity interferometer*, and initially it appeared that there was a discrepancy between the classical and quantum theories describing the workings of the device. The classical description given by Hanbury Brown and Twiss correctly predicted the experimental observations [Brown and Twiss, 1956b] whilst others suggested that if their experimental results were correct, the quantum theory of the photon would require elaboration [Brannen and Ferguson, 1956]. Purcell was one of few researchers who strongly believed that the effect would adhere to a quantum description and was the first to give a successful quantum analysis of the effect in terms of the "clumping" of bosons [Purcell, 1956, Silva and Freire Jr, 2013].

Although the effect discovered by Hanbury Brown and Twiss was successfully explained by a completely classical theory, the debate surrounding the validity of their results compelled researchers to develop the currently existing classical discipline of optics into a fully fledged quantum theory [Gerry and Knight, 2006, Walls and Milburn, 2008, Glauber, 2007]. Among the developments spurred by the experiments of Hanbury Brown and Twiss were the seminal works of Glauber on the quantum theory of optical coherence [Glauber, 1963a, Glauber, 1963c, Glauber, 1963b]. Glauber's theory could successfully explain the effects observed in the experiments of Hanbury Brown and Twiss and could also explain the coherence phenomena that were observed in all optics experiments prior to this. The modern theory of optical coherence as pioneered by Glauber, characterises the phenomenon with a set of functions indexed by a pair of positive integers, (n, m). A complete description of the optical field can only be given if the functions are known for all pairs of indices. Before the discovery of intensity interferometry by Hanbury Brown and Twiss (HBT) all optical experiments that relied on coherence exploited only first order coherences, that is, coherences that can be explained by the lowest order coherence function, $n = m = 1$.

Despite the discovery of higher order coherences taking place over fifty years ago, relatively little work has been performed that requires the use of coherence functions beyond the first order. In recent years research that exploits the higher order coherences of optical fields has become more prevalent. In the field of quantum computation the problem of boson sampling is one such example [Aaronson and Arkhipov, 2010, Aaronson and Arkhipov, 2013]. The boson sampling problem requires the calculation of high order photon coincidences and therefore relies on the theory of higher order coherences. It is easy to show that the output of an m port interferometer, interfering n bosons, is inefficient to simulate on a classical computer [Gard et al., 2014b, Motes et al., 2013]. It is unclear whether the presence of noise will affect this result making the distribution efficiently classically simulated [Gogolin et al., 2013, Aaronson and Arkhipov, 2013]. It is for this reason that researchers in this area anticipate that a boson sampling machine may be useful in

implementing quantum computational tasks. However, to date no known use of a boson sampling machine exists [Gard et al., 2014a]. [1]

In 1995 Pittman *et al.*, inspired by the work of Klyshko [Klyshko, 1988], developed an imaging technique that would later become known as *ghost imaging* [Pittman et al., 1995]. Their technique uses a pair of correlated photons, produced via spontaneous parametric down-conversion, one of which is directed towards an object and then recorded by a detector that does not spatially resolve the presence of the photon, the other is not directed towards the object but is measured by a spatially resolving detector. By measuring the coincidence counts of the two photons, a spatially resolved image of the object is obtained despite the spatial information only being obtained for the photon that did not interact with the object. The photons produced by spontaneous parametric downconversion are highly entangled and therefore exhibit a strong type of correlation that can only be observed in quantum systems. Classical systems can also exhibit correlations, and pairs of classically correlated photons can also be used to perform ghost imaging [Bennink et al., 2002, Ferri et al., 2005, Gatti et al., 2006]. Generally, the price paid for using the weaker correlations present in classical systems instead of the stronger quantum correlations, is a decrease in the visibility of the ghost image [Gatti et al., 2007]. Since ghost imaging measures the two photon coincidence rate, it is closely related to the measurements of Hanbury Brown and Twiss and again requires the use of higher order coherence theory to correctly predict the outcome. More recently, researchers have started to examine generalisations of the ghost imaging protocol in which an $m+n$ coincidence is measured, m being the number of non-spatially resolving coincidences measured and n being the number of spatially resolving detectors [Agafonov et al., 2009a, Agafonov et al., 2009b, Chan et al., 2009, Chen et al., 2010, Liu et al., 2009]. The use of higher order coincidences can lead to an increase in the visibility of the ghost image. However, generally the signal to noise ratio decreases as the coincidence order is increased.

Higher order correlations can also be exploited in generalisations of the original Hanbury Brown and Twiss experiments [Cao et al., 2008, Agafonov et al., 2008a, Liu and Shih, 2009, Zhou et al., 2010]. Hanbury Brown and Twiss (HBT) in their seminal experiments demonstrated that the second order intensity correlation function is proportional to the Fourier transform of the intensity distribution of the source [Brown and Twiss, 1956a]. More generally we can measure the coincidences between n photo-detectors, which also conveys information about the Fourier transform of the source. We call this kind of measurement an n^{th}-order intensity correlation measurement since it is implemented by measuring the correlation between n intensities. The visibility of the measurements can also increase by going to higher correlation orders but just as in the ghost imaging regime the signal to noise ratio suffers as a result. Until now it was unclear what, if any, advantage higher order correlations could offer in measurements of this type. In this thesis we explicitly calculate the performance of estimators in the Hanbury Brown-Twiss (HBT) setup and determine how different correlation orders perform in a parameter estimation

[1] Shortly after the submission of this thesis, Huh *et al.* published a proposal for simulating molecular vibronic spectra using boson sampling [Huh et al., 2015]. The proposed scheme uses a generalised version of boson sampling whereby the input states are squeezed coherent and squeezed vacuum states and allows the vibonic spectra of molecules to be determined by the output of a linear optical network. No known classical algorithms can efficiently predict these spectra.

similar to the original HBT measurement. To ensure the optimal performance of such techniques it is necessary to obtain the best estimates possible.

Measurements of the HBT type rely on the phenomenon of photon bunching, an effect caused by the bosonic nature of photons. Although it has been shown that the effect of bunching can occur for statistically independent photons originating from two independent atoms [Fano, 1961], we demonstrate in this thesis that imaging of thermal sources relies on the correlations present between the photons in the source plane. In fact, we find in that in general the statistical independence of the photons prevents us from performing the most general parameter estimation task and therefore cannot be the source of information in HBT type measurements.

This thesis is composed of six chapters. In chapter 2 we give the theoretical background to the various techniques of quantum optics that shall be used throughout this thesis. In particular we focus on the theory of optical coherence as developed by Glauber and we also demonstrate how propagation affects the coherence properties of optical fields.

In chapter 3 we give an overview of the field of parameter estimation in both the classical and quantum regime. Particular attention is paid to the Cramér-Rao bound, which is used extensively throughout the remainder of this thesis.

Chapter 4 investigates fundamental limitations placed upon imaging protocols by considering images as probability distributions to be discriminated. In this chapter we also determine the effects of detector imperfections on the imaging performance and show how the effect of any imperfection can be calculated. We end this chapter by examining multi-photon correlation experiments, enquiring as to their potential in increasing resolution for imaging.

In chapter 5 we explicitly perform a parameter estimation procedure making use of data that simulates n^{th}-order intensity correlation measurements. We make comparisons of the estimation precision for different correlation orders and determine how we can optimise an experiment to achieve the highest precision.

In chapter 6 we use the quantum mechanical description of a source emitting black body radiation to determine the way in which information regarding the spatial distribution of the source is conveyed to the far field. We show how these results can in principle lead to a complete calculation of the quantum Fisher information.

Chapter 7 summarises the main results of the thesis and gives the direction of future research leading on from this work.

Chapter 2

Quantum Optics

In this chapter we introduce a fully quantum mechanical description of the electromagnetic field and discuss some of the states of this field to be used in the forthcoming chapters. We then discuss the concept of optical coherence, first classically and then again using the quantum mechanical formalism of Glauber. Finally we demonstrate how correlations in the field propagate, obtaining expressions for the coherence properties of optical fields far from the radiation sources.

2.1 The quantum electromagnetic field

Classically, the electric field inside a volume of side lengths L_x, L_y, L_z may only exist in certain allowable vibrational modes, known as the modes of the cavity. This restriction results from the necessity for the electric field to vanish at the edges of the cavity. We therefore require that any valid electric field inside the cavity fulfils this condition. From Fourier analysis we know that any such function can be expressed as

$$E(\mathbf{r},t) = \sum_{l=0}^{\infty}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left[\mathcal{E}_{lnm}a_{lnm}u_{lnm}(\mathbf{r},t) + \mathcal{E}^*_{lnm}a^*_{lnm}u^*_{lnm}(\mathbf{r},t)\right], \tag{2.1}$$

where $u_{lnm}(\mathbf{r},t)$ are the mode functions describing a complete orthonormal set of functions for the cavity, a_{lnm} are the complex amplitudes of each mode, and $\mathcal{E}_{lnm}$ are chosen such that the amplitudes are dimensionless. The field described by Eq. (2.1) is a scalar field and the indices l, n, and m can each be associated with an orthogonal spatial component of the field x, y, and z respectively, see Fig. 2.1.

We can impose a vector character on the field by including an additional spatial degree of freedom that describes the components of the vector field in each orthogonal spatial direction. The field then becomes [Kok and Lovett, 2010]

$$\mathbf{E}(\mathbf{r},t) = \sum_{s}\sum_{\mathbf{k}}\left[\boldsymbol{\mathcal{E}}_{s\mathbf{k}}a_{s\mathbf{k}}u_{\mathbf{k}}(\mathbf{r},t) + \boldsymbol{\mathcal{E}}^*_{s\mathbf{k}}a^*_{s\mathbf{k}}u^*_{\mathbf{k}}(\mathbf{r},t)\right], \tag{2.2}$$

where the sum over s accounts for the independent vector components and the sum over l, n, m is now expressed as a sum over the wavevector $\mathbf{k}$ with components

$$k_x = \frac{2\pi l}{L_x}, \quad k_y = \frac{2\pi n}{L_y}, \quad k_z = \frac{2\pi m}{L_z}. \tag{2.3}$$

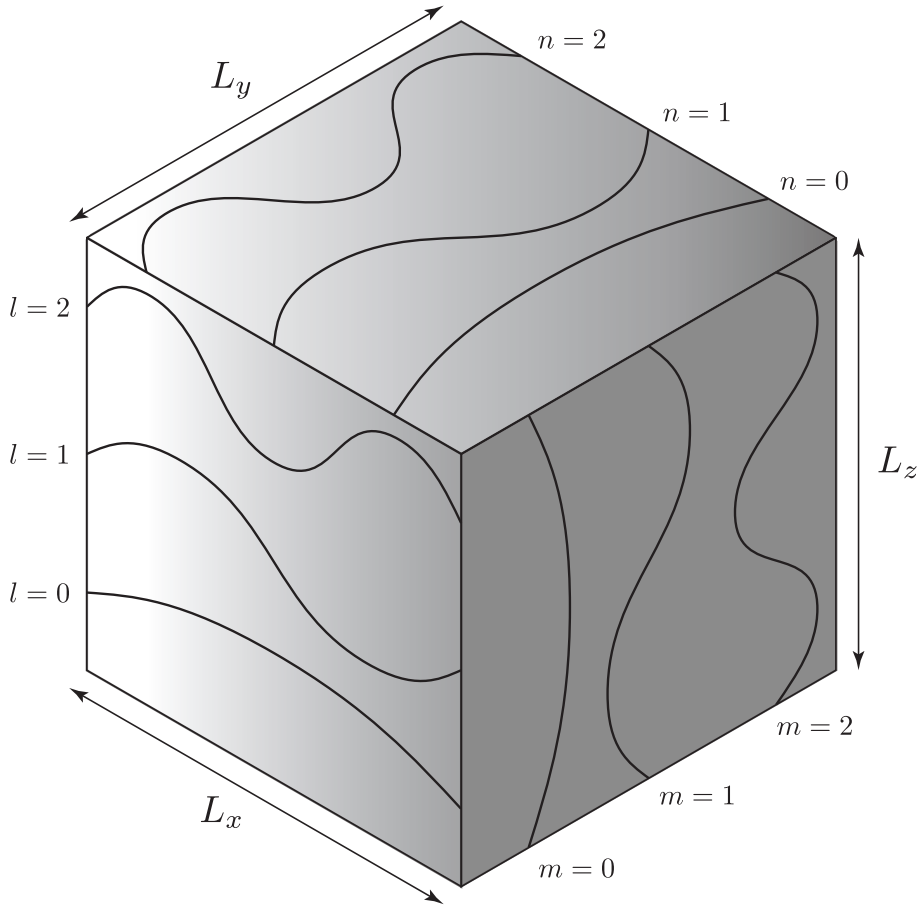

Figure 2.1: The allowable electromagnetic vibrational modes inside a cavity of side lengths L_x, L_y, and L_z. The integers l, n, and m label the modes in the perpendicular x, y, and z directions respectively.

The vectorial character of the field $\mathbf{E}$ is incorporated into the term $\boldsymbol{\mathcal{E}}_{\lambda\mathbf{k}}$. The field $\mathbf{E}$ is now a classical vector field that obeys Maxwell's equations and the wave equation in free space and in the absence of sources [Jackson, 1999]

$$\nabla^2\mathbf{E}(\mathbf{r},t) - \frac{1}{c^2}\frac{\partial^2\mathbf{E}(\mathbf{r},t)}{\partial t^2} = 0. \tag{2.4}$$

where $\nabla^2 = \left(\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial y^2}, \frac{\partial^2}{\partial z^2}\right)$ is the Laplace operator.

In order to quantise the field $\mathbf{E}$, the mode amplitudes $a_{s\mathbf{k}}$, which have so far been complex numbers, now become operators. We denote this by the symbol $\hat{\cdot}$. The quantised electric field is now given by

$$\hat{\mathbf{E}}(\mathbf{r},t) = \sum_s \sum_{\mathbf{k}} \left[\boldsymbol{\mathcal{E}}_{s\mathbf{k}}\hat{a}_{s\mathbf{k}}u_{\mathbf{k}}(\mathbf{r},t) + \boldsymbol{\mathcal{E}}^*_{s\mathbf{k}}\hat{a}^\dagger_{s\mathbf{k}}u^*_{\mathbf{k}}(\mathbf{r},t)\right], \tag{2.5}$$

where $\hat{a}^\dagger_{s\mathbf{k}}$ is the Hermitian conjugate of the operator $\hat{a}_{s\mathbf{k}}$. The operators $\hat{a}^\dagger_{s\mathbf{k}}$ and $\hat{a}_{s\mathbf{k}}$ are the mode operators of the field and, as is implicit in Eq. (2.5), they are not Hermitian. The electric field is a bosonic field and therefore the field operators $\hat{a}^\dagger_{s\mathbf{k}}$, $\hat{a}_{s\mathbf{k}}$ obey the bosonic commutation relations [Walls and Milburn, 2008]

$$\left[\hat{a}_{s\mathbf{k}}, \hat{a}^\dagger_{s'\mathbf{k}'}\right] = \delta_{s,s'}\delta_{\mathbf{k},\mathbf{k}'}, \tag{2.6}$$

and

$$\left[\hat{a}^\dagger_{s\mathbf{k}}, \hat{a}^\dagger_{s'\mathbf{k}'}\right] = [\hat{a}_{s\mathbf{k}}, \hat{a}_{s'\mathbf{k}'}] = 0. \tag{2.7}$$

For convenience we shall write the electric field $\mathbf{E}$ as the sum of two terms

$$\hat{\mathbf{E}}(\mathbf{r},t) = \hat{\mathbf{E}}^{(+)}(\mathbf{r},t) + \hat{\mathbf{E}}^{(-)}(\mathbf{r},t), \tag{2.8}$$

where

$$\hat{\mathbf{E}}^{(+)}(\mathbf{r},t) = \sum_s \sum_{\mathbf{k}} \boldsymbol{\mathcal{E}}_{s\mathbf{k}}\hat{a}_{s\mathbf{k}}u_{\mathbf{k}}(\mathbf{r},t), \tag{2.9}$$

and $\hat{\mathbf{E}}^{(-)} = (\hat{\mathbf{E}}^{(+)})^\dagger$. Combining Eqs. (2.4) and (2.5) we see that the mode functions themselves obey the wave equation and are therefore completely determined by the classical Maxwell theory. For the cavity previously defined, the appropriate mode functions are the plane wave modes [Walls and Milburn, 2008]

$$u_{\mathbf{k}}(\mathbf{r},t) = \frac{1}{\sqrt{V}}\mathrm{e}^{i(\mathbf{k}\cdot\mathbf{r}-\omega_{\mathbf{k}}t)}, \tag{2.10}$$

where $\omega_{\mathbf{k}} = c|\mathbf{k}|$ and $V = L_xL_yL_z$. The mode functions form a complete orthonormal set, which can be seen by taking

$$\begin{aligned}\int_V \mathrm{d}\mathbf{r}\, u^*_{\mathbf{k}}(\mathbf{r},t)u_{\mathbf{k}'}(\mathbf{r},t) &= \frac{1}{V}\int_V \mathrm{d}^3\mathbf{r}\,\mathrm{e}^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}-i(\omega_{\mathbf{k}}-\omega_{\mathbf{k}'})t} \\ &= \delta^3(\mathbf{k}-\mathbf{k}').\end{aligned} \tag{2.11}$$

The mode operators play an important role in quantum optics. For reasons that will become apparent, the operators $\hat{a}^\dagger$ and $\hat{a}$ are referred to as the creation

and annihilation operators, respectively. From Eq. (2.9) and the relation $\hat{\mathbf{E}}^{(-)} = (\hat{\mathbf{E}}^{(+)})^\dagger$, we see that the operators $\hat{\mathbf{E}}^{(-)}$ and $\hat{\mathbf{E}}^{(+)}$ are proportional to the creation and annihilation operators respectively.

Taking the limit $L_x, L_y, L_z \to \infty$, the plane wave mode functions extend over all of space-time and are therefore unphysical as a description of an observable state. We can use superpositions of the plane wave modes to effectively "localise" the electric field modes to regions of space-time. These localised modes are referred to as wave packets and can also form complete orthonormal sets.

2.2 States of the electromagnetic field

We now present some of the more commonly used bases to describe the state of the electromagnetic field and discuss some of their properties. As with all quantum systems, any basis may be used to express the state. However, certain problems become much simpler when expressed in a particular choice of basis. The procedure of quantisation has the effect of discretising the electric field. As such, the field cannot be incrementally increased and can only increase by certain finite amounts. These amounts are the *quanta* of the field, in the case of the electromagnetic field these quanta are known as photons. A convenient basis that reflects this quantisation is the Fock basis, also referred to as the number basis. Classically, the Hamiltonian density is given by [Kok and Lovett, 2010]

$$\mathcal{H} = \frac{\varepsilon_0}{2}\mathbf{E}\cdot\mathbf{E} + \frac{1}{2\mu_0}\mathbf{B}\cdot\mathbf{B}, \tag{2.12}$$

which gives the total energy of the electromagnetic field. By substituting the quantum mechanical form for the electric field, Eq. (2.5), making use of the relation $\hat{\mathbf{B}} = \frac{1}{\omega}\mathbf{k}\times\hat{\mathbf{E}}$, and taking the integral over the entire volume V, we arrive at the total energy, or Hamiltonian operator,

$$\begin{aligned}\hat{H} &= \int_V \left[\frac{\varepsilon_0}{2}\hat{\mathbf{E}}^2 + \frac{1}{2\mu_0}\hat{\mathbf{B}}^2\right] \mathrm{d}^3\mathbf{r} \\ &= \sum_s \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} \left[\hat{a}^\dagger_{s\mathbf{k}}\hat{a}_{s\mathbf{k}} + \frac{1}{2}\right].\end{aligned} \tag{2.13}$$

To elucidate the role of the operators $\hat{a}^\dagger_{s\mathbf{k}}$ and $\hat{a}_{s\mathbf{k}}$ we write the eigenvalue equation for the Hamiltonian as $\hat{H}|n\rangle = E_n|n\rangle$ and then evaluate

$$\begin{aligned}\hat{H}\hat{a}^\dagger_{s'\mathbf{k}'}|n\rangle &= \sum_s\sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} \left[\hat{a}^\dagger_{s\mathbf{k}}\hat{a}_{s\mathbf{k}} + \frac{1}{2}\right]\hat{a}^\dagger_{s'\mathbf{k}'}|n\rangle \\ &= \sum_s\sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} \left[\hat{a}^\dagger_{s\mathbf{k}}(\delta_{s,s'}\delta_{\mathbf{k},\mathbf{k}'} + \hat{a}^\dagger_{s'\mathbf{k}'}\hat{a}_{s\mathbf{k}}) + \frac{\hat{a}^\dagger_{s'\mathbf{k}'}}{2}\right]|n\rangle \\ &= \hat{a}^\dagger_{s'\mathbf{k}'}\left[\hbar\omega_{\mathbf{k}'} + \sum_s\sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}}\left(\hat{a}^\dagger_{s\mathbf{k}}\hat{a}_{s\mathbf{k}} + \frac{1}{2}\right)\right]|n\rangle \\ &= (\hbar\omega_{\mathbf{k}'} + E_n)\hat{a}^\dagger_{s'\mathbf{k}'}|n\rangle,\end{aligned} \tag{2.14}$$

where we have used Eqs. (2.6) and (2.7). Eq. (2.14) demonstrates that the state $\hat{a}^\dagger_{s'\mathbf{k}'}|n\rangle$ is also an eigenstate of the Hamiltonian with eigenvalue $E_n + \hbar\omega_{\mathbf{k}'}$. Similarly,

we find that the state $\hat{a}_{s'\mathbf{k}'}|n\rangle$ is an eigenstate of the Hamiltonian with eigenvalue $E_n - \hbar\omega_{\mathbf{k}'}$. It is for this reason that the operators $\hat{a}^\dagger_{s'\mathbf{k}'}$ and $\hat{a}_{s'\mathbf{k}'}$ are referred to as the creation and annihilation operators, since they fulfil the role of adding and subtracting an amount of energy equal to $\hbar\omega_{\mathbf{k}'}$ from the system. This is the basis of the quantum theory of light.

We notice from Eq. (2.13) that the total energy is the sum of operators of the form $\hbar\omega_{\mathbf{k}}[\hat{a}^\dagger_{s\mathbf{k}}\hat{a}_{s\mathbf{k}} + 1/2]$. The total energy is therefore the sum of the energy for each mode. Concentrating on a single mode, we can drop the subscripts $\mathbf{k}s$ and write the energy of a single mode as $\hbar\omega[\hat{a}^\dagger\hat{a} + 1/2]$. We now see that the operator $\hat{a}^\dagger\hat{a}$ is proportional to the single mode energy operator. We therefore expect the eigenstates to be the same for both operators. We write

$$\hat{a}^\dagger\hat{a}|n\rangle = n|n\rangle, \tag{2.15}$$

multiplying by $\hbar\omega$ and adding $\frac{1}{2}\hbar\omega|n\rangle$ we find

$$\hbar\omega\hat{a}^\dagger\hat{a}|n\rangle + \frac{1}{2}\hbar\omega|n\rangle = \hbar\omega\left(n + \frac{1}{2}\right)|n\rangle. \tag{2.16}$$

The eigenvalue of the operator $\hat{a}^\dagger\hat{a}$ is therefore equal to the number of photons in the state. We therefore refer to this operator as the number operator and often write it as $\hat{a}^\dagger_{s\mathbf{k}}\hat{a}_{s\mathbf{k}} = \hat{n}_{s\mathbf{k}}$. The eigenstates $|n\rangle$ form a complete orthonormal set

$$\langle n|m\rangle = \delta_{n,m}. \tag{2.17}$$

The effect of creation and annihilation operators on the state $|n\rangle$ can now be determined. Since the annihilation operator destroys a photon, the action of the annihilation operator on the state $|n\rangle$ will be [Gerry and Knight, 2006]

$$\begin{aligned} \hat{a}|n\rangle &\propto |n-1\rangle \\ \Rightarrow \hat{a}|n\rangle &= k_n|n-1\rangle. \end{aligned} \tag{2.18}$$

To determine the constant of proportionality k_n, which is potentially complex due to the non-Hermitian nature of $\hat{a}$, we multiply Eq. (2.18) on the left by its complex conjugate

$$(\hat{a}|n\rangle)^\dagger\hat{a}|n\rangle = \langle n|\hat{a}^\dagger\hat{a}|n\rangle = \langle n|\hat{n}|n\rangle = n\langle n|n\rangle = n = |k_n|^2. \tag{2.19}$$

We therefore have the freedom to choose the phase of k_n and the usual choice is $k_n = \sqrt{n}$. From the above relation $\langle n|\hat{a}^\dagger\hat{a}|n\rangle = n$ we see that n is positive. We also find the self-evident relation $\hat{a}|0\rangle = 0$, which asserts the impossibility of having states with negative photon number. By performing the same calculation as Eq. (2.19) with $\hat{a}^\dagger|n\rangle$ we arrive at the pair of relations

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle \tag{2.20}$$

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle. \tag{2.21}$$

And we see explicitly that the operators $\hat{a}^\dagger$ and $\hat{a}$ play the role of creating and annihilating photons.

Another important state of the electromagnetic field is the thermal state. Unlike the number states, the thermal state is not a state of definite photon number. The

thermal state is also not a pure state and therefore cannot be described by a state vector $|\psi\rangle$ but must instead be written as a density operator ρ. Density operators describe mixed states, which are quantum states with classical statistical uncertainties as opposed to purely quantum uncertainty. To demonstrate the necessity for the density operator formalism we consider the expectation value of an operator $\hat{O}$

$$\langle\hat{O}\rangle = \langle\psi|\hat{O}|\psi\rangle = \sum_{i=1}^{n}\langle\psi|i\rangle\langle i|\hat{O}|\psi\rangle = \sum_{i=1}^{n}\langle i|\hat{O}|\psi\rangle\langle\psi|i\rangle = \mathrm{Tr}[\hat{O}|\psi\rangle\langle\psi|], \tag{2.22}$$

where we have used the complete orthonormal basis $|i\rangle$ and $\mathrm{Tr}[\cdot]$ denotes the trace operation. The statistical uncertainty in our state manifests itself as a statistical uncertainty of the expectation value. Moreover, if the state is a statistical mixture of states $|\psi_j\rangle$ with probability p_j, the expectation value will be

$$\langle\hat{O}\rangle = \sum_j p_j \mathrm{Tr}[\hat{O}|\psi_j\rangle\langle\psi_j|] = \mathrm{Tr}\left[\sum_j p_j|\psi_j\rangle\langle\psi_j|\hat{O}\right] \equiv \mathrm{Tr}[\rho\hat{O}], \tag{2.23}$$

where we have used the linearity of the trace operation and defined $\rho = \sum_j p_j|j\rangle\langle j|$. Eq. (2.23) shows how expectation values are calculated using the density operator. The density operator for any physical state must satisfy the following three conditions [Holevo, 2011]

$$\mathrm{Tr}[\rho] = 1, \tag{2.24a}$$
$$\rho^\dagger = \rho, \tag{2.24b}$$
$$\rho \geq 0. \tag{2.24c}$$

For a state in thermal equilibrium, the density operator is given by [Fano, 1957]

$$\rho = \frac{\exp\left(-\beta\hat{H}\right)}{\mathrm{Tr}\left[\exp\left(-\beta\hat{H}\right)\right]}, \tag{2.25}$$

where $\beta = 1/k_B T$. Using Eq. (2.13) we find

$$\rho = \frac{\exp\left(-\sum_{s,\mathbf{k}}\beta\hbar\omega_{\mathbf{k}}(\hat{n}_{s\mathbf{k}}+\frac{1}{2})\right)}{\mathrm{Tr}\left[\exp\left(-\sum_{s',\mathbf{k}'}\beta\hbar\omega_{\mathbf{k}'}(\hat{n}_{s'\mathbf{k}'}+\frac{1}{2})\right)\right]} = \frac{\exp\left(-\sum_{s,\mathbf{k}}\beta\hbar\omega_{\mathbf{k}}\hat{n}_{s\mathbf{k}}\right)}{\mathrm{Tr}\left[\exp\left(-\sum_{s',\mathbf{k}'}\beta\hbar\omega_{\mathbf{k}'}\hat{n}_{s'\mathbf{k}'}\right)\right]}. \tag{2.26}$$

Since $\langle m|f(\hat{n})|n\rangle = f(n)\delta_{n,m}$, any operator that is a function of the number operator is diagonal in the number basis. We therefore evaluate the density operator in the number basis. First we evaluate the denominator, which, being the trace of an

operator (i.e. a number), is the normalisation of the state ρ

$$\begin{aligned}
\mathrm{Tr}\left[\exp\left(-\sum_{s',\mathbf{k}'}\beta\hbar\omega_{\mathbf{k}'}\hat{n}_{s'\mathbf{k}'}\right)\right] &= \prod_{s,\mathbf{k}}\sum_{n_{s\mathbf{k}}}\langle n_{s\mathbf{k}}|\exp\left(-\sum_{s',\mathbf{k}'}\beta\hbar\omega_{\mathbf{k}'}\hat{n}_{s'\mathbf{k}'}\right)|n_{s\mathbf{k}}\rangle \\
&= \prod_{s,\mathbf{k}}\sum_{n_{s\mathbf{k}}}\langle n_{s\mathbf{k}}|\prod_{s',\mathbf{k}'}\exp\left(-\beta\hbar\omega_{\mathbf{k}'}\hat{n}_{s'\mathbf{k}'}\right)|n_{s\mathbf{k}}\rangle \\
&= \prod_{s,\mathbf{k}}\sum_{n_{s\mathbf{k}}}\prod_{s',\mathbf{k}'}\exp\left(-\beta\hbar\omega_{\mathbf{k}'}n_{s'\mathbf{k}'}\right)\delta_{s,s'}\delta_{\mathbf{k},\mathbf{k}'}\langle n_{s\mathbf{k}}|n_{s\mathbf{k}}\rangle \\
&= \prod_{s,\mathbf{k}}\sum_{n_{s\mathbf{k}}}\exp\left(-\beta\hbar\omega_{\mathbf{k}}n_{s\mathbf{k}}\right) \\
&= \prod_{s,\mathbf{k}}\frac{1}{1-\mathrm{e}^{-\beta\hbar\omega_{\mathbf{k}}}}. \qquad (2.27)
\end{aligned}$$

Therefore, the density operator is

$$\begin{aligned}
\rho &= \frac{\otimes_{s,\mathbf{k}}\exp\left(-\beta\hbar\omega_{\mathbf{k}}\hat{n}_{s\mathbf{k}}\right)}{\prod_{s',\mathbf{k}'}(1-\mathrm{e}^{-\beta\hbar\omega_{\mathbf{k}'}})^{-1}} \\
&= \otimes_{s,\mathbf{k}}\sum_{n_{s\mathbf{k}}}(1-\mathrm{e}^{-\beta\hbar\omega_{\mathbf{k}}})\exp\left(-\beta\hbar\omega_{\mathbf{k}}\hat{n}_{s\mathbf{k}}\right)|n_{s\mathbf{k}}\rangle\langle n_{s\mathbf{k}}| \\
&= \otimes_{s,\mathbf{k}}\sum_{n_{s\mathbf{k}}}(1-\mathrm{e}^{-\beta\hbar\omega_{\mathbf{k}}})\exp\left(-\beta\hbar\omega_{\mathbf{k}}n_{s\mathbf{k}}\right)|n_{s\mathbf{k}}\rangle\langle n_{s\mathbf{k}}|, \qquad (2.28)
\end{aligned}$$

from which we see directly that the density operator takes the form, $\rho = \otimes_{s,\mathbf{k}}\rho_{s\mathbf{k}}$. The density operator for a thermal state naturally separates into a product over density operators associated with each mode. Another useful form for ρ comes from evaluating $\langle\hat{n}_{s\mathbf{k}}\rangle$

$$\begin{aligned}
\langle\hat{n}_{s\mathbf{k}}\rangle &= \mathrm{Tr}\left[\sum_{n_{s\mathbf{k}}}(1-\mathrm{e}^{-\beta\hbar\omega_{\mathbf{k}}})\exp\left(-\beta\hbar\omega_{\mathbf{k}}n_{s\mathbf{k}}\right)\hat{n}_{s\mathbf{k}}|n_{s\mathbf{k}}\rangle\langle n_{s\mathbf{k}}|\right] \\
&= \sum_{n_{s\mathbf{k}}}(1-\mathrm{e}^{-\beta\hbar\omega_{\mathbf{k}}})\exp\left(-\beta\hbar\omega_{\mathbf{k}}n_{s\mathbf{k}}\right)n_{s\mathbf{k}} \\
&= \frac{1}{\mathrm{e}^{\beta\hbar\omega_{\mathbf{k}}}-1}. \qquad (2.29)
\end{aligned}$$

We now find

$$\mathrm{e}^{-\beta\hbar\omega_{\mathbf{k}}} = \frac{\langle\hat{n}_{s\mathbf{k}}\rangle}{1+\langle\hat{n}_{s\mathbf{k}}\rangle}, \qquad (2.30)$$

and therefore

$$\rho_{s\mathbf{k}} = \sum_{n}\frac{\langle\hat{n}_{s\mathbf{k}}\rangle^{n}}{(1+\langle\hat{n}_{s\mathbf{k}}\rangle)^{n+1}}|n\rangle\langle n|. \qquad (2.31)$$

It is interesting to note at this point that the density operator ρ depends only on one parameter, namely the temperature T.

For reasons that will become clear later, the thermal state is often described as incoherent radiation. In contrast, there exists another state of the electromagnetic

field known as the coherent state. Coherent states play a particularly important role in quantum optics as they are the eigenstates of the annihilation operator and are also the output states of a laser operating well above threshold [Barnett and Radmore, 1997]. Unlike thermal states, coherent states are pure and therefore do not require the density operator formalism. In the number basis the coherent state is written as

$$|\alpha\rangle = \mathrm{e}^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle, \tag{2.32}$$

where α is a complex number. With the use of Eq. (2.20) this gives

$$\begin{aligned}
\hat{a}|\alpha\rangle &= \mathrm{e}^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} \hat{a}|n\rangle \\
&= \mathrm{e}^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} \sqrt{n}|n-1\rangle \\
&= \mathrm{e}^{-\frac{|\alpha|^2}{2}} \sum_{n=1}^{\infty} \frac{\alpha^n}{\sqrt{(n-1)!}} |n-1\rangle \\
&= \alpha \mathrm{e}^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \\
&= \alpha|\alpha\rangle.
\end{aligned} \tag{2.33}$$

An important feature of the coherent states is the so called over-completeness. We might expect that the generalisation of the completeness relation to the continuous variable α to be $\int |\alpha\rangle\langle\alpha| \, \mathrm{d}^2\alpha = \mathbb{I}$, however, because of the over-completeness of the states $|\alpha\rangle$ this requires modification [Barnett and Radmore, 1997]. Writing $\alpha = r\mathrm{e}^{i\phi}$ and using Eq. (2.32), we find

$$\int |\alpha\rangle\langle\alpha| \, \mathrm{d}^2\alpha = \int_0^{\infty} r \, \mathrm{d}r \int_0^{2\pi} \mathrm{d}\phi \, \mathrm{e}^{-r^2} \sum_{n,m=0}^{\infty} r^{n+m} \frac{\mathrm{e}^{i\phi(n-m)}}{\sqrt{n!m!}} |n\rangle\langle m| \tag{2.34a}$$

$$= 2\pi \int_0^{\infty} r \, \mathrm{d}r \, \mathrm{e}^{-r^2} \sum_{n,m=0}^{\infty} r^{n+m} \frac{\delta_{n,m}}{\sqrt{n!m!}} |n\rangle\langle m| \tag{2.34b}$$

$$= 2\pi \sum_{n=0}^{\infty} \int_0^{\infty} r \, \mathrm{d}r \, \mathrm{e}^{-r^2} \frac{r^{2n}}{n!} |n\rangle\langle n| \tag{2.34c}$$

$$= \pi \sum_{n=0}^{\infty} |n\rangle\langle n| \tag{2.34d}$$

$$= \pi\mathbb{I}, \tag{2.34e}$$

where we have used the definition of the Kronecker delta in Eq. (2.34b) and the standard integral $\int_0^{\infty} \exp(-ar^2) r^n \, \mathrm{d}r = k!/(2a^{k+1})$, $n = 2k+1$, $k \in \mathbb{Z}$, $a > 0$. This proves that

$$\frac{1}{\pi} \int |\alpha\rangle\langle\alpha| \, \mathrm{d}^2\alpha = \mathbb{I}, \tag{2.35}$$

which is a consequence of the over-completeness of the coherent states $|\alpha\rangle$.

Another interesting and useful consequence of the over-completeness is that any traceable operator $\hat{O}$ can be expressed in the coherent state basis using only the matrix elements $\langle\alpha|\hat{O}|\alpha\rangle$ [Jordan, 1964], the over-completeness of $|\alpha\rangle$ allowing it to probe the off diagonal elements of $\hat{O}$. The representation of operators in this form is known as the diagonal coherent state representation or the P-representation. The operator $\hat{O}$ is expressed in terms of the P-representation by

$$\hat{O} = \int P(\alpha)|\alpha\rangle\langle\alpha| \, \mathrm{d}^2\alpha \,, \tag{2.36}$$

where $P(\alpha)$ is a quasi-probability distribution which determines the weighting of the operator $\hat{O}$ across the complex α plane. Using Eq. (2.36) we can express the density operator in the P-representation, in this form evaluation of expectation values is reduced to integration, sometimes simplifying the problem (see appendix A). We will make use of this representation later when evaluating expectation values. We note in passing that the coherent states presented here are part of a broader class of *generalised* coherent states, which can be constructed for an arbitrary Lie algebra, all of which share the over-completeness property [Perelomov, 1986].

2.3 Optical coherence

The concept of optical coherence is closely related to the concept of statistical correlation. When we measure radiation, we typically only come to certain statistical conclusions. Whether the statistical uncertainties come from the inability to characterise the huge number of interactions in the system, such as for thermal light, or whether it is due to fundamental quantum uncertainties, such as trying to measure the coherent state in the Fock basis, these uncertainties are almost always present. We begin by examining optical coherence for classical fields and then develop the theory for quantum fields. The most famous experiment demonstrating optical coherence is Young's double slit experiment. The simplicity of the experimental arrangement allows for a simple yet effective description to be made. Consider Young's arrangement, with two secondary sources of radiation and a screen some distance R away on which to measure the intensity, see Fig. 2.2. The instantaneous intensity of the light on the screen as a function of the transverse coordinate x is given by the modulus squared of the electric field

$$I(x,t) = |E(x,t)|^2. \tag{2.37}$$

Due to the superposition principle [Hecht and Zajac, 1980], the electric field at the point (x,t) is the sum of the fields from the two sources

$$I(x,t) \propto |E_1(x,t) + E_2(x,t)|^2 \,, \tag{2.38}$$

where the subscript in E_i denotes that the field originated in source i. Assuming that the sources are point-like, producing spherical waves, we can simply relate the fields in Eq. (2.38) to the fields emitted in the source plane at earlier points in time [Born and Wolf, 1980]

$$I(x,t) = |\kappa_1 E_1(\mathbf{r}_{S_1}, t - R_1/c) + \kappa_2 E_2(\mathbf{r}_{S_2}, t - R_2/c)|^2 \,, \tag{2.39}$$

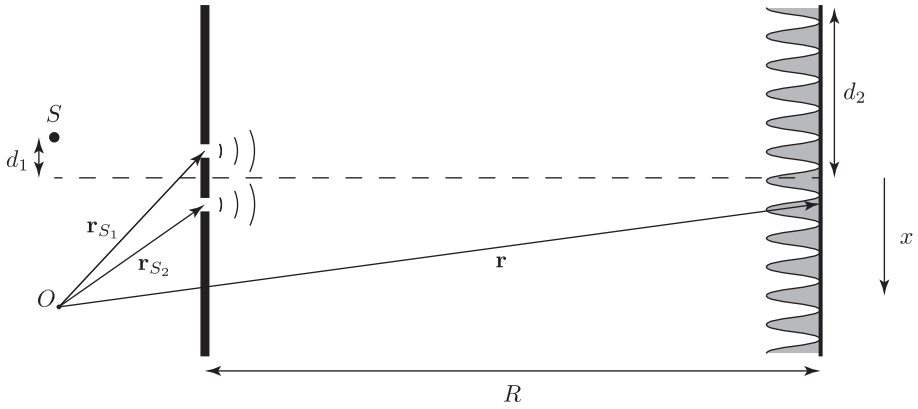

Figure 2.2: Young's double slit experiment. Two secondary sources emit radiation into the far field which is measured on the screen at a distance R. The vector $\mathbf{r}_{S_i}$ gives the position of the i^{th} source relative to the origin O and the vector $\mathbf{r}$ gives the position on the screen. Choosing the screen to parallel to the x axis we can uniquely determine the position on the screen by the single continuous variable x, which labels the displacement from the central axis (dashed line). The distances d_1 and d_2 measure the displacement of the primary source S and the maximum value of x from the central axis respectively.

where $R_i = |\mathbf{r}_{S_i} - \mathbf{r}|$ and κ_1, κ_2 are purely imaginary, time independent factors that depend on the geometry of the arrangement [Mandel and Wolf, 1995] [1]. Taking the average with respect to the time variable t, and assuming that the sources are stationary[2], we find that the average of the intensity is

$$\begin{aligned}\langle I(x)\rangle = {} & |\kappa_1|^2\langle I_1(\mathbf{r}_{S_1})\rangle + |\kappa_2|^2\langle I_2(\mathbf{r}_{S_2})\rangle \\ & + 2\,\mathrm{Re}\left(\kappa_1^*\kappa_2\langle E_1^*(\mathbf{r}_{S_1}, t - R_1/c)E_2(\mathbf{r}_{S_2}, t - R_2/c)\rangle\right). \end{aligned} \tag{2.40}$$

The term $\langle E_1^*(\mathbf{r}_{S_1}, t-R_1/c)E_2(\mathbf{r}_{S_2}, t-R_2/c)\rangle$ is the cross-correlation of the fields from the two sources. Unlike the terms $\langle I_i(\mathbf{r}_{S_i})\rangle$, which correspond to the autocorrelation of the electric field from the i^{th} source with itself, the cross correlation term will in general be a complex number. The first two terms give the contribution to the far field intensity from the individual sources at $\mathbf{r}_{S_1}$ and $\mathbf{r}_{S_2}$. As such we may write $\langle I_i(x)\rangle = |\kappa_i|^2\langle I_i(\mathbf{r}_{S_i})\rangle$. Additionally, since κ_1 and κ_2 are purely imaginary with identical phase, we find

$$\kappa_1^*\kappa_2 = \sqrt{|\kappa_1^*\kappa_2|^2} = \sqrt{|\kappa_1|^2|\kappa_2|^2} = \sqrt{\frac{\langle I_1(x)\rangle}{\langle I_1(\mathbf{r}_{S_1})\rangle}\frac{\langle I_2(x)\rangle}{\langle I_2(\mathbf{r}_{S_2})\rangle}} \tag{2.41}$$

This then gives us for the intensity at position x

$$\langle I(x)\rangle = \langle I_1(x)\rangle + \langle I_2(x)\rangle + 2[\langle I_1(x)\rangle\langle I_2(x)\rangle]^{\frac{1}{2}}\,\mathrm{Re}\left(\gamma(\mathbf{r}_{S_1}, \mathbf{r}_{S_2}, \tau)\right), \tag{2.42}$$

[1] From Kirchoff's diffraction formula we find $E_j(x,t) = -\frac{i}{2\lambda R_j}(1+\cos\chi_j)E_j(\mathbf{r}_{S_j}, t - R_j/c)$, where the prefactor is κ_j, χ_j is the angle between the vectors $\mathbf{r}_{S_j} - OS$ and $\mathbf{r}_{S_j} - \mathbf{r}$. For small displacements d_1 and d_2 (see Fig. 2.2) we find $\cos\chi_{1,2} \approx 1$.

[2] Stationarity implies that the statistical properties of the field are independent of time.

where $\tau = \frac{1}{c}|R_1 - R_2|$ and the function $\gamma(y_1, y_2, \tau)$ is defined as

$$\gamma(\mathbf{r}_{S_1}, \mathbf{r}_{S_2}, \tau) = \frac{\langle E_1^*(\mathbf{r}_{S_1}, t) E_2(\mathbf{r}_{S_2}, t+\tau)\rangle}{[\langle E_1^*(\mathbf{r}_{S_1}) E_1(\mathbf{r}_{S_1})\rangle]^{\frac{1}{2}} [\langle E_2^*(\mathbf{r}_{S_2}) E_2(\mathbf{r}_{S_2})\rangle]^{\frac{1}{2}}} \equiv \frac{\langle E_1^*(\mathbf{r}_{S_1}, t) E_2(\mathbf{r}_{S_2}, t+\tau)\rangle}{[\langle I_1(\mathbf{r}_{S_1})\rangle]^{\frac{1}{2}} [\langle I_2(\mathbf{r}_{S_2})\rangle]^{\frac{1}{2}}}. \tag{2.43}$$

The function $\gamma(\mathbf{r}_{S_1}, \mathbf{r}_{S_2}, \tau)$ is referred to as the *complex degree of coherence* and is used extensively to characterise coherence phenomena. If the two sources are independent then $\langle E_1^*(\mathbf{r}_{S_1}, t) E_2(\mathbf{r}_{S_2}, t+\tau)\rangle = \langle E_1^*(\mathbf{r}_{S_1}, t)\rangle \langle E_2(\mathbf{r}_{S_2}, t+\tau)\rangle$, which is equal to 0 for zero mean fields. The complex degree of coherence also obeys the relation

$$0 \leq |\gamma(\mathbf{r}_{S_1}, \mathbf{r}_{S_2}, \tau)| \leq 1, \tag{2.44}$$

where the extremal values 0 and 1 correspond to completely uncorrelated and completely correlated, respectively.

2.4 Quantum theory of coherence

Consider the detection of a single photon of polarisation λ by a detector at position $\mathbf{r}$ and time t. Typically, the measurement of a photon is achieved by its absorption. The photon is therefore destroyed and the measurement is described by the action of the operator $\hat{E}_\lambda^{(+)}(\mathbf{r}, t)$ on the initial state $|\psi_i\rangle$. If we are ignorant of the final state of the system, $|\psi_f\rangle$, then the total probability of detecting the photon is proportional to [Glauber, 1963c]

$$\sum_{\psi_f} |\langle \psi_f | \hat{E}^{(+)}(\mathbf{r}, t) | \psi_i \rangle|^2, \tag{2.45}$$

where the sum is over all possible final states of the system and we have dropped the subscript λ, assuming that we are only considering photons of a certain polarisation. This leads to

$$\begin{aligned} \sum_{\psi_f} |\langle \psi_f | \hat{E}^{(+)}(\mathbf{r}, t) | \psi_i \rangle|^2 &= \sum_{\psi_f} \langle \psi_i | \hat{E}^{(-)}(\mathbf{r}, t) | \psi_f \rangle \langle \psi_f | \hat{E}^{(+)}(\mathbf{r}, t) | \psi_i \rangle \\ &= \langle \psi_i | \hat{E}^{(-)}(\mathbf{r}, t) \hat{E}^{(+)}(\mathbf{r}, t) | \psi_i \rangle, \end{aligned} \tag{2.46}$$

where we have assumed that the set of final states is complete. If the final states do not constitute a complete set, then we may arbitrarily extend the set to a complete set, since the inner product of any additional states with the state $\hat{E}^{(+)}(\mathbf{r}, t)|\psi_i\rangle$ will be zero. In accordance with Eqs. (2.22) and (2.23), the probability of detection for a mixed state is

$$\mathrm{Tr}[\rho \hat{E}^{(-)}(\mathbf{r}, t) \hat{E}^{(+)}(\mathbf{r}, t)]. \tag{2.47}$$

Measurement of the operator $\hat{E}^{(-)}(\mathbf{r}, t)\hat{E}^{(+)}(\mathbf{r}, t)$ evaluates the photon intensity of the field at the point $\mathbf{r}$ and time t. It is often the case that measurements of the photon intensity do not actually require the quantum mechanical description of light. In these instances, any prediction of the outcome of a photon intensity

measurement can be replicated by treating the electromagnetic field as a classical field (see appendix A). However, in general a fully quantum mechanical description of the field is necessary and predictions of intensity measurement outcomes *cannot* be predicted with a classical description of the electromagnetic field.

Measurements of the intensity are not the only measurements we can make. In fact, intensity measurements belong to a more general class of measurements, of which the intensity operator is merely the simplest. We will examine these measurements in detail shortly but first we will describe the first measurement of this kind and then generalise the theory. In 1956 Hanbury Brown and Twiss discovered a new type of intensity measurement known as intensity interferometry or coincidence counting [Brown and Twiss, 1956a]. Instead of simply determining the photon intensity at a set of points independently, a coincidence measurement simultaneously measures the photon intensity at two separate points, $\mathbf{r}_1$ and $\mathbf{r}_2$, and registers a "click" when the two detectors register photons *coincidentally*. As pointed out by Glauber [Glauber, 1963c], the total probability of detection, analogously to Eq. (2.45), is proportional to

$$\sum_{\psi_f} |\langle \psi_f | \hat{E}^{(+)}(\mathbf{r}_1, t) \hat{E}^{(+)}(\mathbf{r}_2, t) | \psi_i \rangle|^2. \tag{2.48}$$

We now find that the probability of a coincidence detection event is given by

$$\langle \psi_i | \hat{E}^{(-)}(\mathbf{r}_2, t) \hat{E}^{(-)}(\mathbf{r}_1, t) \hat{E}^{(+)}(\mathbf{r}_1, t) \hat{E}^{(+)}(\mathbf{r}_2, t) | \psi_i \rangle, \tag{2.49}$$

or, more generally,

$$\mathrm{Tr}[\rho \hat{E}^{(-)}(\mathbf{r}_2, t) \hat{E}^{(-)}(\mathbf{r}_1, t) \hat{E}^{(+)}(\mathbf{r}_1, t) \hat{E}^{(+)}(\mathbf{r}_2, t)]. \tag{2.50}$$

The detection events described by Eqs. (2.49) and (2.50) are two spatially delocalised photon intensity measurements. We can immediately generalise these measurements to allow for a temporal delay between detections obtaining

$$\mathrm{Tr}[\rho \hat{E}^{(-)}(\mathbf{r}_2, t_2) \hat{E}^{(-)}(\mathbf{r}_1, t_1) \hat{E}^{(+)}(\mathbf{r}_1, t_1) \hat{E}^{(+)}(\mathbf{r}_2, t_2)]. \tag{2.51}$$

If the fields $\hat{E}^{(+)}(\mathbf{r}_1, t_1)$ and $\hat{E}^{(+)}(\mathbf{r}_2, t_2)$ are stationary then the probability depends only on the difference $\tau = t_2 - t_1$.

Following an identical line of reasoning, we find that the detection probability for an n-fold detection at the space-time points $\mathbf{r}_1, t_1, \ldots, \mathbf{r}_n, t_n$ is proportional to

$$\langle \hat{E}^{(-)}(\mathbf{r}_n, t_n) \ldots \hat{E}^{(-)}(\mathbf{r}_1, t_1) \hat{E}^{(+)}(\mathbf{r}_1, t_1) \ldots \hat{E}^{(+)}(\mathbf{r}_n, t_n) \rangle. \tag{2.52}$$

These expectation values appear frequently in quantum optics and therefore have their own name and notation. The expectation in Eq. (2.52) is known as the n^{th} order intensity correlation function and is denoted as

$$G^{(n)}(\mathbf{r}_1, t_1; \ldots; \mathbf{r}_n, t_n) = \langle \hat{E}^{(-)}(\mathbf{r}_n, t_n) \ldots \hat{E}^{(-)}(\mathbf{r}_1, t_1) \hat{E}^{(+)}(\mathbf{r}_1, t_1) \ldots \hat{E}^{(+)}(\mathbf{r}_n, t_n) \rangle. \tag{2.53}$$

To ease the notation slightly, we can make use of the normal ordering notation. Normal ordering of operators places all annihilation operators to the right and all

creation operators to the left. For example, the normal ordering of the product $\hat{a}^\dagger\hat{a}\hat{a}\hat{a}^\dagger\hat{a}$ is

$$:\hat{a}^\dagger\hat{a}\hat{a}\hat{a}^\dagger\hat{a}: = \hat{a}^\dagger\hat{a}^\dagger\hat{a}\hat{a}\hat{a}, \tag{2.54}$$

where the $:\cdot:$ denotes that the operator product is to be normally ordered. The re-ordering is done without the use of the commutation relations and is therefore different to the result obtained when the commutation relations are used to reorder the operators. Using the normal ordering notation we can write the n^{th} order intensity correlation as

$$G^{(n)}(\mathbf{r}_1,t_1;\ldots;\mathbf{r}_n,t_n) = \left\langle : \prod_{i=1}^{n} \hat{E}^{(-)}(\mathbf{r}_i,t_i)\hat{E}^{(+)}(\mathbf{r}_i,t_i) : \right\rangle. \tag{2.55}$$

The intensity correlations are actually a special case of a more general class of correlation functions. Following the notation of Mandel and Wolf we write these correlation functions as [Mandel and Wolf, 1995]

$$\begin{aligned}\Gamma^{(N,M)}(\mathbf{r}_1,t_1,\ldots,\mathbf{r}_N,t_N;\mathbf{r}'_M,t'_M,\ldots,\mathbf{r}'_1,t'_1) = \\ \langle\hat{E}^{(-)}(\mathbf{r}_1,t_1)\ldots\hat{E}^{(-)}(\mathbf{r}_N,t_N)\hat{E}^{(+)}(\mathbf{r}'_M,t'_M)\ldots\hat{E}^{(+)}(\mathbf{r}'_1,t'_1)\rangle.\end{aligned} \tag{2.56}$$

The function $\Gamma^{(N,M)}(\mathbf{r}_1,t_1,\ldots,\mathbf{r}_N,t_N;\mathbf{r}'_M,t'_M,\ldots,\mathbf{r}'_1,t'_1)$ is the expectation of the normally ordered set of N creation operators at the space-time points $\mathbf{r}_1,t_1,\ldots\mathbf{r}_N,t_N$ and M annihilation operators at the space-time points $\mathbf{r}'_1,t'_1,\ldots\mathbf{r}'_M,t'_M$. We see immediately that the intensity correlation functions $G^{(n)}$ are related to the function $\Gamma^{(N,M)}$ by

$$G^{(n)}(\mathbf{r}_1,t_1;\ldots;\mathbf{r}_n,t_n) = \Gamma^{(n,n)}(\mathbf{r}_1,t_1,\ldots,\mathbf{r}_n,t_n;\mathbf{r}_n,t_n,\ldots,\mathbf{r}_1,t_1). \tag{2.57}$$

We can also introduce the normalised n^{th} order intensity correlations, defined as

$$g^{(n)}(\mathbf{r}_1,t_1;\ldots;\mathbf{r}_n,t_n) = \frac{\left\langle : \prod_{i=1}^{n} \hat{E}^{(-)}(\mathbf{r}_i,t_i)\hat{E}^{(+)}(\mathbf{r}_i,t_i) : \right\rangle}{\prod_{j=1}^{n}\langle\hat{E}^{(-)}(\mathbf{r}_j,t_j)\hat{E}^{(+)}(\mathbf{r}_j,t_j)\rangle}, \tag{2.58}$$

or in terms of the generalised correlation functions,

$$g^{(n)}(\mathbf{r}_1,t_1;\ldots;\mathbf{r}_n,t_n) = \frac{\Gamma^{(n,n)}(\mathbf{r}_1,t_1,\ldots,\mathbf{r}_n,t_n;\mathbf{r}_n,t_n,\ldots,\mathbf{r}_1,t_1)}{\prod_{j=1}^{n}\Gamma^{(1,1)}(\mathbf{r}_j,t_j;\mathbf{r}_j,t_j)}. \tag{2.59}$$

We saw earlier in section 2.3 that the important quantity in explaining the presence of interference fringes in a double slit experiment is the complex degree of coherence $\gamma(x_1,x_2,\tau)$. In the quantum theory, the analogous function is given by

$$\gamma^{(1,1)}(\mathbf{r}_1,t_1;\mathbf{r}_2,t_2) = \frac{\Gamma^{(1,1)}(\mathbf{r}_1,t_1;\mathbf{r}_2,t_2)}{[\Gamma^{(1,1)}(\mathbf{r}_1,t_1;\mathbf{r}_1,t_1)\Gamma^{(1,1)}(\mathbf{r}_2,t_2;\mathbf{r}_2,t_2)]^{\frac{1}{2}}}. \tag{2.60}$$

From the operator form of the Cauchy-Schwarz inequality [Titulaer and Glauber, 1965], $|\langle A^\dagger B\rangle|^2 \leq \langle A^\dagger A\rangle\langle B^\dagger B\rangle$, we find

$$|\Gamma^{(1,1)}(\mathbf{r}_1,t_1;\mathbf{r}_2,t_2)| \leq [\Gamma^{(1,1)}(\mathbf{r}_1,t_1;\mathbf{r}_1,t_1)\Gamma^{(1,1)}(\mathbf{r}_2,t_2;\mathbf{r}_2,t_2)]^{\frac{1}{2}}, \tag{2.61}$$

from which it follows that

$$0 \leq |\gamma^{(1,1)}(\mathbf{r}_1, t_1; \mathbf{r}_2, t_2)| \leq 1. \tag{2.62}$$

Again, for stationary fields the temporal dependence of $\gamma^{(1,1)}(\mathbf{r}_1, t_1; \mathbf{r}_2, t_2)$ will be strictly through the difference $\tau = t_2 - t_1$ and we may write

$$\gamma^{(1,1)}(\mathbf{r}_1, t_1; \mathbf{r}_2, t_2) = \gamma^{(1,1)}(\mathbf{r}_1, \mathbf{r}_2, \tau). \tag{2.63}$$

The maximal value of $|\gamma^{(1,1)}(\mathbf{r}_1, t_1; \mathbf{r}_2, t_2)|$ is obtained when equality holds in Eq. (2.61). From Eq. (2.56) we find that this condition is equivalent to

$$|\langle \hat{E}^{(-)}(\mathbf{r}_1, t_1)\hat{E}^{(+)}(\mathbf{r}_2, t_2)\rangle|^2 = \langle \hat{E}^{(-)}(\mathbf{r}_1, t_1)\hat{E}^{(+)}(\mathbf{r}_1, t_1)\rangle\langle \hat{E}^{(-)}(\mathbf{r}_2, t_2)\hat{E}^{(+)}(\mathbf{r}_2, t_2)\rangle, \tag{2.64}$$

which is satisfied if

$$\begin{aligned} |\langle \hat{E}^{(-)}(\mathbf{r}_1, t_1)\hat{E}^{(+)}(\mathbf{r}_2, t_2)\rangle|^2 &= |c^*(\mathbf{r}_1, t_1)c(\mathbf{r}_2, t_2)|^2 \\ &= c^*(\mathbf{r}_1, t_1)c(\mathbf{r}_2, t_2)c(\mathbf{r}_1, t_1)c^*(\mathbf{r}_2, t_2) \\ &= c^*(\mathbf{r}_1, t_1)c(\mathbf{r}_1, t_1)c^*(\mathbf{r}_2, t_2)c(\mathbf{r}_2, t_2) \\ &= \langle \hat{E}^{(-)}(\mathbf{r}_1, t_1)\hat{E}^{(+)}(\mathbf{r}_1, t_1)\rangle\langle \hat{E}^{(-)}(\mathbf{r}_2, t_2)\hat{E}^{(+)}(\mathbf{r}_2, t_2)\rangle, \end{aligned} \tag{2.65}$$

where the c's are complex numbers. We see that this holds for any state $|\psi\rangle$ which is an eigenstate of the operator $\hat{E}^{(+)}$

$$\hat{E}^{(+)}(\mathbf{r}_2, t_2)|\psi\rangle = c(\mathbf{r}_2, t_2)|\psi\rangle. \tag{2.66}$$

Since $\hat{E}^{(+)}$ is proportional to the annihilation operator, $|\gamma^{(1,1)}(\mathbf{r}_1, t_1; \mathbf{r}_2, t_2)| = 1$ for the eigenstates of the annihilation operator.

2.5 The Gaussian moment theorem

The Gaussian moment theorem decomposes the expectation of a product of zero mean, Gaussian distributed (normally distributed) random variables into a sum of products of the expectation of pairs of the variables. We will find it particularly useful in the forthcoming chapters to make use of the Gaussian moment theorem when evaluating correlation functions. We will now prove the Gaussian moment theorem in the context of normally ordered quantum expectation values. First we state the theorem in its original form for completeness.

The statement of the theorem requires the concept of pairings, which we briefly explain before stating the theorem. Given n objects $x_1, \ldots, x_n$ we denote by $\mathcal{P}_n$ all possible pairings of the n objects. In order for this to make sense we assume that n is even such that none of the objects are left unpaired. For example for the four objects i_1, i_2, i_3 and i_4 the set of all pairings is

$$\mathcal{P}_4 = \{[(i_1, i_2), (i_3, i_4)], [(i_1, i_3), (i_2, i_4)], [(i_1, i_4), (i_2, i_3)]\}. \tag{2.67}$$

An element of this set is understood to mean one of the terms in the square brackets $[\cdot]$ and we index the terms in an element by (j), $j = 1 \ldots n$, for example if σ is the second element of $\mathcal{P}_4$ we write $\sigma = \mathcal{P}_4(2) = [(i_1, i_3), (i_2, i_4)]$ and $\sigma(1) = i_1$, $\sigma(2) = i_3$, $\sigma(3) = i_2$ and $\sigma(4) = i_4$. Using this definition we can now write the Gaussian moment theorem as follows [Mandel and Wolf, 1995],

Theorem 1. *Given n normally distributed random variables $x_{i_1}, \ldots, x_{i_n}$ with first moments $\langle x_{i_1}\rangle, \ldots, \langle x_{i_n}\rangle$, the expectation value of the product of centred variables $\Delta x_{i_j} = x_{i_j} - \langle x_{i_j}\rangle$ is given by*

$$\langle \Delta x_{i_1} \Delta x_{i_2} \ldots \Delta x_{i_n}\rangle = \begin{cases} 0 & \text{if } n \text{ odd} \\ \sum_{\sigma\in\mathcal{P}_n} \langle \Delta x_{\sigma(1)}\Delta x_{\sigma(2)}\rangle \ldots \langle \Delta x_{\sigma(n-1)}\Delta x_{\sigma(n)}\rangle & \text{if } n \text{ even} \end{cases}, \tag{2.68}$$

where $\mathcal{P}_n$ is the set of all $(2n)!/(2^n n!)$ pairings of the n objects $\{i_1, \ldots, i_n\}$.

To demonstrate the usefulness of the Gaussian moment theorem for normally ordered quantum expectation values, it is convenient to first prove the optical equivalence theorem. Consider a function of the creation and annihilation operators,

$$f(\hat{a}, \hat{a}^\dagger) = \sum_{n,m} b_{nm} \hat{a}^{\dagger n} \hat{a}^m, \tag{2.69}$$

where each term is normally ordered. If we evaluate the expectation value of this operator making use of the P-representation, we find

$$\begin{aligned} \langle f(\hat{a}, \hat{a}^\dagger)\rangle &= \mathrm{Tr}[\rho f(\hat{a}, \hat{a}^\dagger)] \\ &= \mathrm{Tr}\left[\int \mathrm{d}^2\alpha\, P_\rho(\alpha) |\alpha\rangle\langle\alpha| \sum_{n,m} b_{nm} \hat{a}^{\dagger n}\hat{a}^m\right] \\ &= \int \mathrm{d}^2\alpha\, P_\rho(\alpha) \sum_{n,m} b_{nm} \mathrm{Tr}[\hat{a}^m|\alpha\rangle\langle\alpha|\hat{a}^{\dagger n}] \\ &= \int \mathrm{d}^2\alpha\, P_\rho(\alpha) \sum_{n,m} b_{nm} \alpha^m \alpha^{*n} \\ &= \int \mathrm{d}^2\alpha\, P_\rho(\alpha) f(\alpha, \alpha^*). \end{aligned} \tag{2.70}$$

We see that the expectation value is equivalent to the averaging of the function $f(\alpha, \alpha^*)$ over the complex plane with the $P_\rho(\alpha)$ function playing the role of a weighting function. This is reminiscent of the classical average of the function $f(\alpha, \alpha^*)$ with $P_\rho(\alpha)$ the probability density function. This is known as the *optical equivalence theorem*, and states that the expectation value of any normally ordered product operator can be replaced by the average of the function produced by replacing all creation and annihilation operators by complex random variables $\alpha*$ and α respectively.

The P-representation is of particular importance for evaluating the expectation value of normally ordered operators. Sometimes we may find ourselves confronted with an operator which is not in normal ordered form. It may turn out that applying the commutation relations to normal order the operator is straightforward, generally however, this may not be the case. Therefore it is worth noting the corresponding procedures used for other common orderings. Firstly, when the operator is anti-normally ordered (all creation operators to the right of annihilation operators) we find that the analogous version of Eq. (2.70) requires the use of the weighting function $Q_\rho(\alpha)$, the Husimi or Q-representation. Similarly, when the operators are symmetrically ordered, the weighting function becomes the Wigner function, $W_\rho(\alpha)$ [Cahill and Glauber, 1969]. All of these functions, $P_\rho(\alpha)$, $Q_\rho(\alpha)$ and $W_\rho(\alpha)$,

represent probability density-like functions which describe the distribution of the state in the complex α plane. Due to the fact that the coherent states are not orthogonal, neither $P_\rho(\alpha)$, $Q_\rho(\alpha)$ or $W_\rho(\alpha)$ corresponds to the probability density of finding the state in the coherent state $|\alpha\rangle\langle\alpha|$. The functions P, Q and W therefore cannot correspond to true probability distribution functions [Gnedenko, 1962]. In addition, under certain conditions the functions P and W can take negative values in certain regions of the complex α plane, which is also not permitted for a true probability distribution function [Gnedenko, 1962]. We therefore refer to P, Q and W as quasi-probability functions. A full account of the usefulness of these quasi-probability distributions in quantum optics is given in appendix A.

With the P-representation in mind, we consider again the intensity correlation functions of section 2.4. By the optical equivalence theorem we may write [Mandel and Wolf, 1995]

$$\begin{aligned} G^{(n)}(\mathbf{r}_1,t_1;\ldots;\mathbf{r}_n,t_n) &= \left\langle : \prod_{i=1}^{n} \hat{E}^{(-)}(\mathbf{r}_i,t_i)\hat{E}^{(+)}(\mathbf{r}_i,t_i) : \right\rangle \\ &= \left\langle \prod_{i=1}^{n} E^{(-)}(\mathbf{r}_i,t_i)E^{(+)}(\mathbf{r}_i,t_i) \right\rangle_P , \end{aligned} \tag{2.71}$$

where $E^{(\pm)}(\mathbf{r}_i,t_i)$ are now the eigenvalues of the operator $\hat{E}^{(\pm)}(\mathbf{r}_i,t_i)$, and we have dropped the normal ordering notation since all of the eigenvalues commute. The notation $\langle\cdot\rangle_P$ reminds us that the expectation value is now the classical expectation value taken with respect to the weighting function $P_\rho(\alpha)$. In accordance with the optical equivalence theorem, these eigenvalues are complex random variables. If the function $P_\rho(\alpha)$ takes on the form of a Gaussian distribution, then the random variables $E^{(+)}(\mathbf{r}_i,t_i)$ and $E^{(-)}(\mathbf{r}_i,t_i)$ will be Gaussian random variates. For complex, Gaussian random variables z_j [Reed, 1962] we have

$$\begin{aligned} \langle \Delta z_1^* \ldots \Delta z_N^* \Delta z_{N+1} \ldots \Delta z_{N+M} \rangle &= \begin{cases} 0 & \text{if } N \neq M \\ \sum_{\sigma\in S_N} \langle \Delta z_1^* \Delta z_{\sigma(1)} \rangle \ldots \langle \Delta z_n^* \Delta z_{\sigma(n)} \rangle & \text{if } N = M \end{cases} \\ &= \delta_{N,M} \sum_{\sigma\in S_N} \prod_{i=1}^{N} \langle \Delta z_i^* \Delta z_{\sigma(i)} \rangle, \end{aligned} \tag{2.72}$$

where S_N is the symmetric group containing all $N!$ permutations of N objects. This then leads to the Gaussian moment theorem in the context of quantum mechanical expectation values and we find, for a state with a zero mean, Gaussian P-representation,

$$G^{(n)}(\mathbf{r}_1,t_1;\ldots;\mathbf{r}_n,t_n) = \sum_{\sigma\in S_n} \prod_{i=1}^{n} \Gamma^{(1,1)}(\mathbf{r}_i,t_i;\mathbf{r}_{\sigma(i)},t_{\sigma(i)}). \tag{2.73}$$

The higher order intensity correlations for quantum states with a Gaussian P-representation are therefore entirely determined by the second order degree of coherence. The normalised intensity correlations can also be decomposed into lower

order moments,

$$
\begin{aligned}
g^{(n)}(\mathbf{r}_1,t_1;\ldots;\mathbf{r}_n,t_n) &= \frac{G^{(n)}(\mathbf{r}_1,t_1;\ldots;\mathbf{r}_n,t_n)}{\prod_{j=1}^{n}\Gamma^{(1,1)}(\mathbf{r}_j,t_j;\mathbf{r}_j,t_j)} \\
&= \sum_{\sigma\in S_n}\frac{\prod_{i=1}^{n}\Gamma^{(1,1)}(\mathbf{r}_i,t_i;\mathbf{r}_{\sigma(i)},t_{\sigma(i)})}{\prod_{j=1}^{n}\Gamma^{(1,1)}(\mathbf{r}_j,t_j;\mathbf{r}_j,t_j)} \\
&= \sum_{\sigma\in S_n}\prod_{i=1}^{n}\frac{\Gamma^{(1,1)}(\mathbf{r}_i,t_i;\mathbf{r}_{\sigma(i)},t_{\sigma(i)})}{\Gamma^{(1,1)}(\mathbf{r}_i,t_i;\mathbf{r}_i,t_i)} \\
&= \sum_{\sigma\in S_n}\prod_{i=1}^{n}\gamma^{(1,1)}(\mathbf{r}_i,t_i;\mathbf{r}_{\sigma(i)},t_{\sigma(i)}),
\end{aligned}
\tag{2.74}
$$

where we see that it is now the complex degree of coherence $\gamma^{(1,1)}(\mathbf{r}_i,t_i;\mathbf{r}_{\sigma(i)},t_{\sigma(i)})$ that determines all the higher order normalised intensity correlation functions. Although superficially different from the form given in Theorem 1, the complex version of the Gaussian moment theorem Eq. (2.72), is particularly useful in quantum optics where the variables we encounter are frequently complex.

2.6 Correlations in the far field

So far we have considered how to express the various types of correlations that may exist in the electromagnetic field. We now consider the effect of propagation on these correlations and derive formulas that can be used to determine the correlations of the field once it has propagated away from the radiation source. To derive the propagation formulas for the mutual coherence function $\Gamma(\mathbf{r}_1,t_1;\mathbf{r}_2,t_2)$, we first derive a pair of wave equations obeyed by $\Gamma(\mathbf{r}_1,t_1;\mathbf{r}_2,t_2)$ and also a pair of Helmholtz equations obeyed by the Fourier transform of $\Gamma(\mathbf{r}_1,t_1;\mathbf{r}_2,t_2)$. Starting from the wave equation Eq. (2.4), we take the complex conjugate and multiply from the right by $\mathbf{E}(\mathbf{r}',t')$

$$
\nabla^2\mathbf{E}^*(\mathbf{r},t)\mathbf{E}(\mathbf{r}',t') = \frac{1}{c^2}\frac{\partial^2\mathbf{E}^*(\mathbf{r},t)}{\partial t^2}\mathbf{E}(\mathbf{r}',t'). \tag{2.75}
$$

Since the differential operators in the previous equation are taken with respect to $\mathbf{r}$ and t, we can place the factor $\mathbf{E}(\mathbf{r}',t')$ underneath the derivatives. Simultaneously we take the expectation value to arrive at [Beran and Parrent, 1964]

$$
\nabla^2\langle\mathbf{E}^*(\mathbf{r},t)\mathbf{E}(\mathbf{r}',t')\rangle = \frac{1}{c^2}\frac{\partial^2\langle\mathbf{E}^*(\mathbf{r},t)\mathbf{E}(\mathbf{r}',t')\rangle}{\partial t^2}, \tag{2.76}
$$

where $\langle\mathbf{E}^*(\mathbf{r},t)\mathbf{E}(\mathbf{r}',t')\rangle$ is the mutual coherence function. In a similar way we can derive the equation

$$
\nabla'^2\langle\mathbf{E}^*(\mathbf{r},t)\mathbf{E}(\mathbf{r}',t')\rangle = \frac{1}{c^2}\frac{\partial^2\langle\mathbf{E}^*(\mathbf{r},t)\mathbf{E}(\mathbf{r}',t')\rangle}{\partial t'^2}, \tag{2.77}
$$

which gives us the two wave equations

$$
\nabla^2\Gamma(\mathbf{r},t;\mathbf{r}',t') = \frac{1}{c^2}\frac{\partial^2\Gamma(\mathbf{r},t;\mathbf{r}',t')}{\partial t^2} \tag{2.78}
$$

$$
\nabla'^2\Gamma(\mathbf{r},t;\mathbf{r}',t') = \frac{1}{c^2}\frac{\partial^2\Gamma(\mathbf{r},t;\mathbf{r}',t')}{\partial t'^2}, \tag{2.79}
$$

which the mutual coherence function $\Gamma(\mathbf{r}, t; \mathbf{r}', t')$ satisfies in free space [Wolf, 1955]. For stationary processes it is typical to write the mutual coherence function as $\Gamma(\mathbf{r}, \mathbf{r}', \tau)$ and, using the definition $\tau = t' - t$ we find

$$\frac{\partial}{\partial t'} = \frac{\partial}{\partial \tau} \qquad \text{and} \qquad \frac{\partial}{\partial t} = -\frac{\partial}{\partial \tau} \tag{2.80}$$

which allows us to replace the differential operators $\partial^2/\partial t'^2$ and $\partial^2/\partial t^2$ with $\partial^2/\partial \tau^2$. We also define the cross spectral density function $W(\mathbf{r}, \mathbf{r}', \nu)$ as the Fourier transform of $\Gamma(\mathbf{r}, \mathbf{r}', \tau)$

$$W(\mathbf{r}, \mathbf{r}', \nu) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi\nu\tau} \Gamma(\mathbf{r}, \mathbf{r}', \tau)\, \mathrm{d}\tau\,. \tag{2.81}$$

The cross spectral density satisfies the pair of Helmholtz equations in free space [Klauder and Sudarshan, 1968]

$$\nabla^2 W(\mathbf{r}, \mathbf{r}', \nu) = -k^2 W(\mathbf{r}, \mathbf{r}', \nu) \tag{2.82a}$$
$$\nabla'^2 W(\mathbf{r}, \mathbf{r}', \nu) = -k^2 W(\mathbf{r}, \mathbf{r}', \nu), \tag{2.82b}$$

where $k = 2\pi\nu/c$ is the wavenumber. Similarly to the mutual coherence function, the cross spectral density function can also be expressed as an expectation of two complex fields [Mandel and Wolf, 1995]

$$W(\mathbf{r}, \mathbf{r}', \nu) = \langle \mathscr{E}^*(\mathbf{r}, \nu) \mathscr{E}(\mathbf{r}', \nu') \rangle \delta(\nu - \nu') \tag{2.83}$$

where $\mathscr{E}(\mathbf{r}, \nu)$ is a solution to the Helmholtz equation and is the Fourier transform of $\mathbf{E}(\mathbf{r}, t)$.

2.7 The Rayleigh diffraction formula

The Rayleigh diffraction formula gives the solution to the Helmholtz equation in the half space $z > 0$ for a field propagating in the positive z direction as a function of the values of the field across the entire $z = 0$ plane, assuming that any sources of the field are in the opposing half space $z \leq 0$ [Mandel and Wolf, 1995]. Fig.2.3 illustrates the arrangement of the key quantities in the formula. The Rayleigh diffraction formula is

$$\mathscr{E}(x, y, z) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathscr{E}(\tilde{x}, \tilde{y}, 0) \frac{\partial}{\partial z} \left(\frac{\mathrm{e}^{ikR}}{R} \right) \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{y} \tag{2.84}$$

where $R^2 = (x - \tilde{x})^2 + (y - \tilde{y})^2 + z^2$, $k = 2\pi\nu/c$ and $\mathscr{E}$ is a solution to the Helmholtz equation. Since $W(\mathbf{r}, \mathbf{r}', \nu)$ satisfies the Helmholtz Eqs. (2.82a)(2.82b) we can use the Rayleigh diffraction formula to find the cross spectral density in the half space $z > 0$ assuming we know the values of the cross spectral density in the entire plane $z = 0$. Applying the Rayleigh diffraction formula to the first argument of $W(\mathbf{r}, \mathbf{r}', \nu)$ i.e. $\mathbf{r}$, and keeping $\mathbf{r}'$ constant, we find

$$W(x, y, z, \mathbf{r}', \nu) = -\frac{1}{2\pi} \int\int W(\tilde{x}, \tilde{y}, 0, \mathbf{r}', \nu) \frac{\partial}{\partial z} \left(\frac{\mathrm{e}^{ikR}}{R} \right) \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{y}\,, \tag{2.85}$$

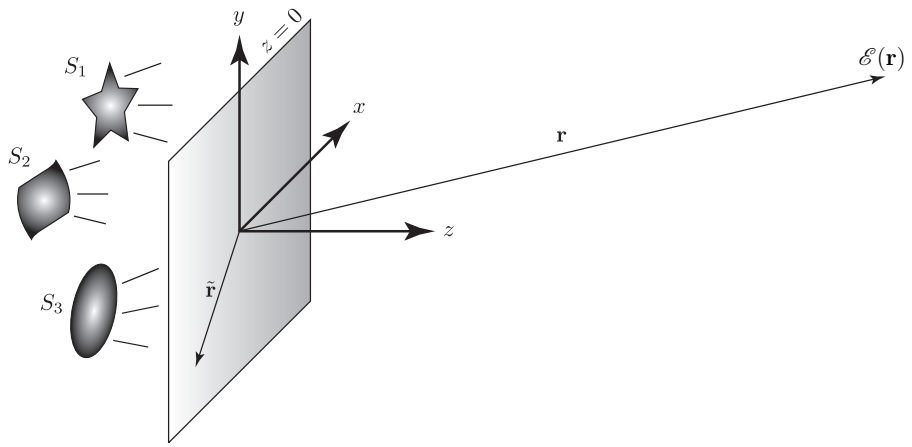

Figure 2.3: The Rayleigh diffraction formula. The sources, S_1, S_2 and S_3, in the half space $z \leq 0$ create the field $\mathscr{E}$ which propagates into the half space $z > 0$. Given knowledge of the field across the whole plane $z = 0$ the Rayleigh diffraction formula gives the value of the field $\mathscr{E}(\mathbf{r})$ consistent with the Helmholtz equation for all values of $\mathbf{r}$ in the half space $z > 0$.

where we have dropped the limits on the integral for brevity. Similarly for $\mathbf{r}'$

$$W(\mathbf{r}, x', y', z', \nu) = -\frac{1}{2\pi} \int\int W(\mathbf{r}, \tilde{x}', \tilde{y}', 0, \nu) \frac{\partial}{\partial z'} \left(\frac{\mathrm{e}^{-ikR'}}{R'} \right) \mathrm{d}\tilde{x}' \, \mathrm{d}\tilde{y}' \,, \tag{2.86}$$

where $R' = |\mathbf{r}' - \tilde{\mathbf{r}}'|$ and the minus sign in the exponential comes from taking the complex conjugate of Eq.(2.84). Substituting Eq. (2.85) into Eq. (2.86) we obtain

$$W(\mathbf{r},\mathbf{r}', \nu) = \frac{1}{4\pi^2} \int\int\int\int W(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}', \nu) \frac{\partial}{\partial z} \left(\frac{\mathrm{e}^{ikR}}{R} \right) \frac{\partial}{\partial z'} \left(\frac{\mathrm{e}^{-ikR'}}{R'} \right) \mathrm{d}\tilde{x} \, \mathrm{d}\tilde{y} \, \mathrm{d}\tilde{x}' \, \mathrm{d}\tilde{y}' \,. \tag{2.87}$$

The above equation is an exact formula allowing us to calculate the cross spectral density at any point in the half space $z > 0$. To find an approximate formulation of Eq. (2.86) valid in the far field ($R \gg \lambda$) we first evaluate the derivatives with respect to z and z' in the far field

$$\frac{\partial}{\partial z} \left(\frac{\mathrm{e}^{ikR}}{R} \right) \simeq \frac{ikz\mathrm{e}^{ikR}}{R^2} \qquad \text{and} \qquad \frac{\partial}{\partial z'} \left(\frac{\mathrm{e}^{-ikR'}}{R'} \right) \simeq \frac{-ikz'\mathrm{e}^{-ikR'}}{R'^2}. \tag{2.88}$$

Taking the Fourier transform gives

$$\Gamma(\mathbf{r}, \mathbf{r}', \tau) = \frac{1}{4\pi^2} \int\int\int\int \frac{zz'}{R^2 R'^2} \left[\int_0^\infty \mathrm{e}^{-i2\pi\nu\tau} k^2 W(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}', \nu) \mathrm{e}^{-ik(R'-R)} \, \mathrm{d}\nu \right] \mathrm{d}\tilde{x} \, \mathrm{d}\tilde{y} \, \mathrm{d}\tilde{x}' \, \mathrm{d}\tilde{y}' \,. \tag{2.89}$$

Concentrating on the term in the square brackets we find

$$\int_0^\infty \exp\left[-i2\pi\nu \left(\tau + \frac{R' - R}{c} \right) \right] \left(\frac{2\pi\nu}{c} \right)^2 W(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}', \nu) \, \mathrm{d}\nu \; = -\frac{1}{c^2} \int_0^\infty \frac{\partial^2}{\partial \tau^2} \exp\left[-i2\pi\nu \left(\tau + \frac{R' - R}{c} \right) \right] W(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}', \nu) \, \mathrm{d}\nu \,. \tag{2.90}$$

Exchanging the order of differentiation and integration leads to

$$-\frac{1}{c^2}\int_0^\infty \frac{\partial^2}{\partial\tau^2}\exp\left[-i2\pi\nu\left(\tau+\frac{R'-R}{c}\right)\right]W(\tilde{\mathbf{r}},\tilde{\mathbf{r}}',\nu)\,\mathrm{d}\nu = -\frac{1}{c^2}\frac{\partial^2}{\partial\tau^2}\Gamma(\tilde{\mathbf{r}},\tilde{\mathbf{r}}',\tau+(R'-R)/c), \tag{2.91}$$

and we find for the mutual degree of coherence in the far field

$$\Gamma(\mathbf{r},\mathbf{r}',\tau) = -\frac{1}{4\pi^2c^2}\int\int\int\int \frac{zz'}{R^2R'^2}\frac{\partial^2}{\partial\tau^2}\Gamma(\tilde{\mathbf{r}},\tilde{\mathbf{r}}',\tau+(R'-R)/c)\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{y}\,\mathrm{d}\tilde{x}'\,\mathrm{d}\tilde{y}'\,. \tag{2.92}$$

The integrals over the variables $\tilde{x}$, $\tilde{y}$, $\tilde{x}'$, and $\tilde{y}'$ all take place over the $z=0$ plane where the mutual degree of coherence is known for all pairs of coordinates.

2.8 Monochromatic approximation

If the fields are monochromatic, or at least approximately so, Eq. (2.92) takes a much simpler form. Looking back to the term in parenthesis from Eq. (2.89) we make the approximation

$$\int_0^\infty \exp\left[-i2\pi\nu\left(\tau+\frac{R'-R}{c}\right)\right]\left(\frac{2\pi\nu}{c}\right)^2 W(\tilde{\mathbf{r}},\tilde{\mathbf{r}}',\nu)\,\mathrm{d}\nu \approx \left(\frac{2\pi\bar{\nu}}{c}\right)^2\int_0^\infty \exp\left[-i2\pi\nu\left(\tau+\frac{R'-R}{c}\right)\right]W(\tilde{\mathbf{r}},\tilde{\mathbf{r}}',\nu)\,\mathrm{d}\nu \tag{2.93}$$

where $\bar{\nu}$ is the central frequency of the field. Which leads to the much simpler expression for $\Gamma(\mathbf{r},\mathbf{r}',\tau)$

$$\Gamma(\mathbf{r},\mathbf{r}',\tau) = \left(\frac{\bar{\nu}}{c}\right)^2\int\int\int\int \frac{zz'}{R^2R'^2}\Gamma(\tilde{\mathbf{r}},\tilde{\mathbf{r}}',\tau+(R'-R)/c)\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{y}\,\mathrm{d}\tilde{x}'\,\mathrm{d}\tilde{y}' \tag{2.94}$$

where we no longer require the second derivative of Γ with respect to τ. This covers the basic ideas of coherence and quantum optics that we shall make frequent use of in later chapters. In the next chapter we give an account of estimation theory, which we will again make frequent use of later.

Chapter 3

Estimation Theory

In this section we give an overview of the field of estimation theory, first examining the classical theory and then exploring the quantum formalism. The fundamental task of estimation theory is to estimate certain values or *parameters*, from a set of data which exhibit a statistical nature. In the classical theory the statistical aspects of the data can be regarded as occurring due to any number of experimental uncertainties whereas in the quantum formalism the statistical behaviour is directly related to the fundamental uncertainty present in a quantum state. A key quantity in estimation theory is the Fisher information. Here we will derive the form of the Fisher information and explain it's relevance to parameter estimation problems. Analogously, in the quantum regime we encounter the quantum Fisher information. Again we derive this quantity and explore its relevance.

3.1 Classical estimation theory

The problem of estimating a set of parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_n\}$ from a set of observation data $\mathbf{x} = \{\mathrm{x}[1], \ldots, \mathrm{x}[N]\}$ is the object of classical estimation theory. The equations which govern how the observed data is used to provide a value for the parameters $\boldsymbol{\theta}$ are known as *estimators*. Symbolically we write

$$\check{\boldsymbol{\theta}} = f(\mathbf{x}), \tag{3.1}$$

where we use the caron $\check{\cdot}$ to denote an estimator. Typically in estimation theory we use the caret $\hat{\cdot}$ to denote an estimator, however since this was used in the previous chapter to denote a quantum operator here we will use the caron instead. The distinction between the actual value of the parameters $\boldsymbol{\theta}$ and an estimate $\check{\boldsymbol{\theta}}$ is crucially important. Whereas the actual value of the parameter $\boldsymbol{\theta}$ is a number or set of numbers, the estimate of the parameter is, in general, a random variable or set of random variables. Two sets of observation data may be obtained for which the parameters are equal, $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, but, due to the probabilistic nature of the data, the exact values of the observations are not equal $\mathbf{x}_1 \neq \mathbf{x}_2$. In general the estimates will not be equal since they are derived from two different data sets using the same procedure.

The example above demonstrates an important feature of estimators. Since the values of the parameters are equal, we would like the estimates to be equal. This would correspond to a perfect estimator, one which, given any data set of length N, returns the exact values of the parameters. This is clearly an unrealistic

idealisation. However, when the above example is performed, we expect that the estimates are at least similar. This leads us to a natural characterisation of the estimators performance, namely the *variance* of the estimator.

The aim in estimation theory is to identify an estimator with the smallest possible variance, therefore leading to estimates that are, on average, as close as they can be to the actual values $\boldsymbol{\theta}$. In order to achieve this goal, we must know the joint probability distribution function (PDF) of our data $p(\mathbf{x}|\boldsymbol{\theta}) = p(\mathrm{x}[1], \ldots, \mathrm{x}[N]|\theta)$, which gives the probability of obtaining all observations $\mathrm{x}[1], \ldots, \mathrm{x}[N]$ given the values of the parameters $\boldsymbol{\theta}$.

3.1.1 The Cramér-Rao bound

In general it is not possible to state exactly what the variance of our estimator is. However, it is usually possible to obtain a lower bound on the variance. Various methods for bounding the estimators variance exist, the most commonly used being the Cramér-Rao bound.

Consider the single parameter estimation problem, $\boldsymbol{\theta} = \theta$. We define the variation in an estimate of θ to be $\delta\check{\theta} = \check{\theta} - \langle\check{\theta}\rangle$, where $\langle\check{\theta}\rangle$ is the expectation value of the estimator defined by

$$\langle\check{\theta}\rangle = \int \mathrm{d}\mathbf{x}\, p(\mathbf{x}|\theta)\check{\theta}. \tag{3.2}$$

We purposefully avoid the notation $\Delta\check{\theta}$ which should be reserved for the error $\Delta\check{\theta} = \check{\theta} - \theta$. The variance of the estimator should not be confused with the mean square error, which is defined by

$$\begin{aligned}\langle(\Delta\check{\theta})^2\rangle &= \langle(\check{\theta} - \theta)^2\rangle = \langle\check{\theta}^2\rangle - 2\langle\check{\theta}\rangle\theta + \theta^2 = \langle\check{\theta}^2\rangle - \langle\check{\theta}\rangle^2 + \langle\check{\theta}\rangle^2 - 2\langle\check{\theta}\rangle\theta + \theta^2 \\ &= \mathrm{Var}(\check{\theta}) + \mathrm{B}(\check{\theta})^2,\end{aligned} \tag{3.3}$$

where $B(\check{\theta})$ is the bias of the estimator defined as $B(\check{\theta}) = \langle\check{\theta}\rangle - \theta$ and we have defined the variance of the estimator to be $\mathrm{Var}(\check{\theta}) = \langle\check{\theta}^2\rangle - \langle\check{\theta}\rangle^2$. We see directly from Eq. (3.3) that the mean square error is equal to the variance if and only if the the bias is 0. We refer to such an estimator as unbiased, for which the equation $\langle\check{\theta}\rangle = \theta$ holds.

Now, since

$$\int \mathrm{d}\mathbf{x}\, p(\mathbf{x}|\theta)\langle\check{\theta}\rangle = \langle\check{\theta}\rangle, \tag{3.4}$$

we find

$$\langle\delta\check{\theta}\rangle = \int \mathrm{d}\mathbf{x}\, p(\mathbf{x}|\theta)(\check{\theta} - \langle\check{\theta}\rangle) = 0. \tag{3.5}$$

For M independent measurements with outcomes $\mathbf{x}_1, \ldots, \mathbf{x}_M$ we can write

$$\int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M\, p(\mathbf{x}_1|\theta) \ldots p(\mathbf{x}_M|\theta)\delta\check{\theta} = 0. \tag{3.6}$$

Example: Mach-Zehnder Interferometer

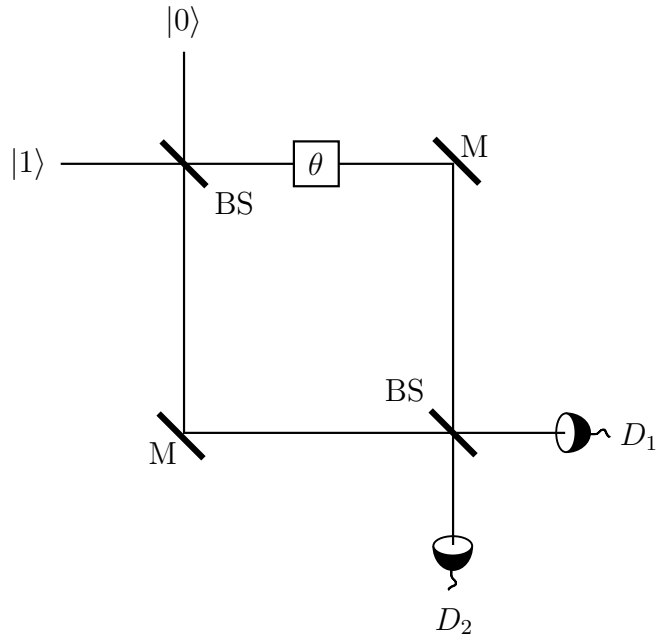

Figure 3.1: A Mach-Zehnder interferometer with a phase shift θ along the upper arm consisting of two beamsplitters (BS), two mirrors (M) and two single photon detectors $D_{1,2}$.

Here we give an example to ease the reader into the notation. Consider a Mach-Zehnder interferometer into which we send a single photon into one port and the vacuum into the other port as shown in Fig. 3.1. Imagine we run the experiment N times, sending a single photon in each time and monitoring the output of the detectors. Each run of the experiment returns two pieces of data x[1] and x[2], which are the photon counts in each detector D_1 and D_2 respectively. Assuming no losses in the interferometer and ideal detectors, the probability distribution function, $p(\mathrm{x}[1], \mathrm{x}[2]|\theta)$, can be summarised as

$$\begin{aligned} p(1,1|\theta) &= 0, \qquad p(1,0|\theta) = \cos^2(\theta/2), \\ p(0,0|\theta) &= 0, \qquad p(0,1|\theta) = \sin^2(\theta/2), \end{aligned}$$

If we count a total N_1 photons in detector D_1 and N_2 in D_2 then a reasonable estimator for θ is

$$\check{\theta} = \arccos\left(\frac{N_1 - N_2}{N}\right).$$

Taking the derivative with respect to θ yields

$$\int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{j=1}^{M} p(\mathbf{x}_j|\theta) \left(\sum_{i=1}^{M} \frac{1}{p(\mathbf{x}_i|\theta)} \frac{\partial p(\mathbf{x}_i|\theta)}{\partial \theta} \delta\check{\theta} + \frac{\partial \delta\check{\theta}}{\partial \theta} \right) = 0. \tag{3.7}$$

Since $\check{\theta} = f(\mathbf{x})$ is an estimator, it does not depend on θ, therefore

$$\frac{\partial \delta\check{\theta}}{\partial \theta} = -\frac{\partial \langle\check{\theta}\rangle}{\partial \theta}, \tag{3.8}$$

and

$$\begin{aligned} \int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{j=1}^{M} p(\mathbf{x}_j|\theta) \frac{\partial \delta\check{\theta}}{\partial \theta} &= -\left\langle \frac{\partial \langle\check{\theta}\rangle}{\partial \theta} \right\rangle \\ &= -\frac{\partial \langle\check{\theta}\rangle}{\partial \theta}, \end{aligned} \tag{3.9}$$

where the last equality holds because $\partial_\theta \langle\check{\theta}\rangle$ is independent of $\mathbf{x}_1, \ldots, \mathbf{x}_M$. We now use the Cauchy-Schwarz inequality

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle \langle y, y\rangle, \tag{3.10}$$

making the substitutions

$$\begin{aligned} x &= \sum_{i=1}^{M} \frac{1}{p(\mathbf{x}_i|\theta)} \frac{\partial p(\mathbf{x}_i|\theta)}{\partial \theta}, \\ y &= \delta\check{\theta}, \end{aligned}$$

we find

$$\begin{aligned} \int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{j=1}^{M} p(\mathbf{x}_j|\theta) \left(\sum_{i=1}^{M} \frac{1}{p(\mathbf{x}_i|\theta)} \frac{\partial p(\mathbf{x}_i|\theta)}{\partial \theta} \right)^2 \\ \times \int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{j=1}^{M} p(\mathbf{x}_j|\theta) (\delta\check{\theta})^2 \geq \left| \frac{\partial \langle\check{\theta}\rangle}{\partial \theta} \right|^2, \end{aligned} \tag{3.11}$$

where the right hand side is obtained by noticing that $\langle x, y\rangle = -\langle \partial_\theta x\rangle$ and $-\langle \partial_\theta x\rangle = \partial_\theta \langle\check{\theta}\rangle$, which come directly from Eq. (3.7) and Eq. (3.9) respectively.

We can simplify the first term by noticing that

$$\begin{aligned} \int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{j=1}^{M} p(\mathbf{x}_j|\theta) \left(\sum_{i=1}^{M} \frac{1}{p(\mathbf{x}_i|\theta)} \frac{\partial p(\mathbf{x}_i|\theta)}{\partial \theta} \right) \left(\sum_{k=1}^{M} \frac{1}{p(\mathbf{x}_k|\theta)} \frac{\partial p(\mathbf{x}_k|\theta)}{\partial \theta} \right) = \\ \sum_{i,k=1}^{M} \int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{j=1}^{M} \frac{p(\mathbf{x}_j|\theta)}{p(\mathbf{x}_i|\theta) p(\mathbf{x}_k|\theta)} \frac{\partial p(\mathbf{x}_i|\theta)}{\partial \theta} \frac{\partial p(\mathbf{x}_k|\theta)}{\partial \theta}, \end{aligned} \tag{3.12}$$

which for $i \neq k$ gives

$$\begin{aligned}\int \mathrm{d}\mathbf{x}_i \int \mathrm{d}\mathbf{x}_k \, \frac{\partial p(\mathbf{x}_i|\theta)}{\partial\theta}\frac{\partial p(\mathbf{x}_k|\theta)}{\partial\theta} &= \int \mathrm{d}\mathbf{x}_i \, \frac{\partial p(\mathbf{x}_i|\theta)}{\partial\theta} \int \mathrm{d}\mathbf{x}_k \, \frac{\partial p(\mathbf{x}_k|\theta)}{\partial\theta} \\ &= \frac{\partial}{\partial\theta}\left[\int \mathrm{d}\mathbf{x}_i \, p(\mathbf{x}_i|\theta)\right]\frac{\partial}{\partial\theta}\left[\int \mathrm{d}\mathbf{x}_k \, p(\mathbf{x}_k|\theta)\right] \\ &= \left(\frac{\partial}{\partial\theta}1\right)^2 \\ &= 0. \end{aligned} \tag{3.13}$$

Eq. (3.11) now simplifies to

$$M\int \mathrm{d}\mathbf{x} \, \frac{1}{p(\mathbf{x}|\theta)}\left(\frac{\partial p(\mathbf{x}|\theta)}{\partial\theta}\right)^2 \langle(\delta\check{\theta})^2\rangle \geq \left|\frac{\partial\langle\check{\theta}\rangle}{\partial\theta}\right|^2. \tag{3.14}$$

The integral in the previous expression plays a particularly important role in estimation theory. It is referred to as the Fisher information and it can be expressed as the expectation value of the derivative of the *log likelihood*, $\ln[p(\mathbf{x}|\theta)]$,

$$\begin{aligned}\mathcal{I}(\theta) = \int \mathrm{d}\mathbf{x} \, \frac{1}{p(\mathbf{x}|\theta)}\left(\frac{\partial p(\mathbf{x}|\theta)}{\partial\theta}\right)^2 &= \int \mathrm{d}\mathbf{x} \, p(\mathbf{x}|\theta)\left(\frac{\partial \ln[p(\mathbf{x}|\theta)]}{\partial\theta}\right)^2 \\ &= \left\langle\left(\frac{\partial \ln[p(\mathbf{x}|\theta)]}{\partial\theta}\right)^2\right\rangle. \end{aligned} \tag{3.15}$$

Physically, the Fisher information represents the average amount of information about the parameter θ that we can access through a measurement for which the outcome probability distribution is $p(\mathbf{x}|\theta)$. From the definition Eq. (3.15) we see that the Fisher information is always positive and from Eq. (3.14) the factor M implies that the Fisher information is additive for independent measurements. These are both important properties for any physical measure of information. Re-arranging Eq. (3.14) and writing the variance as $\mathrm{Var}(\check{\theta})$ leads directly to the Cramér-Rao bound

$$\mathrm{Var}(\check{\theta}) \geq \frac{\left|\frac{\partial\langle\check{\theta}\rangle}{\partial\theta}\right|^2}{M\mathcal{I}(\theta)}, \tag{3.16}$$

where $|\partial_\theta\langle\check{\theta}\rangle|^2$ accounts for a possible difference in units between $\check{\theta}$ and θ. This expression proves that the variance of the estimator $\check{\theta}$ is bounded from below by the Fisher information.

3.1.2 The multi-parameter Cramér-Rao bound

The Cramér-Rao bound as derived in the previous section provides a lower bound for the variance of a single parameter estimator. We will now derive the multi-parameter Cramér-Rao bound that provides a lower bound not only on the variances of the estimators but also on the covariances of the estimators. Starting again from

$$\int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \, p(\mathbf{x}_1|\boldsymbol{\theta}) \ldots p(\mathbf{x}_M|\boldsymbol{\theta})\delta\check{\theta}_i = 0, \tag{3.17}$$

where $\delta\check{\theta}_i = \check{\theta}_i - \langle\check{\theta}_i\rangle$, we now take the derivative with respect to the parameter θ_j. This gives

$$\int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{k=1}^{M} p(\mathbf{x}_k|\boldsymbol{\theta}) \left(\sum_{l=1}^{M} \frac{\partial \ln[p(\mathbf{x}_l|\boldsymbol{\theta})]}{\partial \theta_j} \right) \delta\check{\theta}_i = -\frac{\partial\langle\check{\theta}_i\rangle}{\partial\theta_j}. \tag{3.18}$$

For n parameters Eq. (3.18) defines $n \times n$ equations, writing

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}, \tag{3.19}$$

we can write these in vectorial form

$$\int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{k=1}^{M} p(\mathbf{x}_k|\boldsymbol{\theta}) \left(\sum_{l=1}^{M} \frac{\partial \ln[p(\mathbf{x}_l|\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right) \delta\check{\boldsymbol{\theta}}^{\mathrm{T}} = -\frac{\partial\langle\check{\boldsymbol{\theta}}\rangle}{\partial\boldsymbol{\theta}}. \tag{3.20}$$

Multiplying from the left by $\mathbf{a}^{\mathrm{T}}$ and from the right by $\mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are arbitrary n vectors, allows us to again use the Cauchy-Schwarz inequality. Making the substitutions

$$\begin{aligned} x &= \sum_{l=1}^{M} \mathbf{a}^{\mathrm{T}} \cdot \frac{\partial \ln[p(\mathbf{x}_l|\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}, \\ y &= (\delta\check{\boldsymbol{\theta}})^{\mathrm{T}} \cdot \mathbf{b}, \end{aligned}$$

gives

$$\begin{aligned} \mathbf{a}^{\mathrm{T}} \cdot \int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{k=1}^{M} p(\mathbf{x}_k|\boldsymbol{\theta}) \left(\sum_{l=1}^{M} \frac{\partial \ln[p(\mathbf{x}_l|\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right) \left(\sum_{l'=1}^{M} \frac{\partial \ln[p(\mathbf{x}_{l'}|\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \cdot \mathbf{a} \\ \times \mathbf{b}^{\mathrm{T}} \cdot \int \mathrm{d}\mathbf{x}_1 \ldots \int \mathrm{d}\mathbf{x}_M \prod_{k'=1}^{M} p(\mathbf{x}_{k'}|\boldsymbol{\theta}) (\delta\check{\boldsymbol{\theta}})(\delta\check{\boldsymbol{\theta}})^{\mathrm{T}} \cdot \mathbf{b} \geq \left| \mathbf{a}^{\mathrm{T}} \cdot \frac{\partial\langle\check{\boldsymbol{\theta}}\rangle}{\partial\boldsymbol{\theta}} \cdot \mathbf{b} \right|^2. \end{aligned} \tag{3.21}$$

As in the case for a single parameter, we find that the first term evaluates to 0 unless $l = l'$.

Defining the Fisher information matrix as

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \int \mathrm{d}\mathbf{x}\, p(\mathbf{x}|\boldsymbol{\theta}) \left(\frac{\partial \ln[p(\mathbf{x}|\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ln[p(\mathbf{x}|\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}}, \tag{3.22}$$

with elements

$$[\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]_{ij} = \int \mathrm{d}\mathbf{x}\, p(\mathbf{x}|\boldsymbol{\theta}) \left(\frac{\partial \ln[p(\mathbf{x}|\boldsymbol{\theta})]}{\partial \theta_i} \right) \left(\frac{\partial \ln[p(\mathbf{x}|\boldsymbol{\theta})]}{\partial \theta_j} \right), \tag{3.23}$$

the inequality in Eq. (3.21) now simplifies to

$$M\mathbf{a}^{\mathrm{T}} \cdot \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) \cdot \mathbf{a}\mathbf{b}^{\mathrm{T}} \cdot \langle(\delta\check{\boldsymbol{\theta}})(\delta\check{\boldsymbol{\theta}})^{\mathrm{T}}\rangle \cdot \mathbf{b} \geq \left| \mathbf{a}^{\mathrm{T}} \cdot \frac{\partial\langle\check{\boldsymbol{\theta}}\rangle}{\partial\boldsymbol{\theta}} \cdot \mathbf{b} \right|^2. \tag{3.24}$$

Assuming that the Fisher information matrix can be inverted and since $\mathbf{a}$ is arbitrary, we make the assumption

$$\mathbf{a} = [\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1}]^{\mathrm{T}} \frac{\partial \langle \check{\boldsymbol{\theta}} \rangle}{\partial \boldsymbol{\theta}} \cdot \mathbf{b}, \tag{3.25}$$

which leads to

$$M\mathbf{b}^{\mathrm{T}} \cdot \left[\frac{\partial \langle \check{\boldsymbol{\theta}} \rangle}{\partial \boldsymbol{\theta}} \right]^{\mathrm{T}} \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1} \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) [\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1}]^{\mathrm{T}} \frac{\partial \langle \check{\boldsymbol{\theta}} \rangle}{\partial \boldsymbol{\theta}} \cdot \mathbf{b}\mathbf{b}^{\mathrm{T}} \cdot \langle (\delta\check{\boldsymbol{\theta}})(\delta\check{\boldsymbol{\theta}})^{\mathrm{T}} \rangle \cdot \mathbf{b} \geq \left| \mathbf{b}^{\mathrm{T}} \cdot \left[\frac{\partial \langle \check{\boldsymbol{\theta}} \rangle}{\partial \boldsymbol{\theta}} \right]^{\mathrm{T}} \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1} \frac{\partial \langle \check{\boldsymbol{\theta}} \rangle}{\partial \boldsymbol{\theta}} \cdot \mathbf{b} \right|^2. \tag{3.26}$$

From the definition (3.22) we see that the Fisher information matrix is symmetric, therefore $\boldsymbol{\mathcal{I}}^{\mathrm{T}} = \boldsymbol{\mathcal{I}}$ and

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})[\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1}]^{\mathrm{T}} = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{\mathrm{T}}[\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1}]^{\mathrm{T}} = [\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1}\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]^{\mathrm{T}} = \mathbb{1}. \tag{3.27}$$

The Fisher information matrix is positive definite and therefore so is its inverse. For a set of real parameters the term in the absolute value is real and we find

$$M\mathbf{b}^{\mathrm{T}} \cdot \langle (\delta\check{\boldsymbol{\theta}})(\delta\check{\boldsymbol{\theta}})^{\mathrm{T}} \rangle \cdot \mathbf{b} \geq \mathbf{b}^{\mathrm{T}} \cdot \left[\frac{\partial \langle \check{\boldsymbol{\theta}} \rangle}{\partial \boldsymbol{\theta}} \right]^{\mathrm{T}} \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1} \frac{\partial \langle \check{\boldsymbol{\theta}} \rangle}{\partial \boldsymbol{\theta}} \cdot \mathbf{b}. \tag{3.28}$$

Remembering that $\mathbf{b}$ is arbitrary we have

$$\langle (\delta\check{\boldsymbol{\theta}})(\delta\check{\boldsymbol{\theta}})^{\mathrm{T}} \rangle \geq \frac{1}{M} \left[\frac{\partial \langle \check{\boldsymbol{\theta}} \rangle}{\partial \boldsymbol{\theta}} \right]^{\mathrm{T}} \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1} \frac{\partial \langle \check{\boldsymbol{\theta}} \rangle}{\partial \boldsymbol{\theta}}. \tag{3.29}$$

The term $\langle (\delta\check{\boldsymbol{\theta}})(\delta\check{\boldsymbol{\theta}})^{\mathrm{T}} \rangle$ is the covariance matrix between our estimators $\check{\theta}_1, \ldots, \check{\theta}_n$, whose elements are bounded by

$$\mathrm{Cov}(\check{\theta}_i, \check{\theta}_j) \geq \frac{1}{M} \left(\frac{\partial \langle \check{\theta}_j \rangle}{\partial \theta_i} \frac{\partial \langle \check{\theta}_i \rangle}{\partial \theta_j} \right) [\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1}]_{ij}. \tag{3.30}$$

This is the Cramér-Rao bound for multiple parameters and reduces to Eq. (3.16) when $n = 1$.

3.1.3 Estimating parameters

As previously mentioned, the Cramér-Rao bound places a lower bound on the variance of an estimator. In order to achieve the Cramér-Rao bound a suitable estimator must be chosen. Numerous methods for choosing an estimator exist and in order to ensure that we obtain an estimator that performs well, we must ensure that the right method is applied. Unfortunately there is no single method for choosing an estimator that applies to all situations and we must make a decision of which method to use based on our understanding of the situation. In general the Cramér-Rao bound is dependent on the parameters $\boldsymbol{\theta}$. Any estimator that achieves the Cramér-Rao bound for all values of $\boldsymbol{\theta}$ is called an *efficient* estimator.

We first consider the procedure of maximum likelihood estimation which, due to its simplicity is used in numerous applications. The maximum likelihood procedure

attempts to find the values of the parameters $\boldsymbol{\theta}$ that maximise the log likelihood function $\ln[p(\mathbf{x}|\boldsymbol{\theta})]$. In some circumstances this may be as simple as taking the derivative of the likelihood function and equating it to zero to find the maximum. However, this is not always a straightforward operation and alternative methods for obtaining the maximum likelihood estimator must be used.

The maximum likelihood estimator may also be obtained numerically via an iterative algorithm. The scoring algorithm is one such procedure which proceeds according to the iterative equation [Kay, 1993]

$$\boldsymbol{\mathcal{I}}(\check{\boldsymbol{\theta}}^{(k)})\check{\boldsymbol{\theta}}^{(k+1)} = \boldsymbol{\mathcal{I}}(\check{\boldsymbol{\theta}}^{(k)})\check{\boldsymbol{\theta}}^{(k)} + \left.\frac{\partial \ln[p(\mathbf{x}|\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\check{\boldsymbol{\theta}}^{(k)}}, \tag{3.31}$$

where $\check{\boldsymbol{\theta}}^{(k)}$ is the k^{th} iteration of the estimator $\check{\boldsymbol{\theta}}$, $X(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\check{\boldsymbol{\theta}}^{(k)}}$ denotes evaluation of X at the point $\boldsymbol{\theta} = \check{\boldsymbol{\theta}}^{(k)}$ and $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ is the Fisher information. This method takes an initial guess of the parameter, $\check{\boldsymbol{\theta}}^{(0)}$ and via successive iterations gradually hones in on the actual values of parameters $\boldsymbol{\theta}$. When evaluating the Cramér-Rao bound, if we calculate the Fisher information via Eq. (3.22) then the term $\partial_{\boldsymbol{\theta}} \ln[p(\mathbf{x}|\boldsymbol{\theta})]$ has already been determined and the scoring algorithm proceeds without difficulty.

3.2 Quantum estimation theory

In the previous section we demonstrated how the variance of an unbiased estimator can be bounded from below by the Fisher information via the Cramér-Rao bound. The probabilistic nature of the data we measure only allows us to make statistical inferences about any parameters on which the data depend. At this stage a natural question arises: *what happens when the probabilistic nature of the data is a direct consequence of quantum uncertainty as opposed to specific detection uncertainties?* In this section we show that the quantum Fisher information, $\boldsymbol{\mathcal{I}}_Q$, provides a natural generalisation to the Fisher information of the previous section, which shall henceforth be referred to as the classical Fisher information or just $\boldsymbol{\mathcal{I}}$.

3.2.1 The quantum Fisher information

Although not the first researchers to define the quantum Fisher information, Braunstein and Caves demonstrated that the quantum Fisher information can be considered as the solution to two distinct optimization problems [Braunstein and Caves, 1994]. By considering the discrimination of neighbouring quantum states as a parameter estimation problem where the parameter we wish to measure, θ, defines a path through the state space of the system, they showed that the optimal performance of any estimator for θ is given by the quantum Fisher information. The quantum Fisher information is therefore independent of the measurement procedure chosen and is in fact only a function of the state itself. It therefore provides a lower bound which is dictated by the geometry of state space for the system and therefore cannot be overcome without modifying the state itself. In some cases, modification of the state is not an issue and we can seek to find the best state to optimise measurements of θ, however in some circumstances this is not possible and the best we can do is seek a measurement with achieves the minimum variance via the quantum Cramér-Rao bound.

We start with the single parameter version of the classical Fisher information as given by Eq. (3.15) only now we use the Born rule to find the probability distribution $p(x|\theta)$

$$p(x|\theta) = \mathrm{Tr}\Big[\rho(\theta)\hat{E}_x\Big], \tag{3.32}$$

where $\rho(\theta)$ is the state which depends on the parameter θ that we want to estimate and $\hat{E}_x$ is the POVM element describing the measurement which produces outcome x. A detailed account of the use of POVMs is given in section 4.1. To calculate the Fisher information we need to take the derivative of this probability distribution with respect to the parameter θ. Taking the derivative of Eq. (3.32) gives

$$\frac{\partial}{\partial\theta}p(x|\theta) = \mathrm{Tr}\left[\frac{\partial}{\partial\theta}\rho\hat{E}_x\right] \equiv \mathrm{Tr}\Big[\partial_\theta\rho\hat{E}_x\Big], \tag{3.33}$$

where we have dropped the ρ dependence on θ and henceforth this shall be assumed. We now introduce the symmetric logarithmic derivative super-operator (SLD) implicitly defined by the equation

$$\rho' = \frac{\mathscr{L}_\rho(\rho')\rho + \rho\mathscr{L}_\rho(\rho')}{2}, \tag{3.34}$$

where $\mathscr{L}_\rho(\rho')$ is the symmetric logarithmic derivative and we have defined $\rho' = \partial_\theta\rho$. Eq. (3.34) takes the form of a Lyapunov equation, for which the solution is [Paris, 2009]

$$\mathscr{L}_\rho(\rho') = 2\int_0^\infty \mathrm{d}t\, \mathrm{e}^{-\rho t}\rho'\mathrm{e}^{-\rho t}. \tag{3.35}$$

In the basis that diagonalises ρ we find

$$\mathscr{L}_\rho(\rho') = 2\sum_{i,j}\frac{\rho'_{ij}}{p_i + p_j}|i\rangle\langle j|, \tag{3.36}$$

where p_i and p_j are the eigenvalues of ρ and the sum is taken over all i, j that satisfy $p_{i,j} \neq 0$.

Using the fact that ρ is Hermitian, and therefore $p_{i,j}$ are real, we see that the symmetric logarithmic derivative is also Hermitian

$$\mathscr{L}^\dagger_\rho(\rho') = 2\sum_{i,j}\frac{(\rho'_{ij})^*}{p_i + p_j}|j\rangle\langle i| = 2\sum_{i,j}\frac{\rho'_{ji}}{p_i + p_j}|j\rangle\langle i| = \mathscr{L}_\rho(\rho'). \tag{3.37}$$

Inserting the form of the derivative given by the SLD, Eq. (3.34), into Eq. (3.33) gives us

$$\partial_\theta p(x|\theta) = \frac{1}{2}\mathrm{Tr}\Big[\rho\hat{E}_x\mathscr{L}_\rho(\rho')\Big] + \frac{1}{2}\mathrm{Tr}\Big[\rho\mathscr{L}_\rho(\rho')\hat{E}_x\Big], \tag{3.38}$$

and using the property, $\mathrm{Tr}[ABC] = (\mathrm{Tr}[ACB])^*$ for Hermitian A, B and C [1], we find

$$\partial_\theta p(x|\theta) = \mathrm{Re}\left\{\mathrm{Tr}\left[\rho\hat{E}_x\mathscr{L}_\rho(\rho')\right]\right\}. \tag{3.39}$$

From the definition of the Fisher information Eq. (3.15) and the expressions just obtained, we find that the classical Fisher information for a measurement of the operator $\hat{E}_x$ on the state ρ

$$\mathcal{I} = \int \mathrm{d}x\, \frac{\mathrm{Re}\{\mathrm{Tr}[\rho\hat{E}_x\mathscr{L}_\rho(\rho')]\}^2}{\mathrm{Tr}[\rho\hat{E}_x]}. \tag{3.40}$$

We now want to maximise the classical Fisher information over all possible measurements $\hat{E}_x$. To do this we first find a measurement-independent upper bound for $\mathcal{I}$ and then show that this upper bound is achievable for an appropriate choice of $\hat{E}_x$. We therefore will conclude that the Fisher information found is an intrinsic property of the state ρ meaning that it provides a lower bound for the variance of an estimator for θ (via the Cramér-Rao bound) which a change of $\hat{E}_x$ cannot overcome.

To find the measurement independent form of $\mathcal{I}$ we first make use of the fact that for complex α, $\mathrm{Re}\{\alpha\}^2 \leq |\alpha|^2$, where equality implies that α is real. This allows us to express the Fisher information in Eq. (3.40) as

$$\begin{aligned}\mathcal{I} &\leq \int \mathrm{d}x\, \frac{|\mathrm{Tr}[\rho\hat{E}_x\mathscr{L}_\rho(\rho')]|^2}{\mathrm{Tr}[\rho\hat{E}_x]} \\ &= \int \mathrm{d}x \left|\frac{\mathrm{Tr}[\rho\hat{E}_x\mathscr{L}_\rho(\rho')]}{\mathrm{Tr}[\rho\hat{E}_x]^{\frac{1}{2}}}\right|^2,\end{aligned} \tag{3.41}$$

where equality holds iff $\mathrm{Im}\{\mathrm{Tr}[\rho\hat{E}_x\mathscr{L}_\rho(\rho')]\} = 0$. Now, from the Cauchy-Schwarz inequality we find

$$\begin{aligned}\int \mathrm{d}x \left|\frac{\mathrm{Tr}[\rho\hat{E}_x\mathscr{L}_\rho(\rho')]}{\mathrm{Tr}[\rho\hat{E}_x]^{\frac{1}{2}}}\right|^2 &= \int \mathrm{d}x \left|\frac{\mathrm{Tr}[\rho^{\frac{1}{2}}\hat{E}_x^{\frac{1}{2}}\hat{E}_x^{\frac{1}{2}}\mathscr{L}_\rho(\rho')\rho^{\frac{1}{2}}]}{\mathrm{Tr}[\rho\hat{E}_x]^{\frac{1}{2}}}\right|^2 \\ &= \int \mathrm{d}x \left|\mathrm{Tr}\left[\frac{\rho^{\frac{1}{2}}\hat{E}_x^{\frac{1}{2}}}{\mathrm{Tr}[\rho\hat{E}_x]^{\frac{1}{2}}}\hat{E}_x^{\frac{1}{2}}\mathscr{L}_\rho(\rho')\rho^{\frac{1}{2}}\right]\right|^2 \\ &\leq \int \mathrm{d}x\, \mathrm{Tr}\left[\hat{E}_x^{\frac{1}{2}}\mathscr{L}_\rho(\rho')\rho^{\frac{1}{2}}(\hat{E}_x^{\frac{1}{2}}\mathscr{L}_\rho(\rho')\rho^{\frac{1}{2}})^\dagger\right] \qquad (3.42a) \\ &= \int \mathrm{d}x\, \mathrm{Tr}\left[\rho\mathscr{L}_\rho(\rho')\hat{E}_x\mathscr{L}_\rho(\rho')\right] \\ &= \mathrm{Tr}\left[\rho\mathscr{L}_\rho^2(\rho')\right]. \qquad (3.42b)\end{aligned}$$

[1] To prove this assertion we start from right hand side and work back. First we write the left hand side in index notation

$$(\mathrm{Tr}[ACB])^* = \sum_{i,j,k}(A_{ij}C_{jk}B_{ki})^* = \sum_{i,j,k} A^*_{ij}C^*_{jk}B^*_{ki}.$$

Now using the fact the A, B, and C are Hermitian we find

$$(\mathrm{Tr}[ACB])^* = \sum_{i,j,k} A_{ji}C_{kj}B_{ik} = \sum_{i,j,k} A_{ji}B_{ik}C_{kj} = \mathrm{Tr}[ABC],$$

and we see that $(\mathrm{Tr}[ACB])^* = \mathrm{Tr}[ABC]$ as we set out to prove.

In order to obtain equality in Eq. (3.42a) requires that the equation

$$\frac{\hat{E}_x^{\frac{1}{2}}\rho^{\frac{1}{2}}}{\mathrm{Tr}[\rho\hat{E}_x]} = \frac{\hat{E}_x^{\frac{1}{2}}\mathscr{L}_\rho(\rho')\rho^{\frac{1}{2}}}{\mathrm{Tr}[\rho\hat{E}_x\mathscr{L}_\rho(\rho')]}, \tag{3.43}$$

is satisfied for all values of θ [Paris, 2009]. This is achieved when the operator $\hat{E}_x$ is a one dimensional projector that projects onto a complete set of eigenstates of $\mathscr{L}_\rho(\rho')$ [Braunstein et al., 1996].

The upper bound for $\mathcal{I}$ is therefore an achievable. The quantum Fisher information is therefore given by

$$\mathcal{I}_Q = \mathrm{Tr}\left[\rho\mathscr{L}_\rho^2(\rho')\right], \tag{3.44}$$

or by substituting from Eq. (3.34) into Eq. (3.42b)

$$\mathrm{Tr}\left[\rho\mathscr{L}_\rho^2(\rho')\right] = 2\mathrm{Tr}[\rho'\mathscr{L}_\rho(\rho')] - \mathrm{Tr}\left[\rho\mathscr{L}_\rho^2(\rho')\right] \implies \mathcal{I}_Q = \mathrm{Tr}[\rho'\mathscr{L}_\rho(\rho')]. \tag{3.45}$$

We can now lower bound the variance of any estimator of θ by the quantum Cramér-Rao bound

$$\mathrm{Var}(\check{\theta}) \geq \frac{\frac{\partial\langle\check{\theta}\rangle}{\partial\theta}}{M\mathcal{I}_Q}, \tag{3.46}$$

which is achievable for a measurement satisfying the conditions given above.

As pointed out by various authors [Braunstein and Caves, 1994, Braunstein et al., 1996, Paris, 2009], the quantum Fisher information adheres to a geometrical interpretation. The quantum Fisher information is proportional to the Bures metric, which is used to calculate distances in the state space of mixed quantum systems. When viewed as a metric the quantum Fisher information tells us how quickly an infinitesimal change of the parameter moves the family of states parameterised by θ through the space. A large quantum Fisher information implies that small changes of the parameter can move the state to a new state far apart from the initial one. The further apart these states live in the state space the more distinguishable they are and therefore the easier it is to tell them apart in an experiment. If small changes in the parameter can easily be determined then the parameter can be measured with high precision, this is the basis of the quantum Cramér-Rao bound.

3.2.2 The multi-parameter quantum Cramér-Rao bound

Just as in the classical treatment, the quantum theory of parameter estimation can also be extended to the problem of estimating multiple parameters simultaneously. Again writing the n parameters to be estimated as $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^{\mathrm{T}}$ we find that the bound for the covariance matrix of our estimators is

$$\mathrm{Cov}(\check{\theta}_i, \check{\theta}_j) \geq \frac{1}{M}\left(\frac{\partial\langle\check{\theta}_j\rangle}{\partial\theta_i}\frac{\partial\langle\check{\theta}_i\rangle}{\partial\theta_j}\right)[\boldsymbol{\mathcal{I}}_Q^{-1}]_{ij}, \tag{3.47}$$

where the elements of the quantum Fisher information matrix are given by

$$[\boldsymbol{\mathcal{I}}_Q]_{ij} = \mathrm{Tr}[\rho(\mathscr{L}_i\mathscr{L}_j + \mathscr{L}_j\mathscr{L}_i)], \tag{3.48}$$

and the SLD for the parameter θ_i is defined via the equation

$$\frac{\partial \rho}{\partial \theta_i} = \frac{\mathscr{L}_i \rho + \rho \mathscr{L}_i}{2}. \tag{3.49}$$

The subscript and argument have been dropped from $\mathscr{L}$ for brevity. Using the definition Eq. (3.49) in Eq. (3.48) we find

$$[\boldsymbol{\mathcal{I}}_Q]_{ij} = \mathrm{Tr}[(\partial_i \rho)\mathscr{L}_j], \tag{3.50}$$

where $\partial_i \rho \equiv \frac{\partial}{\partial \theta_i}\rho$.

In general the non-commutativity of quantum mechanics renders this bound unachievable except when [Monras and Illuminati, 2011]

$$\mathrm{Tr}[\rho[\mathscr{L}_i, \mathscr{L}_j]] = 0. \tag{3.51}$$

Finding the SLD for a set of parameters therefore allows us to determine if we can perform a joint efficient estimation and, from the previous section, allows us to find the optimal estimator. For these reasons determining the SLD is of paramount importance for a general quantum parameter estimation problem.

Chapter 4

Fundamental Limits of Classical and Quantum Imaging

This chapter is largely based on [Pérez-Delgado et al., 2012]. In this chapter we develop a theoretical framework that will allow us to calculate the Fisher information, and also the resolution, for a number of well known experimental arrangements. We find that the theory allows us to derive a number of resolution limits up to a multiplicative factor, where the criterion for resolution comes directly from the distinguishability of probability distributions. We also show how this model is well equipped to deal with imperfect measurements and demonstrate the effect of certain imperfections on the resolution. The general nature of this approach allows for wide application, of which a small subclass of problems is considered here.

4.1 Positive Operator Valued Measures

In the forthcoming sections we will make extensive use of imperfect measurement devices. To ensure that we can accurately describe the action of such devices we give a brief overview of the quantum description of imperfect measurements. Imperfect measurements are incorporated into quantum measurement theory in much the same way the density operators are incorporated to describe imperfect knowledge about a quantum state. Whereas perfect measurements are described by projection operators of the form $|n\rangle\langle n|$, an imperfect measurement is described by a positive operator-valued measure (POVM) with elements given by [Helstrom, 1976]

$$\hat{E}_k = \sum_n q_k(n)|n\rangle\langle n|, \tag{4.1}$$

where $q_k(n)$ is the conditional probability of obtaining measurement outcome k given the state $|n\rangle$. We notice from Eq. (4.1) that multiple states can lead to the same measurement outcome. Also, the same input state can lead to different measurement outcomes. The set of $\hat{E}_k$ for all measurement outcomes k form the POVM and any valid POVM must have the following three properties [Nielsen and Chuang, 2000]

$$\hat{E}_k \geq 0, \tag{4.2a}$$

$$\sum_k \hat{E}_k = \mathbb{I}, \tag{4.2b}$$

$$\hat{E}_k^\dagger = \hat{E}_k, \tag{4.2c}$$

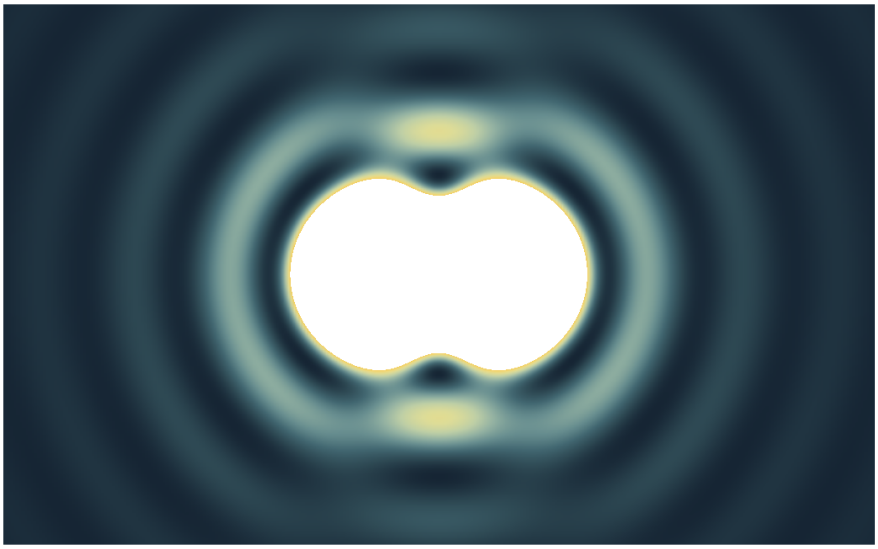

Figure 4.1: The intensity distribution for two point sources as viewed through a circular aperture. The diffraction of light by the aperture causes the point intensity profile to spread.

whereby $\hat{E}_k \geq 0$ is a shorthand way of writing $z^{\mathrm{T}}\hat{E}_k z \geq 0$ for all real, non-zero column vectors z. A POVM is uniquely defined by the set of conditional probabilities $q_k(n)$ and a choice of basis $|n\rangle$. In the following sections we shall therefore seek to determine the conditional probabilities such that the desired POVMs are described.

4.2 Image resolution

The resolution of an image is defined as the smallest feature details that can be obtained from the image. Statements about the resolution limit for an imaging system usually involve the following definition:

> *Two points in the object can be reliably resolved if the maxima from one of the point spread functions falls beyond the first minimum from the other.*

The point spread function describes how a point in the object plane blurs into a dot on the image plane due to the diffraction of the light by the imaging system. Fig. 4.1 shows the overlap of two point spread functions defined by a circular aperture. The two points in the figure are just resolved as their principle maxima fall on the first minima of the other distribution. In the following sections we shall consider a more general definition of resolution based on the distinguishability of probability distributions. The generality of this approach allows our method to be utilised in a wide range of physically interesting situations.

4.3 The imaging observable

Fig. 4.2 shows a general imaging system complete with detection apparatus for measuring the state $\rho(\theta)$ which depends on the parameter θ that we wish to measure.

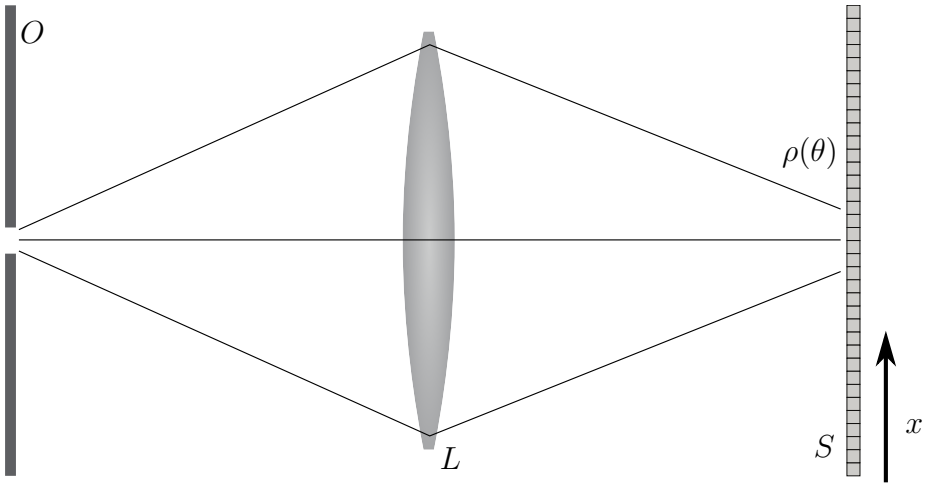

Figure 4.2: General imaging system. Light from the object O is manipulated by some optical system L and is then incident on the substrate S. The substrate makes a measurement (possibly imperfect) of the state $\rho(\theta)$ which depends upon the parameter we wish to measure θ.

We consider the detector, S, to be granular, consisting of N pixels of length ℓ making the total length of the substrate $L = N\ell$. The pixels are distinguished by their position x on the substrate. We denote the image intensity by i_k and the probability of measuring i_k at position x is given by $p_k(x)$. The average intensity recorded at position x is given by $I(x) = \sum_k i_k p_k(x)$. We can calculate the probabilities $p_k(x)$ from the Born rule $p_k(x) = \mathrm{Tr}[\rho(\theta)\hat{E}_k(x)]$ where $\hat{E}_k(x)$ is the measurement operator associated with the outcome i_k. The set of operators $\hat{E}_k(x)$ form a positive operator-valued measure (POVM) for each pixel, $\sum_k \hat{E}_k(x) = \mathbb{I}(x)$.

The image $I(x)$ need not be restricted to the expectation value of the intensity at pixel x. In principle any observable quantity can be identified as the "image", for example we may choose as our observable the equal-time, n^{th} order intensity correlation function $G^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ [Oppel et al., 2012]. Using the Born rule and the definition of the image we find

$$I(x) = \sum_k i_k \mathrm{Tr}\Big[\rho(\theta)\hat{E}_k(x)\Big] = \mathrm{Tr}\left[\sum_k i_k \rho(\theta)\hat{E}_k(x)\right] \equiv \mathrm{Tr}\Big[\rho(\theta)\hat{I}(x)\Big], \tag{4.3}$$

where we have defined the imaging operator to be $\hat{I}(x) = \sum_k i_k \hat{E}_k(x)$. In the simplest case, the imaging operator simply records the intensity at the position x, which is directly proportional to the photon number k. The POVM for such a substrate is simply given by $\hat{E}_k(x) = |k\rangle_x\langle k|$ and $i_k = k$. Later we will make use of more complicated POVM's when we wish to consider the effects of imperfect measurements.

Typically, measurements in imaging can be described by operators that are diagonal in the Fock basis since they usually consist of photon counts at each pixel.

This allows us to write the POVM as

$$\begin{aligned}\hat{E}_k(x) &= \sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} q_k(x|n_1,\ldots,n_N)|n_1\rangle_1\langle n_1|\otimes\cdots\otimes|n_N\rangle_N\langle n_N| \\ &\equiv \sum_{\vec{n}} q_k(x|\vec{n})|\vec{n}\rangle\langle\vec{n}|, \end{aligned} \tag{4.4}$$

where $\sum_{\vec{n}} = \sum_{n_1}\cdots\sum_{n_N}$, $|n_i\rangle_i$ being the n_i photon Fock state at position i, and $|\vec{n}\rangle = |n_1\rangle_1\otimes\cdots\otimes|n_N\rangle_N$ and $q_k(x|\vec{n})$ is the conditional probability of getting outcome k at position x given the photon number distribution $\vec{n} = (n_1,\ldots,n_N)$. We now identify the image with an unnormalised probability distribution. We can normalise $I(x)$ to define the probability distribution $\Pr(x)$

$$\Pr(x) \equiv \frac{I(x)}{\sum_{x'} I(x')} = \frac{I(x)}{I_0}, \tag{4.5}$$

where we have divided the image value at position x by the total image I_0 so that $\Pr(x)$ can be truly identified as a probability. As expected for a probability distribution $\sum_x \Pr(x) = 1$. It is not difficult to see that $\Pr(x)$ is positive real and σ-additive and therefore can justifiably be referred to as a probability distribution. By defining our probability distribution in this way we become insensitive to statistical fluctuations of the intensity at various points and also the effects of loss. Effectively we are only sensitive to variations in the random variable x that represents the position at which the signal is measured

In imaging the aim is to measure some or all of the spatial properties of the source, thus gaining information about the spatial distribution of the source. Parameterising these spatial properties by θ we can write the conditional probability of measuring outcome x given the parameter θ as $\Pr(x|\theta)$. To distinguish between two possible images is then formally equivalent to distinguishing two probability distributions $\Pr(x|\theta = 0)$ and $\Pr(x|\theta)$ where we have arbitrarily set $\theta = 0$ for one of the distributions. To make the discussion of probability distribution distinguishability more formal, we introduce the statistical distance, defined in infinitesimal form as [Braunstein and Caves, 1994]

$$\mathrm{d}s^2 = \sum_x \frac{\mathrm{d}\Pr(x|\theta)^2}{\Pr(x|\theta)} = \sum_x \frac{[\Pr(x|\mathrm{d}\theta) - \Pr(x|0)]^2}{\Pr(x|\theta)}. \tag{4.6}$$

Integrating Eq. (4.6) leads to an expression for $s(\theta)$, Taylor expanding $s(\theta)$ to first order we find

$$\delta s \equiv s(\delta\theta) - s(0) = \left.\frac{\mathrm{d}s}{\mathrm{d}\theta}\right|_{\theta=0}\delta\theta. \tag{4.7}$$

Two probability distributions are distinguishable if $(\delta s)^2 \geq 1$ [Wootters, 1981]. Using the definition of the Fisher information Eq. (3.15) we can rewrite this condition as

$$\mathcal{I}(0)(\delta\theta)^2 \geq 1. \tag{4.8}$$

This bears similarity to the Cramér-Rao bound but it is in fact our distinguishability criterion. Here the Fisher information is strictly being used as a metric on the space

of probabilities that are defined by Eq. (4.5) and as such should not be confused with the Fisher information associated with a measurement of the intensity along the substrate. Whilst we would expect the Fisher information for such a measurement to depend on the noise, the quantity $\mathcal{I}$ defined here is independent of such noise due to the normalisation in Eq. (4.5). Using the probability distribution given in Eq. (4.5) to calculate the Fisher information leads to

$$\begin{aligned}\mathcal{I}(\theta) &= \sum_x \frac{I_0}{\langle\hat{I}(x)\rangle}\left[\frac{1}{I_0}\frac{\partial\langle\hat{I}(x)\rangle}{\partial\theta} - \frac{\langle\hat{I}(x)\rangle}{I_0^2}\frac{\partial I_0}{\partial\theta}\right]^2 \\ &= \sum_x \left[\frac{1}{\langle\hat{I}(x)\rangle I_0}\left(\frac{\partial\langle\hat{I}(x)\rangle}{\partial\theta}\right)^2 - \frac{2}{I_0^2}\frac{\partial I_0}{\partial\theta}\frac{\partial\langle\hat{I}(x)\rangle}{\partial\theta} + \frac{\langle\hat{I}(x)\rangle}{I_0^3}\left(\frac{\partial I_0}{\partial\theta}\right)^2\right] \\ &= \sum_x \frac{1}{\langle\hat{I}(x)\rangle I_0}\left(\frac{\partial\langle\hat{I}(x)\rangle}{\partial\theta}\right)^2 - \frac{1}{I_0^2}\left(\frac{\partial I_0}{\partial\theta}\right)^2, \end{aligned} \tag{4.9}$$

where we have used

$$\sum_x \frac{\partial\langle\hat{I}(x)\rangle}{\partial\theta} = \frac{\partial}{\partial\theta}\sum_x\langle\hat{I}(x)\rangle = \frac{\partial I_0}{\partial\theta}. \tag{4.10}$$

Now we have an expression for the Fisher information in terms of the image $\langle\hat{I}(x)\rangle$ and the first derivative of the image with respect to θ, which we can use along with Eq. (4.8) to determine the statistical resolution.

4.4 Examples

We now use the theory outlined above to show that the results are consistent with well known resolution limits. First we consider the double slit experiment, in which the image is the interference pattern in the far field

$$\langle\hat{I}(x)\rangle = \cos^2\left(\frac{k\theta x}{2R}\right), \tag{4.11}$$

where k is the wavenumber of the radiation used and R is the distance from the slit to the image plane. Taking the limit of a continuous detection medium, we find the probability distribution $\Pr(x|\theta)$ to be

$$\Pr(x|\theta) = \frac{k\theta\cos^2\left(\frac{k\theta x}{2R}\right)}{k\theta\mathcal{A} + \sin(k\theta\mathcal{A})} \tag{4.12}$$

where $\mathcal{A} = x_{\max}/R$ is the numerical aperture of the system [Lipson et al., 2011]. From this we derive the Fisher information

$$\mathcal{I}(\theta) = \frac{1}{\theta^2} - \frac{2\mathcal{A}^3\theta k^3}{3[\mathcal{A}\theta k + \sin(\mathcal{A}\theta k)]} + \frac{\mathcal{A}^2k^2\left[\mathcal{A}^2\theta^2k^2 - 2\cos(\mathcal{A}\theta k) - 2\right]}{[\mathcal{A}\theta k + \sin(\mathcal{A}\theta k)]^2}. \tag{4.13}$$

In order to resolve the two slits the distance θ between them must be greater than the uncertainty $\delta\theta$. We therefore find that the two slits can be resolved if $\mathcal{I}(\theta)\theta^2 \geq 1$. Multiplying Eq. (4.13) by θ^2 and making the substitution $z = k\theta\mathcal{A}$ we find

$$\mathcal{I}(\theta)\theta^2 = 1 - \frac{2z^3}{3(z+\sin z)} + \frac{z^2(z^2 - 2\cos z - 2)}{(z+\sin z)^2}. \tag{4.14}$$

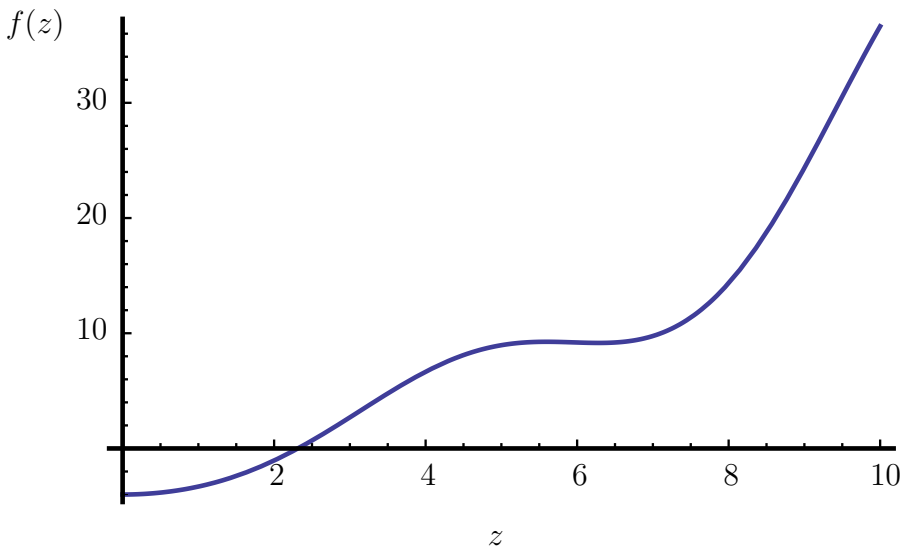

Figure 4.3: The function $f(z)$ plotted against the dimensionless quantity z. The parameter θ is resolved when the curve is greater than or equal to zero.

Demanding $\mathcal{I}(\theta)\theta^2 \geq 1$ leads to

$$f(z) = \frac{z^2}{3} - \frac{2}{3}z\sin z - 2\cos z - 2 \geq 0, \tag{4.15}$$

which we plot in Fig. 4.3. Solving numerically we find $z \geq 2.31$, or

$$\theta_{\min} \geq \frac{2.31}{k\mathcal{A}} = \frac{0.369\lambda}{\mathcal{A}}. \tag{4.16}$$

We find that our result approximately reproduces the Abbe limit, $\theta_{\min} \geq 0.5\lambda/\mathcal{A}$. However, whilst the Abbe limit is a statement about distinguishing two intensity peaks based on the overlap between them, the resolution limit presented here is a precise mathematical statement about the distinguishability of the probability distribution parameterised by θ.

Next, we consider classical and quantum photolithography [Boto et al., 2000] as shown in Fig. 4.4. Boto *et al.* demonstrated that by using an M-photon absorption substrate and entangled NOON states, it is possible to achieve an increase in the resolution. An M-photon detection substrate is a binary detector with possible outputs "click" (M-photons detected) and "no-click" (M-photons not detected). For a general M-photon absorption substrate with single photon detection efficiency η the conditional probabilities are

$$q_0(x|n_x) = \begin{cases} 1 & \text{if } n_x < M \\ (1-\eta)^{n_x} & \text{if } n_x \geq M \end{cases}, \tag{4.17}$$

and $q_1(x|n_x) = 1 - q_0(x|n_x)$, where the outcomes 0 and 1 correspond to "no-click" and "click" in the detectors, respectively. In the limit where the two beams just graze the surface of the substrate, the two beams become counter propagating and we can write $k_a = -k_b \equiv k$. The mode operators for the two beams are given by

$$\hat{a} = \frac{1}{\sqrt{N}}\sum_x \mathrm{e}^{ik\ell x}\hat{a}_x \qquad \text{and} \qquad \hat{b} = \frac{1}{\sqrt{N}}\sum_x \mathrm{e}^{-ik\ell x}\hat{b}_x, \tag{4.18}$$

which we write in terms of the spatially localised modes $\hat{a}_x$ and $\hat{b}_x$, N being the number of pixels.

The mode operator at position x on the substrate is

$$\hat{e}_x = \frac{\hat{a}_x + \hat{b}_x}{\sqrt{2}}. \tag{4.19}$$

For classical, single photon lithography the POVM elements $\hat{E}_0(x)$ and $\hat{E}_1(x)$ are

$$\hat{E}_0(x) = \sum_{n_x=0}^{\infty} (1-\eta)^{n_x} |n_x\rangle_x\langle n_x| \qquad \text{and} \qquad \hat{E}_1(x) = \sum_{n_x=0}^{\infty} [1-(1-\eta)^{n_x}] |n_x\rangle_x\langle n_x|, \tag{4.20}$$

where $|n_x\rangle_x$ are the Fock basis for the mode $\hat{e}_x$, i.e. $\hat{e}_x|n_x\rangle_x = \sqrt{n_x}|n_x - 1\rangle_x$.

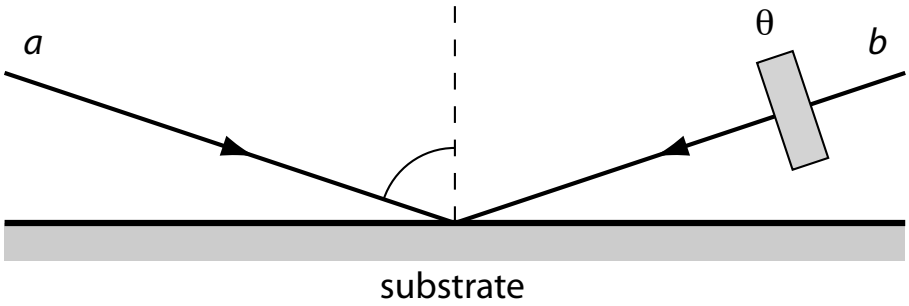

Figure 4.4: Photolithography. The two nearly counter propagating beams a and b interfere producing a pattern on the substrate. The phase shift θ translates the image along the substrate.

To simplify the notation slightly we will drop the x subscript on the bras and kets. Defining the measurement outcomes to be $i_0 = 0$ and $i_1 = 1$ without loss of generality, we find the image is

$$\langle \hat{I}(x) \rangle = \sum_{n_x=0}^{\infty} [1-(1-\eta)^{n_x}] \langle n_x|\rho|n_x\rangle, \tag{4.21}$$

where the term $\langle n_x|\rho|n_x\rangle$, becomes $|\langle n_x|\psi\rangle|^2$ for pure states.

In the classical case the input state is the single photon state

$$|\psi\rangle = \frac{|1,0\rangle_{ab} + \mathrm{e}^{i\theta}|0,1\rangle_{ab}}{\sqrt{2}}. \tag{4.22}$$

It might at first seem strange to identify the highly path entangled state $|\psi\rangle$ with a classical state, however we note that in the canonical example of a classical state, the coherent state, all the photons are completely statistically independent. Therefore the action of a beamsplitter on a coherent state is the sum of the action the action on the individual photons. It is this exact principle that creates the correspondence between the coherent illumination double slit experiment and the single photon double slit experiment and is the origin of Dirac's famous remark: *Each photon interferes only with itself. Interference between two different photons never occurs*

[Dirac, 1989]. The action of a beamsplitter on a single photon state gives Eq. (4.22), which gives

$$|\langle n_x|\psi\rangle|^2 = \frac{1}{N}\cos^2\left(k\ell x + \frac{\theta}{2}\right)\delta_{1,n_x}. \tag{4.23}$$

The image in the case of single photon lithography then becomes

$$\langle\hat{I}(x)\rangle = \frac{\eta}{N}\cos^2\left(k\ell x + \frac{\theta}{2}\right). \tag{4.24}$$

In the limit of large N the normalisation I_0 tends to $\eta/2$, we then calculate the Fisher information to be

$$\begin{aligned}\mathcal{I}(\theta) &= \frac{2}{N}\sum_x \sin^2\left(k\ell x + \frac{\theta}{2}\right)\\ &= \frac{1}{N}\left(N - \cos[k\ell(N+1)+\theta]\frac{\sin(Nk\ell)}{\sin(k\ell)}\right)\\ &\simeq (1-\cos[kL+\theta]\,\mathrm{sinc}(kL))\\ &\leq 2,\end{aligned} \tag{4.25}$$

where we see that the η dependence has vanished as a consequence of the normalisation discussed earlier. Assuming the wavelength to be much smaller than the detection length $L \gg \lambda$, we find $\mathrm{sinc}(kL) \sim 0$ and $\mathcal{I}(\theta) \sim 1$. The resolution is then given by $(\delta\theta)^2 \geq 1/\mathcal{I}(\theta)$. Since the parameter θ translates the pattern along the substrate we can write $\delta\theta = k\ell\delta x$ leading to $\ell\delta x \geq \lambda/2\pi$. This is in close agreement to the Rayleigh criterion $\ell\delta x \geq \lambda/4$ which does not give a rigorous definition of resolution but simply compares the maxima and minima of the image distribution [Lipson et al., 2011].

Next we consider the case of quantum lithography which utilises NOON states and M-photon absorption substrates to achieve a higher resolution. We take as our input state the M-photon NOON state

$$|\psi\rangle = \frac{|M,0\rangle_{ab} + |0,M\rangle_{ab}}{\sqrt{2}}, \tag{4.26}$$

for which we find

$$|\langle n_x|\psi\rangle|^2 = \frac{2}{(2N)^M}\cos^2\left(k\ell M x + \frac{M\theta}{2}\right)\delta_{M,n_x}. \tag{4.27}$$

We see the factor M in the argument of the $\cos^2$ that is the origin of the resolution enhancement in quantum lithography. Again making the choice $i_0 = 0$ and $i_1 = 1$ we find our image is given by

$$\langle\hat{I}(x)\rangle = 2\left(\frac{\eta}{2N}\right)^M\cos^2\left(k\ell M x + \frac{M\theta}{2}\right), \tag{4.28}$$

and in the limit of large N we find $I_0 = N(\eta/2N)^M$. From this we find that the

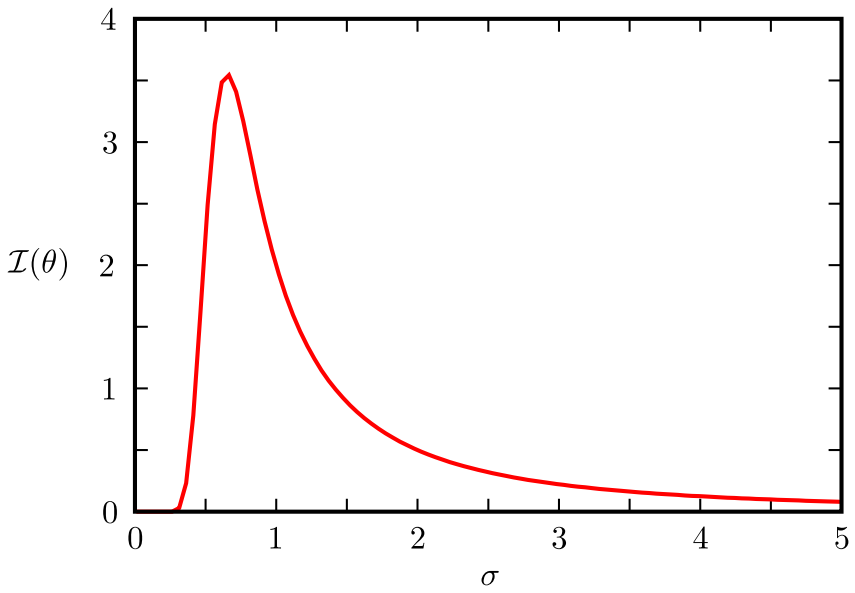

Figure 4.5: The Fisher information as a function of σ.

Fisher information is

$$\begin{aligned}\mathcal{I}(\theta) &= \frac{2M^2}{N}\sum_x \sin^2\left(k\ell Mx + \frac{M\theta}{2}\right) \\ &= \frac{M^2}{N}\left(N - \cos[Mk\ell(N+1) + M\theta]\frac{\sin(MNk\ell)}{\sin(Mk\ell)}\right) \\ &\simeq M^2\left(1 - \frac{1}{M}\cos[MkL + M\theta]\,\mathrm{sinc}(MkL)\right) \\ &\leq 2M^2. \end{aligned} \tag{4.29}$$

Again taking the limit $L \gg \lambda$, we find $\mathcal{I}(\theta) \sim M^2$ which leads to the resolution $\ell\delta x \geq \lambda/2\pi M$, where the M-fold enhancement of the resolution as predicted by Boto et al. is clearly reproduced.

4.5 Detector imperfections

Next we demonstrate how imperfect measurements can be incorporated into the model allowing us to calculate their effects on the resolution. We take as our state a coherent state with a Gaussian intensity profile across the detector

$$|\psi\rangle = \bigotimes_{x=1}^{N} |\alpha_x\rangle_x, \tag{4.30}$$

where $|\alpha_x\rangle_x$ is a coherent state of amplitude $\alpha_x = \alpha_0 \exp[-(x-\theta)^2/2\sigma^2]$. For simplicity we have assumed no coherence between the pixels and we also assume $\sigma \ll L$ and that θ is towards the centre of the substrate. This allows us to make the approximation $\partial_\theta I_0 \approx 0$. Modelling the substrate as a perfect photodetector, our image is given by

$$\langle \hat{I}(x)\rangle = |\alpha_x|^2, \tag{4.31}$$

and the normalisation $I_0 = \sum_x |\alpha_x|^2$. The Fisher information is

$$\mathcal{I}(\theta) = 4 \sum_x \frac{|\alpha_x|^2}{I_0} \frac{(x-\theta)^2}{\sigma^4}. \tag{4.32}$$

The Fisher information is shown as a function of σ in Fig. 4.5. We see from Fig. 4.5 that the Fisher information goes to zero as σ goes to zero since, for fixed α_0, the signal disappears in this limit. To determine the form of $I(\theta)$ for large σ we note that in this limit the pixelation does not have such a noticeable effect, therefore, as long as the substrate length L is much larger than the dot and $\sigma \gg \ell$, we can write

$$\begin{aligned} \mathcal{I}(\theta) &= \frac{4}{\sigma^4} \frac{\sum_x \exp[-\frac{(x-\theta)^2}{\sigma^2}](x-\theta)^2}{\sum_{x'} \exp[-\frac{(x'-\theta)^2}{\sigma^2}]} \\ &\to \frac{4}{\sigma^4} \frac{\int_{-\infty}^{\infty} \mathrm{d}x \, \exp[-\frac{(x-\theta)^2}{\sigma^2}](x-\theta)^2}{\int_{-\infty}^{\infty} \mathrm{d}x' \, \exp[-\frac{(x'-\theta)^2}{\sigma^2}]} \\ &= \frac{2}{\sigma^2}, \end{aligned} \tag{4.33}$$

which gives the resolution $\delta\theta \geq \sigma/\sqrt{2}$. The simplicity of the system allows us to introduce more complex POVMs that include imperfect measurements consistent with the use of physically realisable detectors.

First we consider the effect of an upper limit to the intensity measured by the detector. We call this detector a saturable detector as it has a saturation point, past which it can no longer detect additional photons. Fig. 4.6 shows what the image looks like for such a Gaussian dot measured by such a detector. The conditional probabilities for a POVM describing such a detector are

$$q_k(x|n_x) = \begin{cases} \delta_{k,n_x} & \text{if } k < S \\ \sum_{j=0}^{\infty} \delta_{n_x,S+j} & \text{if } k = S \end{cases}, \tag{4.34}$$

where S is the saturation point of the detector. The sum in Eq. (4.34) ensures that all states with photon number greater than or equal to S are grouped together in the same measurement outcome S modelling the effect of the saturation. From Eq. (4.3) we find

$$\begin{aligned} \langle \hat{I}(x) \rangle &= \sum_{k=0}^{S} i_k \sum_{n_x=0}^{\infty} q_k(x|n_x) |\langle n_x|\psi\rangle|^2 \\ &= \sum_{n_x=0}^{\infty} \sum_{k=0}^{S-1} i_k \delta_{k,n_x} |\langle n_x|\psi\rangle|^2 + \sum_{n_x=0}^{\infty} \sum_{j=0}^{\infty} i_S \delta_{n_x,S+j} |\langle n_x|\psi\rangle|^2 \\ &= \sum_{n_x=0}^{S-1} i_k |\langle n_x|\psi\rangle|^2 + \sum_{n_x=S}^{\infty} i_S |\langle n_x|\psi\rangle|^2. \end{aligned} \tag{4.35}$$

Since $|n_x\rangle$ form a complete basis we can write $\sum_{n_x=0}^{\infty} |\langle n_x|\psi\rangle|^2 = \sum_{n_x=0}^{S-1} |\langle n_x|\psi\rangle|^2 +$

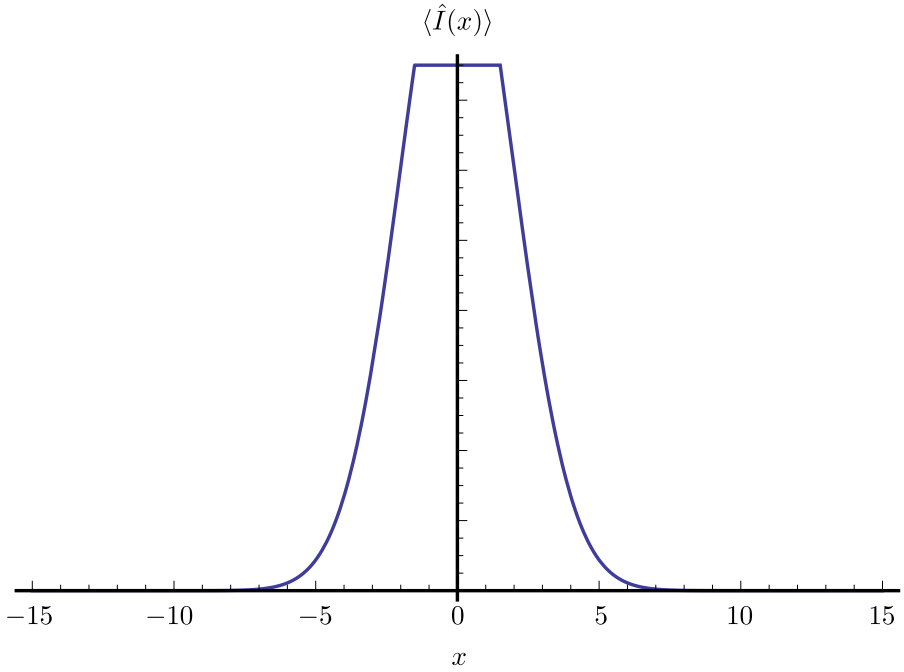

Figure 4.6: A Gaussian intensity distribution as measured by a detector with a saturation point. The effect of the saturation is to chop the top of the Gaussian distribution.

$\sum_{n_x=S}^{\infty} |\langle n_x|\psi\rangle|^2 = 1$ and the image is therefore

$$\begin{aligned}\langle\hat{I}(x)\rangle &= \sum_{n_x=0}^{S-1} i_k |\langle n_x|\psi\rangle|^2 + i_S \left(1 - \sum_{n_x=0}^{S-1} |\langle n_x|\psi\rangle|^2\right) \\ &= \sum_{n_x=0}^{S-1} (i_k - i_S)|\langle n_x|\psi\rangle|^2 + i_S. \end{aligned} \tag{4.36}$$

For the state given by Eq. (4.30) we find

$$|\langle n_x|\psi\rangle|^2 = \mathrm{e}^{-|\alpha_x|^2} \frac{|\alpha_x|^{2n_x}}{n_x!} \tag{4.37}$$

and the probability distribution $\Pr(x|\theta)$ is

$$\begin{aligned}\Pr(x|\theta) &= \frac{\sum_{n_x=0}^{S-1} i_0(n_x - S)\mathrm{e}^{-|\alpha_x|^2}\frac{|\alpha_x|^{2n_x}}{n_x!} + i_0 S}{\sum_x \left\{\sum_{n_x=0}^{S-1} i_0(n_x - S)\mathrm{e}^{-|\alpha_x|^2}\frac{|\alpha_x|^{2n_x}}{n_x!} + i_0 S\right\}} \\ &= \frac{\sum_{n_x=0}^{S-1} (n_x - S)\mathrm{e}^{-|\alpha_x|^2}\frac{|\alpha_x|^{2n_x}}{n_x!} + S}{\sum_x \left\{\sum_{n_x=0}^{S-1} (n_x - S)\mathrm{e}^{-|\alpha_x|^2}\frac{|\alpha_x|^{2n_x}}{n_x!} + S\right\}}, \end{aligned} \tag{4.38}$$

where we have defined $i_0 \equiv i_k/k$ to be the image intensity per photon. We again make the assumption that the image normalisation I_0 is independent of θ. The

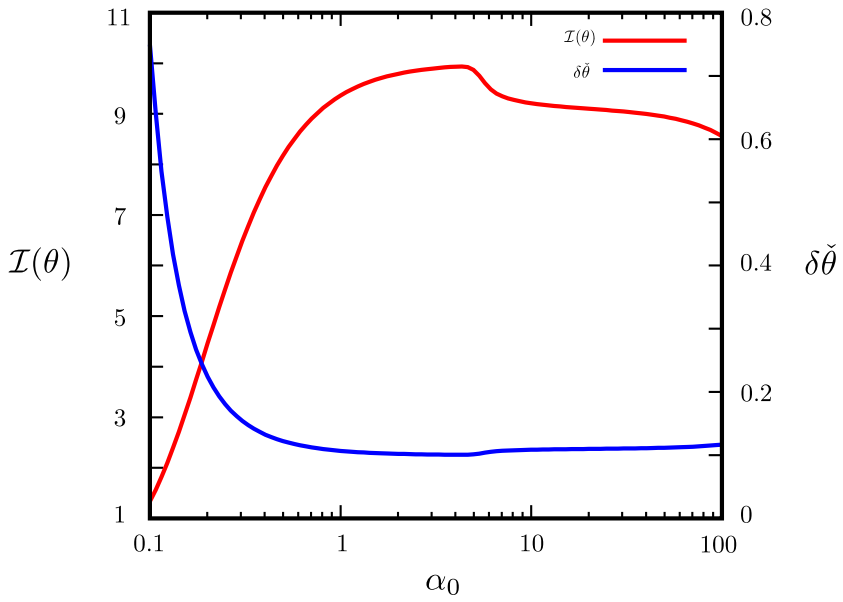

Figure 4.7: The Fisher information and variance of θ as a function of the peak amplitude α_0. In this example $S = 28$, $\varepsilon = 0.005$ and $\sigma = 100$.

derivative of the probability distribution $\Pr(x|\theta)$ is therefore given by

$$\frac{\partial \Pr(x|\theta)}{\partial \theta} = \frac{\sum_{n_x=0}^{S-1} (n_x - S) 2\mathrm{e}^{-|\alpha_x|^2} \frac{|\alpha_x|^{2n_x}}{n_x!} \frac{(x-\theta)}{\sigma^2} \left[n_x - |\alpha_x|^2\right]}{\sum_x \left\{ \sum_{n_x=0}^{S-1} (n_x - S) \mathrm{e}^{-|\alpha_x|^2} \frac{|\alpha_x|^{2n_x}}{n_x!} + S \right\}}, \tag{4.39}$$

and the Fisher information is

$$\mathcal{I}(\theta) = \sum_x \frac{\left(\sum_{n_x=0}^{S-1} (n_x - S) 2\mathrm{e}^{-|\alpha_x|^2} \frac{|\alpha_x|^{2n_x}}{n_x!} \frac{(x-\theta)}{\sigma^2} \left[n_x - |\alpha_x|^2\right] \right)^2}{\left\{ \sum_{n_x=0}^{S-1} (n_x - S) \mathrm{e}^{-|\alpha_x|^2} \frac{|\alpha_x|^{2n_x}}{n_x!} + S \right\} \sum_{x'} \left\{ \sum_{n_{x'}=0}^{S-1} (n_{x'} - S) \mathrm{e}^{-|\alpha_{x'}|^2} \frac{|\alpha_{x'}|^{2n_{x'}}}{n_{x'}!} + S \right\}}. \tag{4.40}$$

Fig. 4.7 shows the Fisher information and the standard deviation $\delta\theta$ as a function of the coherent state amplitude α_0. As expected, the Fisher information is reduced when the peak of the Gaussian reaches the saturation point, however, the effect does not completely destroy our ability to estimate θ. The initial increase in $\mathcal{I}(\theta)$ for low α_0 is an artefact of the small but non-zero noise ϵ, which was added as a background noise term to ensure the probability never gets so small that it causes numerical infinities when calculating the Fisher information. Reducing ϵ makes this slope steeper until the behaviour is essentially a step function before reaching the dip caused by the saturation point.

We now investigate the effect of pixel "bleeding" which we define to be the effect of a signal measured at a pixel being recorded by it's neighbours. The conditional probabilities for such a substrate can be described by $q_k(x|\vec{n}) = \sum_{x'} \delta_{k,n_{x'}} p(d)$, where $p(d)$ is the probability of the signal bleeding a distance $d = |x - x'|$, which we assume

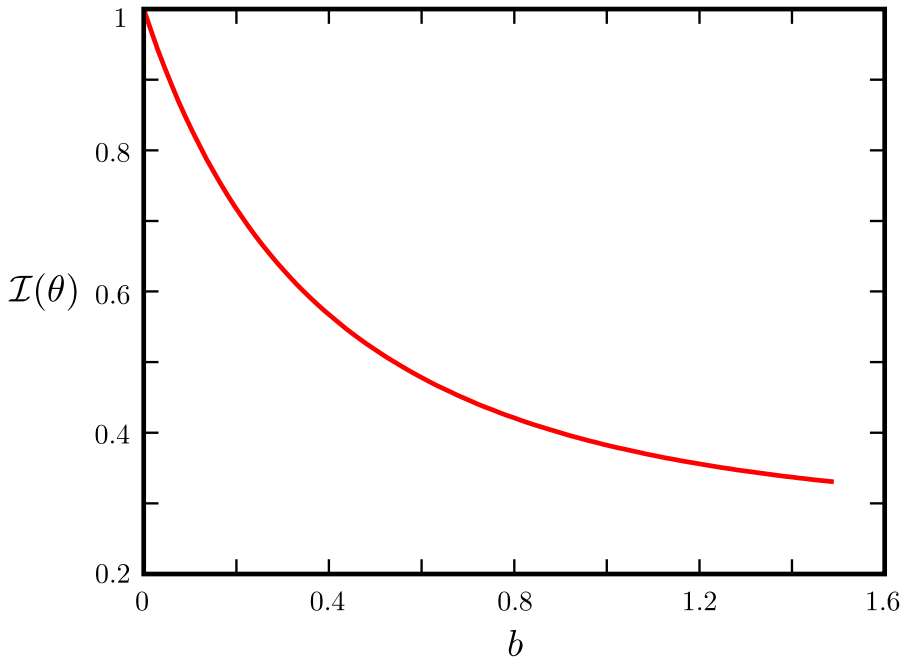

Figure 4.8: The Fisher information as a function of the pixel bleeding b.

to be symmetric. The image for such a detector is

$$\langle \hat{I}(x) \rangle = \sum_{x'} i_0 p(|x - x'|) |\alpha_{x'}|^2, \tag{4.41}$$

and, assuming again $\partial_\theta I_0 = 0$, the Fisher information is

$$\mathcal{I}(\theta) = \sum_{x_1} \frac{\left(\sum_{x_2} 2p(|x_1 - x_2|)(x_2 - \theta) \frac{|\alpha_{x_2}|^2}{\sigma^2} \right)^2}{\left\{ \sum_{x_3} p(|x_1 - x_3|) |\alpha_{x_3}|^2 \right\} \left\{ \sum_{x_4, x_5} p(|x_4 - x_5|) |\alpha_{x_5}|^2 \right\}}. \tag{4.42}$$

For simplicity we take as our probability distribution $p(|x - x'|)$ a Poissonian with bleeding parameter b which characterises the average distance that the signal bleeds. This gives us

$$p(|x - x'|) = \frac{1}{p_0} \frac{b^{|x-x'|} \mathrm{e}^{-b}}{|x - x'|}, \tag{4.43}$$

where

$$p_0 = \sum_{x''} \frac{b^{|x-x''|} \mathrm{e}^{-b}}{|x - x''|}, \tag{4.44}$$

which ensures that the probability is normalised despite the finite size of the substrate. Since the parameter α_0 appears to the fourth power in both the numerator and denominator of Eq. (4.42). The plot shows that the Fisher information is decreased as the bleeding parameter b is increased. Fig. 4.8 shows the variation in the Fisher information as a function of b.

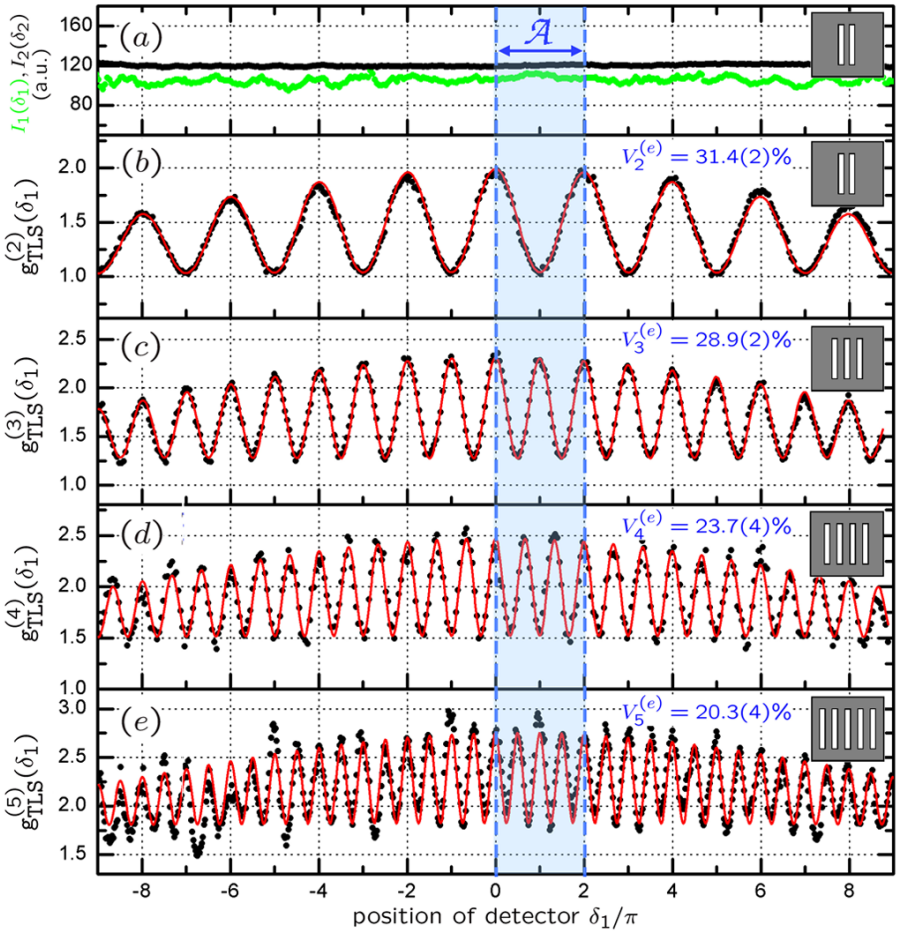

Figure 4.9: Measurements of the first five normalised intensity correlation functions showing the ability for the correlation functions to exhibit sinusoidal oscillations. The frequency of these is seen to increase with the number of sources and detectors. Figure reproduced with permission from the authors and the American Physical Society.

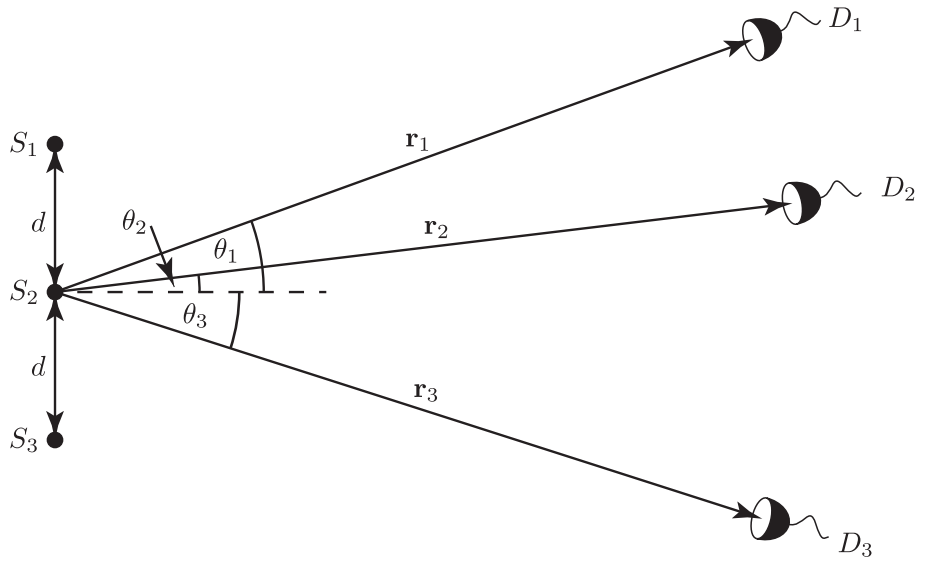

Figure 4.10: Demonstrating the setup used in the paper [Oppel et al., 2012]. Three uniformly distributed sources S_1, S_2 and S_3 emit radiation into the far field which is then detected by the three detectors D_1, D_2 and D_3. Generally n sources and detectors can be used.

4.6 Multi-photon correlations

In this section we analyse the recent experiments by Oppel *et al.* [Oppel et al., 2012], in particular we use the method of the previous section to calculate resolution limits in this context. In their paper *Superresolving Multiphoton Interferences with Independent Light Sources* the authors observe interference fringes in the far field intensity correlation functions using a variable number of statistically independent pseudo-thermal light sources. The experiments demonstrated that the use of multi-photon correlation measurements suggests an increase in resolution when used in imaging. The authors also calculate the far field intensity correlation functions for statistically independent single photon emitters and show that these also exhibit interference fringes. By placing the detectors at particular positions, the intensity correlation functions can achieve a sinusoidal modulation with spatial frequency inversely proportional to the number of independent sources (see Fig. 4.9). The authors demonstrate that the use of higher order correlation functions can provide more information about the separation of the sources as well as providing information about the number of sources present. Therefore by observing higher order intensity correlations we can obtain a more precise estimate of the spatial layout of the sources.

Fig. 4.10 shows the setup for three sources. The statistical independence of the sources allows us to write the density operator as $\rho = \otimes_{i=1}^{N} \rho_{S_i}$ where ρ_{S_i} is the state of the i^{th} source in the far field. Assuming the state to be diagonal in the Fock basis we can write

$$\rho_{S_i} = \sum_{k_i=0}^{\infty} P_{S_i}(k_i) |k_i\rangle \langle k_i| \tag{4.45}$$

where $P_{S_i}(k_i)$ is the probability of finding the field from the source i in the state $|k_i\rangle$.

By the superposition principle the electric field operators in the far field, $\hat{E}(\mathbf{r})$ will be a sum of the field operators from each of the sources $\hat{S}_1(\mathbf{r}), \ldots, \hat{S}_N(\mathbf{r})$. Writing the operator $\hat{\mathcal{G}} = \hat{E}^{(+)}(\mathbf{r}_1)\ldots\hat{E}^{(+)}(\mathbf{r}_n)$ for shorthand [1], we can write

$$\begin{aligned}\hat{\mathcal{G}} &= \prod_{i=1}^{n} \hat{E}^{(+)}(\mathbf{r}_i) \\ &= \prod_{i=1}^{n} [\hat{S}_1^{(+)}(\mathbf{r}_i) + \cdots + \hat{S}_N^{(+)}(\mathbf{r}_i)]\end{aligned} \tag{4.46}$$

where $\hat{S}_j^{(+)}(\mathbf{r}_i)$ is proportional to the photon annihilation operator at the point $\mathbf{r}_i$ from the source S_j. The n-point intensity correlation can be calculated as follows

$$\begin{aligned}&G^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n) \\ &\quad= \mathrm{Tr}\left[\rho\hat{\mathcal{G}}\right] \\ &\quad= \mathrm{Tr}\left[\sum_{k_1=0}^{\infty}\cdots\sum_{k_N=0}^{\infty} P_{S_1}(k_1)\ldots P_{S_N}(k_N)|k_1,\ldots,k_N\rangle\langle k_1,\ldots,k_N|\hat{\mathcal{G}}^\dagger\hat{\mathcal{G}}\right] \\ &\quad= \sum_{k_1=0}^{\infty}\cdots\sum_{k_N=0}^{\infty} P_{S_1}(k_1)\ldots P_{S_N}(k_N)\langle k_1,\ldots,k_N|\hat{\mathcal{G}}^\dagger\hat{\mathcal{G}}|k_1,\ldots,k_N\rangle \\ &\quad= \sum_{k_1=0}^{\infty}\cdots\sum_{k_N=0}^{\infty} P_{S_1}(k_1)\ldots P_{S_N}(k_N)\sum_{m_1=0}^{\infty}\cdots\sum_{m_N=0}^{\infty}|\langle m_1,\ldots,m_N|\hat{\mathcal{G}}|k_1,\ldots,k_N\rangle|^2.\end{aligned} \tag{4.47}$$

Writing $\hat{S}_j^{(+)}(\mathbf{r}_i) = \kappa e^{ikr_{ji}}\hat{a}_{S_j}$, where r_{ji} is the distance from the j^{th} source to i^{th} detector, we can calculate the normalised intensity correlation explicitly for $N = 2$ and 3. The factor κ is strictly a function of the geometry of the arrangement. However, in the far field, paraxial regime, κ can be considered a constant. First we must calculate $G^{(1)}(\mathbf{r}_1)$ for $N = 2$ and 3

$$\begin{aligned}G^{(1)}(\mathbf{r}_1) &= \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty} P_{S_1}(k_1)P_{S_2}(k_2)|\langle m_1,m_2|\hat{S}_1^{(+)}(\mathbf{r}_1)+\hat{S}_2^{(+)}(\mathbf{r}_1)|k_1,k_2\rangle|^2 \\ &= \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} P_{S_1}(k_1)P_{S_2}(k_2)\Big[|\langle k_1-1,k_2|\hat{S}_1^{(+)}(\mathbf{r}_1)|k_1,k_2\rangle|^2 \\ &\quad + |\langle k_1,k_2-1|\hat{S}_2^{(+)}(\mathbf{r}_1)|k_1,k_2\rangle|^2\Big] \\ &= \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} P_{S_1}(k_1)P_{S_2}(k_2)\kappa[k_1+k_2] \\ &= \kappa[\langle\hat{n}_1\rangle + \langle\hat{n}_2\rangle]\end{aligned} \tag{4.48}$$

where $\langle\hat{n}_i\rangle$ is the photon number expectation associated with source i. Likewise for $N = 3$, $G^{(1)}(\mathbf{r}_1) = \kappa\langle\hat{n}_1\rangle + \langle\hat{n}_2\rangle + \langle\hat{n}_3\rangle$. The second order intensity correlation is

[1] We use N for the total number of sources and n for the number of detectors. These do not necessarily have to be the same although in the experiments performed by Oppel *et al.* $N = n$ was used throughout.

calculated to be

$$G^{(2)}(\mathbf{r}_1,\mathbf{r}_2) = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty} P_{S_1}(k_1)P_{S_2}(k_2)|\langle m_1,m_2|[\hat{S}_1^{(+)}(\mathbf{r}_1)+\hat{S}_2^{(+)}(\mathbf{r}_1)] \times [\hat{S}_1^{(+)}(\mathbf{r}_2)+\hat{S}_2^{(+)}(\mathbf{r}_2)]|k_1,k_2\rangle|^2. \tag{4.49}$$

In general all the terms in the expansion of the square brackets will contribute to the correlation function.

Using for the sources single photon emitters (SPE) and thermal light sources (TLS), we find the probabilities $P_{S_i}(k_j)$ are given by

$$\text{SPE}: \qquad P_{S_i}(k_j) = \begin{cases} 1 & \text{if } k_j = 1 \\ 0 & \text{if } k_j \neq 1 \end{cases} \tag{4.50}$$

$$\text{TLS}: \qquad P_{S_i}(k_j) = \frac{\langle\hat{n}_i\rangle^{k_j}}{(1+\langle\hat{n}_i\rangle)^{k_j+1}}, \tag{4.51}$$

where the distribution for the TLS sources is the well known Bose-Einstein distribution. The sources are treated as point like, therefore justifying the assumption that the thermal light sources consist of a single mode in thermal equilibrium.

4.6.1 Single photon sources

For single photon sources, we find by inserting Eq. (4.50) into Eq. (4.49) we find

$$\begin{aligned} G^{(2)}_{\text{SPE}}(\mathbf{r}_1,\mathbf{r}_2) &= |\langle k_1-1,k_2-1|[\hat{S}_1^{(+)}(\mathbf{r}_2)\hat{S}_2^{(+)}(\mathbf{r}_1)+\hat{S}_1^{(+)}(\mathbf{r}_1)\hat{S}_2^{(+)}(\mathbf{r}_2)]|k_1,k_2\rangle|^2 \\ &= \kappa^2|\mathrm{e}^{ik(r_{21}+r_{12})}+\mathrm{e}^{ik(r_{11}+r_{22})}|^2 \\ &= \kappa^2\left[2+\mathrm{e}^{ik(r_{21}+r_{12}-r_{11}-r_{22})}+\mathrm{e}^{-ik(r_{21}+r_{12}-r_{11}-r_{22})}\right] \\ &= \kappa^2\left[2+2\cos(kd\sin\phi_1-kd\sin\phi_2)\right] \\ &\equiv \kappa^2\left[2+2\cos(\delta_1-\delta_2)\right], \end{aligned} \tag{4.52}$$

where we have defined $\delta_i = k\theta\sin\phi_i$, ϕ_i being the angle between the central normal to the sources and the source i and θ being the separation between the two sources. Using Eq. (4.48) we find

$$g^{(2)}_{\text{SPE}}(\delta_1,\delta_2) = \frac{1}{2}[1+\cos(\delta_1-\delta_2)], \tag{4.53}$$

where we see the appearance of a sinusoidal oscillation of the correlation function with spatial frequency kd and we have changed the arguments to δ_1, δ_2 for clarity.

Since the correlation functions depend on all the possible indistinguishable paths that the emitted photons can take to arrive at the detector the general expression for the correlation functions grows factorially with n. For single photon sources the third order correlation is

$$\begin{aligned} g^{(3)}_{\text{SPE}}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) = \frac{2}{27}\Big[&3+2\cos(\delta_1-\delta_2)+2\cos(\delta_1-\delta_3)+2\cos(\delta_2-\delta_3) \\ &+\cos(2\delta_1-2\delta_2)+\cos(2\delta_1-2\delta_3)+\cos(2\delta_2-2\delta_3) \\ &+2\cos(2\delta_1-\delta_2-\delta_3)+\cos(\delta_1+\delta_2-2\delta_3)+\cos(-\delta_1+2\delta_2-\delta_3) \\ &+\cos(\delta_1+\delta_2-2\delta_3)+\cos(\delta_1-2\delta_2-\delta_3)\Big], \end{aligned} \tag{4.54}$$

where we have assumed the sources to be uniformly distributed about a straight line as in Fig 4.10. For $N = 2$ we clearly see that the choice $\delta_2 = 0$ leads to a pure sinusoidal spatial modulation of the correlation function. For $N = 3$ it is less obvious but can be verified by direct substitution that the choice $\delta_2 = \pi/4$, $\delta_3 = 7\pi/4$ also leads to a pure sinusoidal modulation, we find

$$g^{(3)}_{\mathrm{SPE}}\left(\delta_1, \frac{\pi}{4}, \frac{7\pi}{4}\right) = \frac{4}{27}(1 + \cos 2\delta_1) \tag{4.55}$$

where we see that the spatial frequency of the modulation is twice that of $g^{(2)}(\delta_1, 0)$ [Oppel, 2012]. Since k and θ are fixed for a particular experiment, our choice of δ_2 and δ_3 must be achieved by moving the positions of the second and third detectors such that the angles ϕ_2 and ϕ_3 correspond to $\phi_i = \sin^{-1}(\delta_i/k\theta)$.

The general expression for the intensity correlation functions for single photon sources is give by

$$g^{(n)}_{\mathrm{SPE}} = A\{1 + \cos[(n-1)\delta_1]\},. \tag{4.56}$$

Writing Eq. (4.56) as $g^{(n)}_{\mathrm{SPE}} = 2A\cos^2[(n-1)\delta_1]$ we notice the similarity with Eq. (4.11). In fact, taking the integral of Eq. (4.56) over the range $-\mathcal{A} \leq \phi_1 \leq \mathcal{A}$ allows us to define the probability distribution

$$\Pr(\phi_1|\theta) = \frac{g^{(n)}(\delta_1, m_2, \ldots, m_n)}{\int_{-\mathcal{A}}^{\mathcal{A}} g^{(n)}(\delta_1, m_2, \ldots, m_n)\,\mathrm{d}\phi_1} = \frac{k\theta\cos^2\left[\frac{(n-1)}{2}kd\phi_1\right]}{\left\{k\theta\mathcal{A} + \frac{\sin[(n-1)kd\mathcal{A}]}{(n-1)}\right\}}, \tag{4.57}$$

and in an equivalent manner determine the minimum resolvable slit separation

$$\theta^{(n)}_{\mathrm{min}} \geq \frac{0.369\lambda}{(n-1)\mathcal{A}}. \tag{4.58}$$

We see directly that $\theta^{(n)}_{\mathrm{min}} \geq \theta^{(n+1)}_{\mathrm{min}}$ which demonstrates that the higher order intensity correlations allow for the resolution of smaller source separations than the lower orders.

4.6.2 Thermal light sources

For thermal light sources there is also the possibility of both photons originating from the same source, this gives rise to additional terms in the calculation of $G^{(2)}_{\mathrm{TLS}}(\mathbf{r}_1, \mathbf{r}_2)$. As we will show these terms have the effect of reducing the visibility whilst keeping the spatial frequency the same as for single photon sources.

$$\begin{aligned}
&G^{(2)}_{\mathrm{TLS}}(\mathbf{r}_1, \mathbf{r}_2) = \\
&\sum_{k_1,k_2} P_{S_1}(k_1)P_{S_2}(k_2)\left\{|\langle k_1 - 1, k_2 - 1|[\hat{S}^{(+)}_1(\mathbf{r}_2)\hat{S}^{(+)}_2(\mathbf{r}_1) + \hat{S}^{(+)}_1(\mathbf{r}_1)\hat{S}^{(+)}_2(\mathbf{r}_2)]|k_1, k_2\rangle|^2\right. \\
&\quad \left. + |\langle k_1 - 2, k_2|\hat{S}^{(+)}_1(\mathbf{r}_1)\hat{S}^{(+)}_1(\mathbf{r}_2)|k_1, k_2\rangle|^2 + |\langle k_1 - 2, k_2|\hat{S}^{(+)}_2(\mathbf{r}_1)\hat{S}^{(+)}_2(\mathbf{r}_2)|k_1, k_2\rangle|^2\right\} \\
&= \kappa^2\langle\hat{n}_1\rangle\langle\hat{n}_2\rangle[2 + 2\cos(\delta_1 - \delta_2)] + 2\kappa^2\langle\hat{n}_1\rangle^2 + 2\kappa^2\langle\hat{n}_2\rangle^2,
\end{aligned} \tag{4.59}$$

where we see that the interference comes from the two photons coming from individual sources, whereas when the photons originate from the same source, no

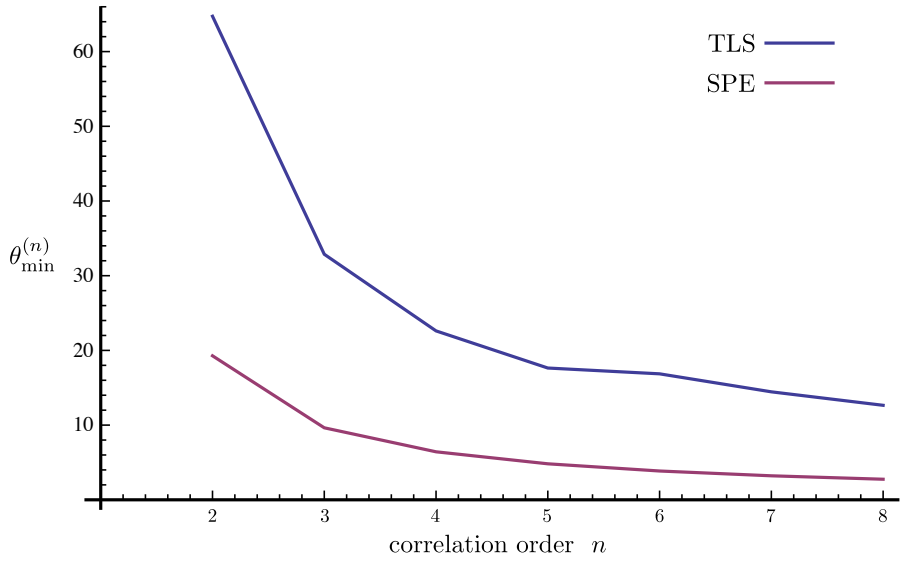

Figure 4.11: The resolution of the correlation functions against the correlation order n.

interference occurs. Assuming the sources to be equal intensities, $\langle \hat{n}_1 \rangle = \langle \hat{n}_2 \rangle \equiv \langle \hat{n} \rangle$ we find the normalised intensity correlation to be

$$g^{(2)}_{\mathrm{TLS}}(\delta_1, \delta_2) = \frac{3}{2}\left[1 + \frac{1}{3}\cos(\delta_1 - \delta_2)\right]. \tag{4.60}$$

We see from Eq. (4.60) that the spatial modulation of the correlation function is the same as for the single photon sources, however the visibility is reduced by 70%. There is a clear interpretation of this reduction in the visibility, since it is caused by the terms that account for the two photons originating from the same source which clearly carry no information about the spatial separation of the two independent sources.

Remarkably, as we observed for $g^{(2)}$, the third order correlation for thermal light sources can also exhibit a pure sinusoidal oscillation of the same spatial frequency as $g^{(3)}_{\mathrm{TLS}}$. As with the second order correlation the visibility for thermal light sources is reduced compared with the single photon emitters. In order to obtain a pure sinusoidal modulation for thermal light sources, we must choose $\delta_2 = 0$ and $\delta_3 = \pi$. For an appropriate choice $\delta_2 \ldots \delta_n$ we find that the intensity correlations for thermal light sources can be described by the expression

$$g^{(n)}_{\mathrm{TLS}} = B\{1 + V_n \cos[(n-1)\delta_1]\}. \tag{4.61}$$

For thermal light sources the appearance of V_n in Eq. (4.61) does not allow us to derive a similar analytic expression for the minimum resolvable source separation. However we can numerically find the resolution for various orders of n. Fig. 4.11 shows the resolution as a function of n, we see that the single photon emitting sources always provide better resolution compared to the thermal light sources, however, the general trend of the resolution decreasing with n is still present even for thermal sources.

4.7 Discussion and conclusions

In conclusion, we have demonstrated how the statistical distance can be used as a measure of distinguishability for probability distributions. By identifying images with probability distributions we have shown how the statistical distance can be used to determine the resolution of various imaging procedures. Detector imperfections can be fully accounted for and the resolution of various setups including imperfections can be calculated via this method. By identifying an image with the expectation value of the imaging observable we ignore the effects of noise in the detection. In order to properly perform a parameter estimation of the spatial parameters of the source we must take noise into account. This is the topic of the next chapter.

Chapter 5

Estimation of Thermal Source Dimensions

The contents of this chapter is based on [Pearce et al., 2015]. In this chapter we use rigorous estimation theory to demonstrate how higher-order intensity interferometry can yield a resolution improvement over the lowest (second) order intensity interferometry. The general setup is shown in Fig. 5.1a. We provide a general framework that allows us to analyse intensity correlations of arbitrary order and discuss the advantages and disadvantages associated with higher-order intensity correlations. We present a rigorous analysis of the HBT set up, explicitly calculating the likelihood function and Fisher information, which allows us to determine the lowest possible variance in estimates of the source and also allows us to perform a maximum likelihood estimation, therefore achieving this bound. An integral part of our procedure is to treat unknown quantities as parameters to be estimated, every unknown therefore increasing the size of the estimation problem. We then demonstrate our method, making use of simulated data to obtain precision estimates of the radius of a thermal source. We also show how these results can be used to estimate the dimensions of any source, regardless of its geometry. Our method is also not limited to thermal sources. We show that as long as the total photon number is not deterministic (i.e. $\Delta\hat{n} > 0$) our method can be used to calculate the Fisher information and also provide maximum likelihood estimates of the source dimensions via the method of scoring.

By explicitly calculating the Fisher information for different correlation orders n, we show that in some instances the Fisher information increases with correlation order. However the exact relationship between the Fisher information and the correlation order depends strongly on the particular arrangement of the detectors and the source geometry. The idea of exploiting higher order intensity correlations to improve resolution has been suggested previously [Thiel et al., 2007, Oppel et al., 2012, Iskhakov et al., 2011] and in the context of ghost imaging, the idea of exploiting higher correlation orders has received great attention [Zhou et al., 2010, Chen et al., 2010, Chan et al., 2009, Agafonov et al., 2009b]. However, to date there has not been a rigorous theoretical explanation which can quantify the possible enhancement of higher order correlation measurements for HBT-type experiments. We provide that explanation here and demonstrate its use. We also include the effects of photon losses in our model and show that higher order correlation can continue to perform well even in the presence of loss.

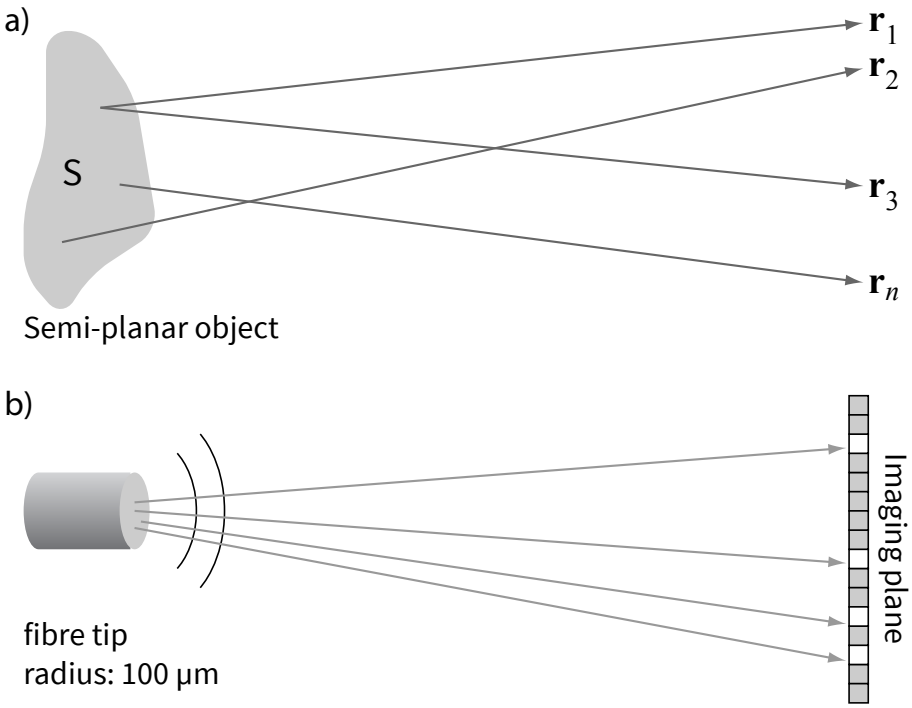

Figure 5.1: (a) Spatially incoherent, planar source S emitting into the far field. The intensity measured at the positions $\mathbf{r}_1, \ldots, \mathbf{r}_n$ can be correlated to give the n-point intensity correlation function, which in turn contains information about the geometry of S. (b) Implementation simulated here: a fibre tip of radius 100 μm emits monochromatic thermal light with a wavelength of 633 nm, which is recorded with a CCD camera. Our task is to estimate the radius.

We consider the particular measurement scheme of an array of pixels measuring the intensity at a discrete number of positions as in Fig. 5.2a. The benefit of such a setup is that we can capture a large amount of correlation data in one exposure. For example, as shown in Fig 5.2 we can calculate the two-point intensity correlation between any pair of pixels as a function of the position of one of the pixels, $G^{(2)}(x, s_2) = G^{(1)}(x)G^{(1)}(s_2)$, where s_2 is the stationary position of the second pixel. When measuring a correlation function in this way the individual data points that are calculated may not be independent. The correlation between any pair of measurements, e.g., $G^{(2)}(x_1, s_2)$ and $G^{(2)}(x_2, s_2)$, depends on the correlations between all the measured intensities $G^{(1)}(x_1)$, $G^{(1)}(x_2)$ and $G^{(1)}(s_2)$. Since the measurement relies on the statistical dependence of these intensities, the resulting values of $G^{(2)}(x_1, s_2)$ and $G^{(2)}(x_2, s_2)$ will not, in general, be independent. The same argument holds for higher order correlations.

We also assume that repeated measurements made by the pixel array are independent, i.e. separate images are independent. For thermal light this shouldn't present a problem because the coherence times of thermal light are intrinsically very short and will typically be far less than the dead time of the detectors. However, for pseudo-thermal light this may no longer be the case. We therefore assume that sufficient time is left between images such that the images can be considered to be statistically independent. This procedure represents a sort of course-graining over the temporal domain, which will generally reduce the amount of information [Ferrie, 2014]. However, since this sort of course-graining is inherent in most devices that

measure intensities, we deem this assumption to be necessary and relevant. The assumption of independent images allows us to use the central limit theorem, which vastly simplifies our model of the probability distribution function (PDF) of our data.

As we will show explicitly in the next section, the visibility for n^{th}-order intensity correlations can increase with larger n. The visibility of a signal s is defined as

$$\mathcal{V} = \frac{s_{\max} - s_{\min}}{s_{\max} + s_{\min}}, \tag{5.1}$$

which we see is bounded by $0 \leq \mathcal{V} \leq 1$. Although the visibility is sometimes used as a figure of merit in imaging, in order to perform a proper parameter estimation, the full PDF must be known. In particular, the noise of the signal (characterised by the second moments of the PDF) must also be taken into account. For signals generated via postselection of the data, such as intensity correlation signals, the signal to noise ratio will generally decrease as less data is postselected (increasing n), which will generally lead to worse parameter estimation. However, whatever figure of merit is chosen, there is no substitute for calculating the Fisher information from the PDF and therefore finding the amount of information that can be extracted from the data.

5.1 n-point intensity correlation functions

Consider a semi-planar source S emitting radiation that is observed in the far field. The setup is described diagrammatically in Fig. 5.1a. We first describe the second order intensity interference for experiments that measure the equal-time two-point intensity correlation

$$\begin{aligned} G^{(2)}(\mathbf{r}_1, \mathbf{r}_2) &= \langle :\hat{E}^{(-)}(\mathbf{r}_1)\hat{E}^{(+)}(\mathbf{r}_1)\hat{E}^{(-)}(\mathbf{r}_2)\hat{E}^{(+)}(\mathbf{r}_2):\rangle \\ &\propto \langle :\hat{a}^\dagger(\mathbf{r}_1)\hat{a}(\mathbf{r}_1)\hat{a}^\dagger(\mathbf{r}_2)\hat{a}(\mathbf{r}_2):\rangle \,, \end{aligned} \tag{5.2}$$

where $E^{(\pm)}$ are the positive and negative frequency parts of the electric field, $\hat{a}(\mathbf{r})$ and $\hat{a}^\dagger(\mathbf{r})$ are the annihilation and creation operators of the field at position $\mathbf{r}$, $::$ denotes normal ordering, and $\mathbf{r}_i$ is the position of the i^{th} detector in the far field. When viewed in the far field of the source, interference fringes can be observed in the correlation functions under certain conditions [Oppel et al., 2012]. More generally, we can consider the equal-time n-point intensity correlation $G^{(n)} \propto \langle :\prod_{i=1}^{n} \hat{a}^\dagger(\mathbf{r}_i)\hat{a}(\mathbf{r}_i):\rangle$. Increasing the correlation order n can also lead to an increased visibility of the interference fringes, suggesting we may be able to extract more information from the higher correlation orders [Oppel et al., 2012, Agafonov et al., 2008b].

To calculate the n-point intensity correlation we make use of the optical equivalence theorem [Sudarshan, 1963], which states that the expectation of a normally ordered product of creation and annihilation operators can be replaced by respectively their left and right eigenvalues if the expectation is replaced by an ensemble average weighted by the P-representation of the state. Mathematically, we write

$$\langle f(\hat{a}^\dagger, \hat{a})\rangle = \int P_\rho(\alpha) f(\alpha^*, \alpha)\, \mathrm{d}^2\alpha \equiv \langle f(\alpha^*, \alpha)\rangle_P, \tag{5.3}$$

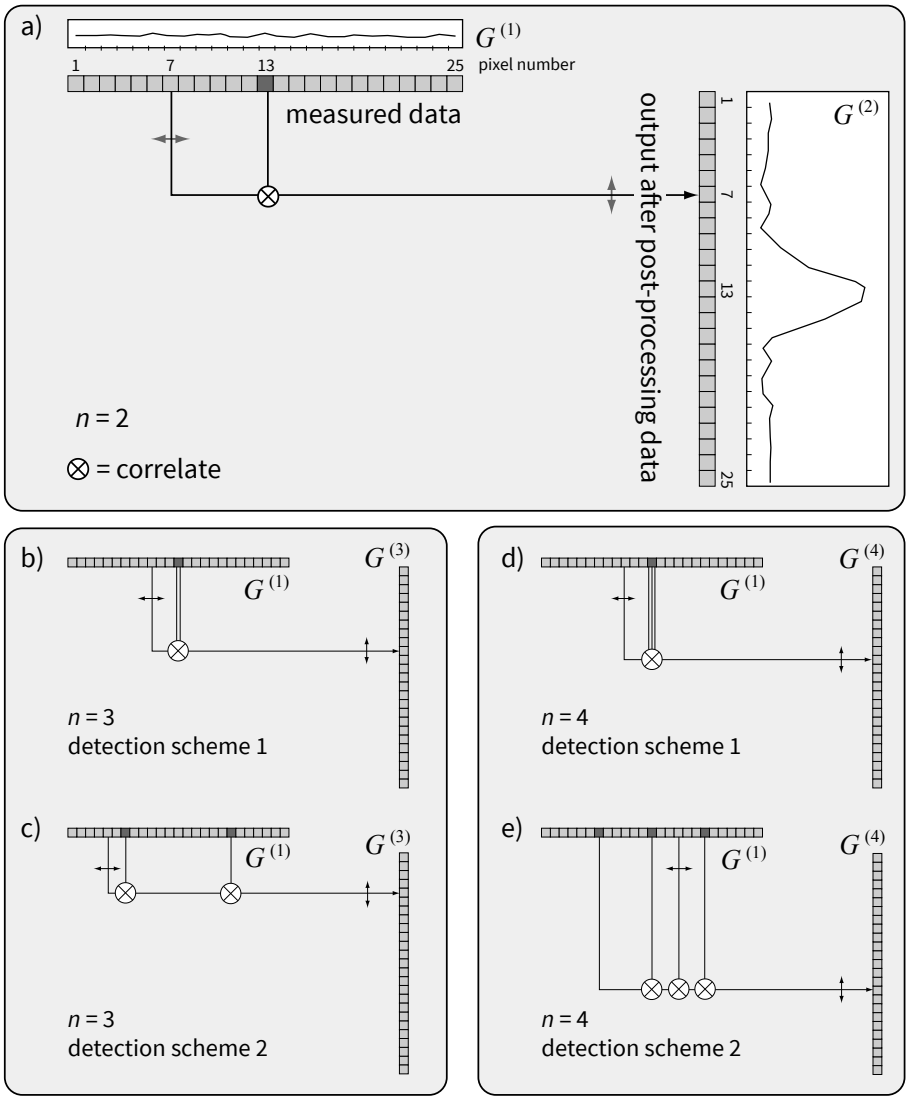

Figure 5.2: Measuring the second order intensity correlation function with an array of pixels. (a) The second order intensity correlation $G^{(2)}(x_i)$ at pixel $x_i = 7$ is calculated as the correlation between the intensities $G^{(1)}(s_2)$ at a fixed pixel $s_2 = 13$ (shown as a darker pixel) and $G^{(1)}(x_i)$ at the pixel $x_i = 7$. (b, c) The third order correlation function $G^{(3)}(x_i)$ is defined using two detection schemes, namely via a single fixed pixel that is correlated twice (detection scheme 1, note the double line from pixel 13), or two fixed pixels (detection scheme 2). (d, e) The fourth order intensity correlation $G^{(4)}(x_i)$ is defined analogous to $G^{(3)}(x_i)$.

where f is any normally ordered function of the creation and annihilation operators, the first expectation is the quantum mechanical average, and the subscript P on the second expectation signifies that it is an ensemble average taken with respect to the quasi-probability function P. With the help of the optical equivalence theorem, the n-point intensity correlation $G^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ can be written as

$$\left\langle :\prod_{i=1}^{n} \hat{a}^\dagger(\mathbf{r}_i)\hat{a}(\mathbf{r}_i): \right\rangle = \left\langle \prod_{i=1}^{n} \alpha^*(\mathbf{r}_i)\alpha(\mathbf{r}_i) \right\rangle_P. \tag{5.4}$$

Thermal light exhibits a Gaussian zero mean P-representation. We can therefore apply the Gaussian moment theorem [Mandel and Wolf, 1995] and make the simplification

$$\left\langle \prod_{i=1}^{n} \alpha^*(\mathbf{r}_i)\alpha(\mathbf{r}_i) \right\rangle_P = \sum_{\sigma \in S_n} \prod_{i=1}^{n} \langle \alpha^*(\mathbf{r}_i)\alpha(\mathbf{r}_{\sigma(i)}) \rangle, \tag{5.5}$$

were S_n is the symmetric group containing all permutations of the set $\{1, \ldots, n\}$. This allows us to write $G^{(n)}$ as

$$G^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n) = |K|^{2n} \sum_{\sigma \in S_n} \prod_{i=1}^{n} \langle \hat{a}^\dagger(\mathbf{r}_i)\hat{a}(\mathbf{r}_{\sigma(i)}) \rangle, \tag{5.6}$$

where we have defined $\hat{E}^{(+)} = K\hat{a}$. We can make a further simplification by introducing the complex degree of coherence, defined as [Mandel and Wolf, 1995] (sometimes referred to as the mutual coherence function [Gerry and Knight, 2006])

$$\gamma(\mathbf{r}_1, \mathbf{r}_2) = \frac{\langle \hat{a}^\dagger(\mathbf{r}_1)\hat{a}(\mathbf{r}_2) \rangle}{\left[\langle \hat{a}^\dagger(\mathbf{r}_1)\hat{a}(\mathbf{r}_1) \rangle \langle \hat{a}^\dagger(\mathbf{r}_2)\hat{a}(\mathbf{r}_2) \rangle \right]^{1/2}}. \tag{5.7}$$

This allows us to write the n-point intensity correlation function as

$$G^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} \langle \hat{a}^\dagger(\mathbf{r}_i)\hat{a}(\mathbf{r}_i) \rangle \gamma(\mathbf{r}_i, \mathbf{r}_{\sigma(i)}) \,, \tag{5.8}$$

where we omitted the constant of proportionality K for brevity. From this expression we see that the n-point correlation for Gaussian light is equal to the permanent of a matrix Γ [Minc, 1978]

$$G^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \mathrm{Perm}(\Gamma) \,, \tag{5.9}$$

where

$$\Gamma_{ij} \equiv \left[\langle \hat{a}^\dagger(\mathbf{r}_i)\hat{a}(\mathbf{r}_i) \rangle \langle \hat{a}^\dagger(\mathbf{r}_j)\hat{a}(\mathbf{r}_j) \rangle \right]^{1/2} \gamma(\mathbf{r}_i, \mathbf{r}_j). \tag{5.10}$$

In general the permanent of a matrix is difficult to calculate. Therefore, for larger n it is increasingly costly to calculate the n-point correlations. This is the main limiting factor for the use of higher-order correlation functions in imaging.

The advantage of writing $G^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ in terms of the complex degree of coherence is that the complex degree of coherence in the far field paraxial regime is given by the two-dimensional Fourier transform of the intensity distribution of the source [Mandel and Wolf, 1995]. This result is known as the Van Cittert-Zernike

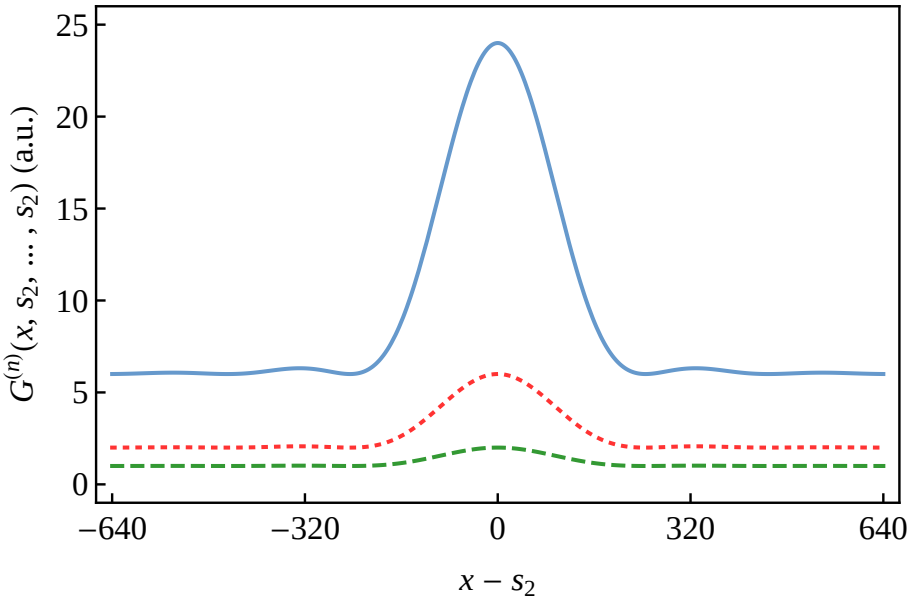

Figure 5.3: The second (dashed), third (dotted) and fourth (solid) order intensity correlation functions of Eq. (5.12) as a function of the separation between the scanning pixel $\mathbf{r}_1 = x$ and the reference pixel s_2 for the remaining arguments, and with normalised intensity $\langle \hat{a}^\dagger(\mathbf{r}_i)\hat{a}(\mathbf{r}_i)\rangle = 1$ and $\vartheta k = 64$. The width of the curves is directly proportional to the angular diameter of the source a. The higher visibilities of $G^{(3)}$ and $G^{(4)}$ over $G^{(2)}$ suggest that higher-order correlations may outperform regular Hanbury Brown and Twiss estimation of the source diameter. Here we show rigorously using estimation theory that this can indeed be the case in some instances.

theorem [van Cittert, 1934, Zernike, 1938]. We can therefore calculate the far field intensity correlations for any source geometry, provided that we can determine the Fourier transform of the intensity distribution.

As an example, consider a monochromatic, thermal, circular source of uniform intensity and angular diameter $\vartheta = \tan^{-1}(a/d) \approx a/d$ where a is the diameter of the source and d is the distance from the source to the observation plane. The far field complex degree of coherence is the two-dimensional Fourier transform of a circle with radius a [Born and Wolf, 1980]

$$\gamma(\mathbf{r}_1, \mathbf{r}_2) = \frac{2J_1\left(\vartheta k|\mathbf{r}_1 - \mathbf{r}_2|\right)}{\left(\vartheta k|\mathbf{r}_1 - \mathbf{r}_2|\right)}, \tag{5.11}$$

where J_1 is the first order Bessel function of the first kind and k is the wavenumber. The n-point correlation function then becomes

$$G^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} \langle \hat{a}^\dagger(\mathbf{r}_i)\hat{a}(\mathbf{r}_i)\rangle \frac{2J_1\left(\vartheta k|\mathbf{r}_i - \mathbf{r}_{\sigma(i)}|\right)}{\left(\vartheta k|\mathbf{r}_i - \mathbf{r}_{\sigma(i)}|\right)}. \tag{5.12}$$

Fig. 5.3 shows the second, third and fourth order correlation functions for a uniform disc along a one-dimensional detector. We have seen previously that the n-point correlation contains a large number of terms (the permanent contains a sum over the symmetric group S_n giving $n!$ terms). In practice this is what limits our ability to calculate arbitrary correlation orders.

5.2 Estimation from intensity correlations

In the previous section we saw that intensity correlation measurements in the far field depend on the parameters describing the source geometry, and in this section we examine how we can use these measurements to practically obtain spatial information about the source. To this end we employ parameter estimation theory, which involves the use of an estimator $\check{\boldsymbol{\theta}}$ that takes the measured data $\mathbf{x} = (x_1, \ldots, x_M)$ in M measurements (the pixels in the detector) and returns estimates of the parameters of interest $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_l)$. In order to extract the spatial information in the most efficient way, we can apply a maximum likelihood estimation procedure. Maximum likelihood estimation relies on maximisation of the joint probability distribution of our data, and we therefore need to characterise the probability distribution from which the correlation functions are sampled [Kay, 1993].

Once we have determined the the conditional probability $p(\mathbf{x}|\boldsymbol{\theta})$ of obtaining the measurement outcomes $\mathbf{x}$ given the values of the parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_l)$, we can determine the performance of our estimates using the Cramér-Rao bound (CRB). The CRB provides a lower bound for the variance of our estimators in terms of the Fisher information matrix

$$\mathrm{Var}(\theta_i) \geq [\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]^{-1}_{ii}, \tag{5.13}$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ is the Fisher information matrix given by

$$[\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]_{ij} = \sum_{\mathbf{x}} p(\mathbf{x}|\boldsymbol{\theta}) \left(\frac{\partial \ln[p(\mathbf{x}|\boldsymbol{\theta})]}{\partial \theta_i} \right) \left(\frac{\partial \ln[p(\mathbf{x}|\boldsymbol{\theta})]}{\partial \theta_j} \right), \tag{5.14}$$

where the sum over $\mathbf{x}$ may be an integral if the values of x_i form a continuum. In practice, intensity measurements in modern optical detectors yield a digital signal with discrete values.

The exact probability distribution $p(\mathbf{x}|\theta)$ of the correlation functions will be a complicated expression depending on the number of images N—not to be confused with the number of pixels M in each image. However, since the measurements of the correlation functions are averages over a (preferably) large data set, the central limit theorem dictates that these measurements will be normally distributed [van der Vaart, 1998]. Assuming that we make M measurements of the correlation functions at M discrete detector positions, in the limit of large N the data will follow an M-dimensional normal distribution:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^{\mathrm{T}} \mathbf{C}^{-1}(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))\right)}{\sqrt{(2\pi)^M |\mathbf{C}(\boldsymbol{\theta})|}}, \tag{5.15}$$

where $\boldsymbol{\mu}$ is the tuple of expectation values of the distribution at each of the sampling points: $\boldsymbol{\mu}^{\mathrm{T}} = (\langle \mathbf{x}_1 \rangle, \ldots, \langle \mathbf{x}_M \rangle)$, $^{\mathrm{T}}$ denotes the transpose, and $\mathbf{C}$ is the covariance matrix between pairs of measurements $\mathbf{C}_{ij} = \mathrm{Cov}(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i \mathbf{x}_j \rangle - \langle \mathbf{x}_i \rangle \langle \mathbf{x}_j \rangle$.

For a multivariate-normal distribution the elements of the Fisher information matrix are given by [Kay, 1993]

$$\begin{aligned} [\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]_{ij} &= \left(\frac{\partial \boldsymbol{\mu}}{\partial \theta_i} \right)^{\mathrm{T}} \mathbf{C}^{-1} \left(\frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \right) + \frac{1}{2} \mathrm{Tr}\left[\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_i} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_j} \right] \\ &\equiv [\boldsymbol{\mathcal{I}}_1(\boldsymbol{\theta})]_{ij} + [\boldsymbol{\mathcal{I}}_2(\boldsymbol{\theta})]_{ij}, \end{aligned} \tag{5.16}$$

	Statistics	Correlation functions
random variable	X_i	$I(x_i)\, I(s_2) \dots I(s_n)$
k^{th} measurement	$X_i^{(k)}$	$I_k(x_i)\, I_k(s_2) \dots I_k(s_n)$
sample mean	$\overline{X}_i = \frac{1}{N}\sum_{k=1}^{N} X_i^{(k)}$	$\mathfrak{G}^{(n)}(x_i) = \frac{1}{N}\sum_{k=1}^{N} I_k(x_i) I_k(s_2) \dots I_k(s_n)$
first moment	$\langle \overline{X}_i \rangle = \langle X_i \rangle$	$\langle \mathfrak{G}^{(n)}(x_i) \rangle = \langle I(x_i)\, I(s_2) \dots I(s_n) \rangle \propto G^{(n)}(x_i, s_2, \dots, s_n)$
second moment	$\langle \overline{X}_i \overline{X}_j \rangle = \frac{1}{N}\langle X_i X_j \rangle$	$\langle \mathfrak{G}^{(n)}(x_i) \mathfrak{G}^{(n)}(x_j) \rangle = \frac{1}{N}\langle I(x_i) I(x_j)\, I(s_2)^2 \dots I(s_n)^2 \rangle \propto \frac{1}{N} G^{(2n)}(x_i, x_j, s_2, s_2, \dots, s_n, s_n)$

Table 5.1: Statistical quantities and their counterparts in the correlation functions. The random variable X_i is the product of the intensity measurements at positions x_i, $s_2, \dots, s_n$, and the index $i \in \{1, \dots, M\}$ runs over all the detector positions (pixels). After N measurements, we define a sample mean $\overline{X}_i$ that is itself a fluctuating quantity. This is not to be confused with the first moment $\langle X_i \rangle$, which is not a random variable. Since the sample mean is an unbiased estimate of the first moment, the expectation value of the sample mean is equal to the first moment. The second moment of any pair of sample averages is *not* equal to the second moment of the variables but instead N^{-1} times the second moment. This quantifies the intuition that taking more data (increasing N) reduces the variation of the sample averages. As $N \to \infty$ the sample averages coincide with the expectation values and the $\overline{X}_i$ are no longer random variables. The first and second moments are only proportional to the correlation functions $G^{(n)}$ and $G^{(2n)}$ due to the efficiency factor in the measured intensity in Eq. (5.18).

where we define the first term (depending on $\boldsymbol{\mu}$) as $[\boldsymbol{\mathcal{I}}_1(\boldsymbol{\theta})]_{ij}$, and the second term as $[\boldsymbol{\mathcal{I}}_2(\boldsymbol{\theta})]_{ij}$. Since the optical field exhibits strong transverse correlations in the detection plane, the covariances will not be negligible, and we must therefore explicitly evaluate these covariances in order to perform maximum likelihood estimation. In the next section we explain how the correlation functions are incorporated into the parameter estimation procedure. Finally, in order to find the maximum likelihood estimate we use the method of scoring as outlined in section 3.1.3. The process is described by the recursion relation Eq. (3.31). In order to begin the scoring algorithm we require an initial value $\boldsymbol{\theta}^{(0)}$. Provided the initial value is sufficiently close to the actual value, the algorithm should continue without difficulty. If no prior knowledge exists about the parameters to be estimated, approximate values can be obtained by simple methods that by no means achieve the CRB that can then be used as the initial values $\boldsymbol{\theta}^{(0)}$.

5.3 Measuring the correlation functions

For simplicity we suppress the y dependence in $\mathbf{r}_i = (x_i, y_i)$ and consider only the one-dimensional problem where a single "moving" detector x_i scans across a set of M discrete positions $x_1, \dots, x_M$ and the remaining $n-1$ detectors are kept fixed

(see Fig. 5.2b-e). We refer to the fixed detectors as the reference pixels and write the reference pixel positions as $x_2 = s_2, \ldots, x_n = s_n$. We distinguish between two detector schemes, namely one where all reference pixels are identical (scheme 1), and one where all reference pixels are different (scheme 2). Taking N images means performing N independent measurements of the intensity $I(x_i)$ at pixel x_i and calculating the sample average of the intensity moments

$$\mathfrak{G}^{(n)}(x_1, \ldots, x_n) = \frac{1}{N} \sum_{k=0}^{N} \prod_{i=1}^{n} I_k(x_i) \,, \tag{5.17}$$

where $I_k(x_i)$ is the k^{th} measurement of the intensity at position x_i. Given the reference pixels $\{s_2, \ldots, s_n\}$, we can abbreviate $\mathfrak{G}^{(n)}(x_i, s_2, \ldots, s_n) \equiv \mathfrak{G}^{(n)}(x_i)$, and the data that is used in the estimation procedure is $\mathbf{x} = (\mathfrak{G}^{(n)}(x_1), \ldots, \mathfrak{G}^{(n)}(x_M))$. This is still quite a general description and includes, for example, the experimental arrangement used in Ref. [Oppel et al., 2012]. We use $\mathfrak{G}^{(n)}$ to denote a measurement of the correlation function. This is not to be confused with the true correlation function $G^{(n)}$ as given by Eq. (5.8). It is an important distinction since the measured correlation $\mathfrak{G}^{(n)}(x_i)$ is a random variable due to the finite size of the sample N, whereas $G^{(n)}(x_i)$ is the expectation value of the correlation function, only in the limit $N \to \infty$ do the two coincide. The relation between the standard statistical quantities and the correlation functions are collated in Table 5.1.

In addition to the statistical noise due to the finite sample size N, any realisable detection scheme will introduce additional noise into the measurements. One important source of noise is reduced detection efficiency of the pixels. Often this is treated as a constant parameter η. However, when calculating the intensity correlations we necessarily sample the higher moments of the detector noise. It is therefore important that we acknowledge the random nature of the noise in order to correctly deduce its effects. Physically we would expect the noise to be sharply peaked around some constant value with some small but non-zero variance. We would also expect the random noise to be independent across the pixel array since the pixels themselves are independent. We model this additional noise as uncorrelated, normally distributed noise with mean and variance $\langle \eta(x_i) \rangle = \nu$ and $\langle \eta(x_i)^2 \rangle - \langle \eta(x_i) \rangle^2 = \varsigma^2$ respectively [1]. We can therefore write

$$I_k(x_i) = \eta_k(x_i) \tilde{I}_k(x_i) \,, \tag{5.18}$$

where $\tilde{I}_k(x_i)$ is the k^{th} realisation of the random intensity at pixel x_i as measured by an ideal detector, and $\eta_k(x_i)$ is the k^{th} realisation of the noise at pixel x_i. The

[1] The particular model chosen for the noise is in some sense arbitrary, as we will show in appendix B it is only necessary to determine up to the $2n^{\text{th}}$ moment of the noise distribution, treating each new moment as a new parameter to be estimated. The use of Gaussian noise allows us to use only two parameters to calculate any moment, a feature not shared by a general distribution. However, the additional parameters introduced by a different noise distribution (higher moments) can be added to the list of estimation parameters, therefore increasing the size of the estimation problem whilst still keeping the problem tractable.

expectation of $\mathfrak{G}^{(n)}(x_i)$ is given by

$$\begin{aligned}\mu_i &= \langle \mathfrak{G}^{(n)}(x_i, s_2 \ldots, s_n)\rangle \\ &= \frac{1}{N}\sum_{k=1}^{N}\langle I_k(x_1) I_k(s_2)\ldots I_k(s_n)\rangle \\ &= \frac{1}{N}\sum_{k=0}^{N}\langle \tilde{I}_k(x_i)\tilde{I}_k(s_2)\ldots\tilde{I}_k(s_n)\rangle\langle \eta_k(x_i)\eta_k(s_2)\ldots\eta_k(s_n)\rangle\,, \end{aligned} \tag{5.19}$$

which is to be used in Eq. (5.15). We assume that the noise and intensity are stationary random variables, so we can immediately perform the sum removing the factor N^{-1}. The first factor in Eq. (5.19) is, by definition, the intensity correlation $\langle \tilde{I}_k(x_i)\tilde{I}_k(s_2)\ldots\tilde{I}_k(s_n)\rangle = G^{(n)}(x_1, s_2, \ldots, s_n)$. The second factor in Eq. (5.19) is in general some combination of moments of the normal distribution characterised by ν and ς.

The elements of the covariance matrix in Eq. (5.15) are

$$\mathbf{C}_{ij} = \mathrm{Cov}\left(\mathfrak{G}^{(n)}(x_i), \mathfrak{G}^{(n)}(x_j)\right), \tag{5.20}$$

where Cov denotes the covariance between the sample average of the intensity correlation at pixel x_i and x_j. Using Eq. (5.17) with the abbreviation of $\mathfrak{G}^{(n)}(x_i, s_2, \ldots, s_n)$, and the definition of μ_i in Eq. (5.19) this yields

$$\begin{aligned}\mathbf{C}_{ij} &= \frac{1}{N}\Big(\langle I(x_i)I(x_j)I(s_2)^2\ldots I(s_n)^2\rangle - \mu_i\mu_j\Big) \\ &= \frac{1}{N}G^{(2n)}(x_i, x_j, s_2, s_2, \ldots, s_n, s_n) \\ &\quad\times\left[\langle \eta(x_i)\eta(x_j)\eta(s_2)\eta(s_2)\ldots\eta(s_n)\eta(s_n)\rangle - \mu_i\mu_j\right]. \end{aligned} \tag{5.21}$$

We see that the covariances between our data $\mathfrak{G}^{(n)}(x_i)$ and $\mathfrak{G}^{(n)}(x_j)$ depend on correlation functions of order $2n$. The term $\langle \eta(x_i)\eta(x_j)\eta(s_2)\eta(s_2)\ldots\eta(s_n)\eta(s_n))\rangle$ is evaluated in appendix B. We now have a complete characterisation of the probability distribution $p(\mathbf{x}|\boldsymbol{\theta})$ and can therefore calculate the Fisher information to determine the lower bound on the variance of our estimates of $\boldsymbol{\theta}$ via the CRB and can also perform a maximum likelihood estimation procedure to estimate the dimensions of the source. In the next section we present numerical simulations of this estimation procedure.

5.4 Numerical Simulations

To determine the performance of our estimates, we produce simulations of the experiment shown in Fig. 5.1b, where a circular fibre tip of radius of $a = 100$ µm emits uniform thermal light of wavelength $\lambda = 633$ nm. The complex degree of coherence of such a source is given by Eq. (5.11). To simulate the experiment we again make use of the optical equivalence theorem and the P-representation. The P-representation for thermal light takes the form of a complex multivariate normal distribution. The simulation of the experiment is then performed by sampling from a $2M$-dimensional normal distribution corresponding to the real and imaginary parts of $\alpha(x_1), \ldots, \alpha(x_M)$, i.e., $\alpha(x) = a(x) + ib(x)$ with covariances

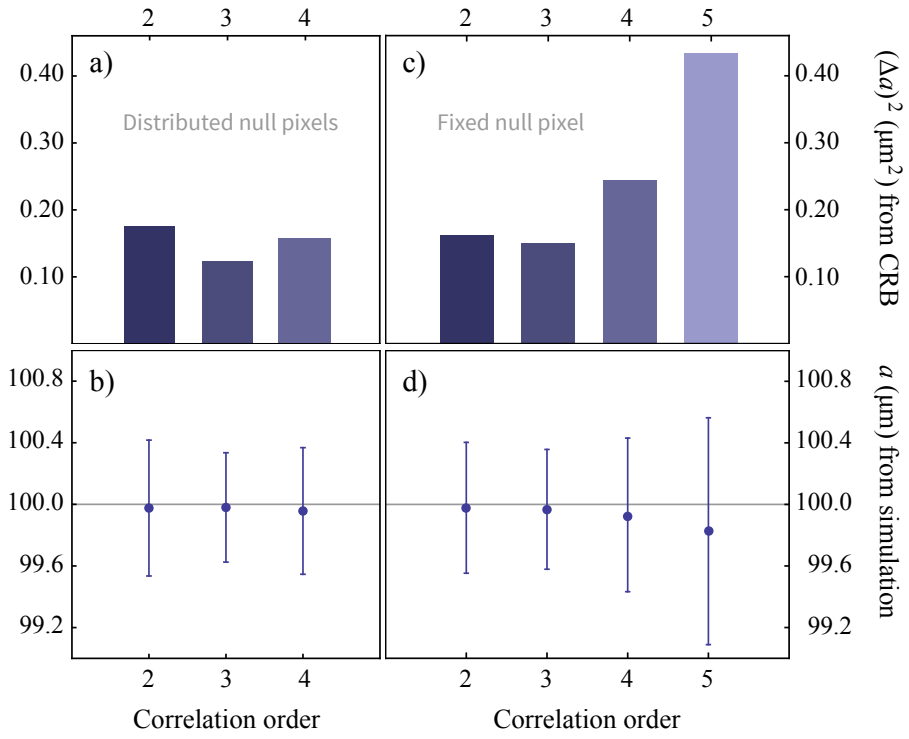

Figure 5.4: Results of the numerical estimation of the source diameter a (μm) and comparison with the Cramér-Rao bound (CRB). Choosing the reference pixels in a distributed manner as in Fig 5.2c, e leads to a CRB value of the variance as shown in (a), with the estimate and standard deviation shown in (b). Here we have chosen a detection efficiency $\nu = 0.5$ with $\varsigma = 0.01$. Due to the computational complexity of the problem only correlation orders up to $n = 4$ can be calculated. Choosing the central pixel as the reference pixel as in Fig 5.2b, d leads to a CRB value of the variance as shown in (c), with the estimate and standard deviation shown in (d). This configuration does not allow us to include detector efficiency as a random variable in the estimation procedure. For comparison, the numerical values in this figure are collated in Tables 5.2 and 5.3.

$$\begin{aligned} \langle a(x_i)a(x_j)\rangle &= \langle b(x_i)b(x_j)\rangle = \frac{1}{2}\langle I\rangle\, \gamma(x_i, x_j) \\ \langle a(x_i)b(x_j)\rangle &= 0\,. \end{aligned} \tag{5.22}$$

We have assumed a uniform far field intensity distribution of thermal sources, $\langle I(x_i)\rangle = \langle I(x_j)\rangle \equiv \langle I\rangle$. This simulates the far field intensity for a thermal source complete with all the intensity correlations.

We include the effect of pixel noise by adding an additional normal random variable to each of the intensities with mean ν and variance ς^2. The intensity correlations are then calculated from the simulated field by means of Eq. (5.17), which are in turn used to estimate the dimensions of the source parameters. We require averaging over a large set of data in order to apply the central limit theorem and treat the data as normally distributed. The parameter $\langle I\rangle$ is also treated as an unknown parameter to be estimated. The importance of this cannot be overstated: if we instead treat $\langle I\rangle$ as a constant, any slight deviation from the exact value can lead to catastrophic failure of the estimation procedure. It is therefore imperative that the unknown parameter $\langle I\rangle$ should be treated as a nuisance parameter in order

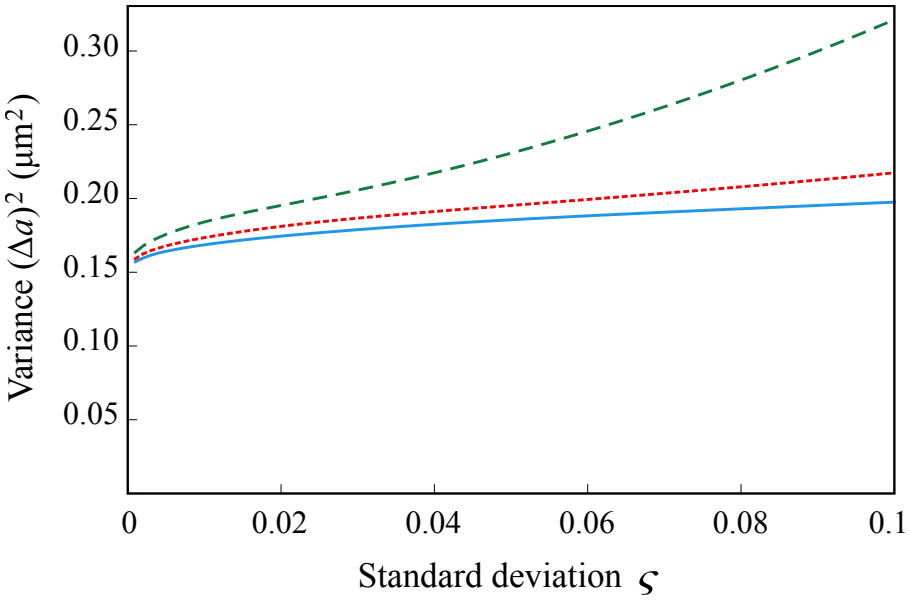

Figure 5.5: Variance $(\Delta a)^2$ of the estimator $\check{a}$ for $G^{(2)}$ against the standard deviation of the noise ς (dimensionless). Average detector efficiencies $\nu = 0.2$ (dashed), $\nu = 0.5$ (dotted), and $\nu = 0.9$ (full).

to perform the maximum likelihood estimation. Although there appear to be four unknown parameters: $\boldsymbol{\theta} = (a, \langle I \rangle, \nu, \varsigma)$, the factorisation of Eq. (5.25) allows us to combine the parameters $\langle I \rangle$ and ν into the new parameter $\langle I_{\text{eff}} \rangle = \nu \langle I \rangle$ which is the effective intensity recorded by the detector in the presence of inefficiencies characterised by $\langle \eta \rangle = \nu$.

Once we generated the simulated data from the $2M$-dimensional normal distribution described above and added the noise, we estimated the parameters a, $\langle I_{\text{eff}} \rangle$ and χ based on different orders of intensity correlations and the scoring method. The maximum likelihood estimation procedure was repeated 1000 times such that a statistical variance $(\Delta a)^2_{\text{Sim}}$ for the simulated data could be calculated, and we can compare this to the lower bound on the variance $(\Delta a)^2_{\text{CRB}}$ based on the Cramér-Rao bound:

$$(\Delta a)^2_{\text{CRB}} \geq [\boldsymbol{\mathcal{I}}]^{-1}_{aa} , \tag{5.23}$$

where we now must find the inverse of the 3×3 Fisher information matrix $\boldsymbol{\mathcal{I}}$. There are two main cases to consider.

5.4.1 Constant detector loss

First, we analyse the special case of a detection system with constant loss for each pixel ($\varsigma = 0$). For this particular noise model, the choice of reference pixel positions does not effect the noise terms in Eqs. (5.19)(5.21), since $\langle \eta_k(x_i) \eta_k(s_2) \dots \eta_k(s_n) \rangle = \nu^n$ for all choices of reference pixel positions. We find the choice $s_2 = s_3 = \cdots = s_n \equiv s = \lfloor M/2 \rfloor$ to be of particular interest since it simply involves taking powers of the measured intensity $I(s)^{n-1}$, and it is the central pixel on a one-dimensional CCD. This allows us to compare the effects of the post processing without the need to consider complications regarding the exact placement of the reference pixels, $s_2, \dots, s_n$. The inherent simplicity of this arrangement also allows us to calculate

the correlations up to arbitrary order since the correlation functions $G^{(n)}$ and $G^{(2n)}$ take the compact forms

$$\begin{aligned} G^{(n)}(x_i, s, \ldots, s) =& (n-1)![1 + (n-1)|\gamma(x_i, s)|^2], \\ G^{(2n)}(x_i, x_j, s, \ldots, s) =& (n-2)!\big\{1 + |\gamma(x_i, x_j)|^2 + (n-2) \\ & \times \big[2\mathrm{Re}(\gamma(x_i, x_j)\gamma(s, x_i)\gamma(x_j, s)) \\ & + |\gamma(x_i, s)|^2 + |\gamma(x_j, s)|^2 \\ & + |\gamma(x_i, s)|^2|\gamma(x_j, s)|^2\big]\big\}. \end{aligned} \tag{5.24}$$

Fig. 5.4c shows the Cramér-Rao lower bound on the variance for our estimate of a for the first four correlation functions $G^{(2)}$ to $G^{(5)}$, and Fig. 5.4d shows the estimate of a with the actual standard deviation Δa_{Sim}. The numerical results are also collated in Table 5.2.

Table 5.2: Results of the maximum likelihood estimation for the correlation functions $G^{(2)}$ to $G^{(5)}$ and the Cramér-Rao lower bounds. The estimation procedure was performed on 1000 simulated data sets.

n	Estimated a (μm)	$(\Delta a)^2_{\mathrm{Sim}}$ (μm^2)	$(\Delta a)^2_{\mathrm{CRB}}$ (μm^2)
2	99.978	0.181	0.162
3	99.968	0.151	0.150
4	99.932	0.249	0.244
5	99.826	0.543	0.434

Surprisingly, the best estimates of the spatial dimensions of the source occur for $n = 3$, and the estimates get progressively worse as the correlation order is increased beyond third order. Therefore, to extract the maximum amount of spatial information from our data, correlations of third order should be used. We stress the importance of this finding as it requires no additional measurements to be made other than those made to measure $G^{(2)}$. Indeed, in principle it would be possible to use the exact data used by Hanbury Brown and Twiss to measure the angular diameter of Sirius [Brown and Twiss, 1956a]. We also note that there is nothing in our treatment that uniquely picks out the spatial correlation functions, in the same manner we could equally discuss temporal correlations. Interestingly, estimates of the effective intensity $\langle I_{\mathrm{eff}} \rangle$ do not follow the same pattern as those for a. If we wish to estimate $\langle I_{\mathrm{eff}} \rangle$ the best performance is given by $G^{(2)}$, with higher orders performing worse.

5.4.2 Detector loss as a random variable

Second, we demonstrate the effect of a small non-zero ς, representing a system with uncertainty in the detector loss mechanism. The effect of this additional noise is shown in Fig. 5.5, where we plot the variance of the estimator for the second order intensity correlation function against the standard deviation ς. As expected, the addition of noise in the detection process reduces the precision in our estimator, but the question is whether this detector noise as a random variable is a better model than the fixed detector efficiency with $\varsigma = 0$. The results of the maximum likelihood estimation and the Cramér-Rao bound are given in Table 5.3.

In order to perform the estimation, first we must evaluate the second term in Eq. (5.19), which is an n^{th} moment of the noise distribution. Having considered the case where all reference pixels are the same in the previous section, we now restrict ourselves to only considering cases where no two reference pixel positions are the same, i.e. $s_2 \neq s_3 \neq \cdots \neq s_n$, such that we can determine the effect of separating the reference pixels. In this regime the n^{th} moment of the noise distribution, $\langle \eta(x_i)\eta(s_2)\ldots\eta(s_n)\rangle$ is given by

$$\begin{aligned}\langle \eta_k(x_i)\eta_k(s_2)\ldots\eta_k(s_n)\rangle &= \nu^n + \nu^{n-2}\sum_{j=2}^{n}\delta_{x_i,s_j}\varsigma^2 \\ &= \nu^n\left(1+\sum_{j=2}^{n}\delta_{x_i,s_j}\chi^2\right) \end{aligned} \tag{5.25}$$

where $\chi = \varsigma/\nu$. We therefore find it necessary to re-parameterise the problem using the parameters $\boldsymbol{\theta} = (a, \langle I\rangle, \nu, \chi)$. Eq. (5.25) represents an n^{th} moment of the noise distribution, and the Kronecker deltas arise from the independence of the distribution for individual pixels. The $2n^{\text{th}}$ moment of the noise distribution appearing in Eq. (5.21) is calculated in appendix B.

In order to find the optimum position of the reference pixels, we define $|s_i - s_{i+1}| = d$ and plot the standard deviation as a function of d, the separation between adjacent reference pixels. Fig. 5.6 shows the standard deviation for $G^{(2)}$ to $G^{(4)}$ as a function of d. Interestingly, the higher order correlations do not outperform $G^{(2)}$ for all values of d. For $G^{(3)}$ we find that the optimum positions correspond to separations where the two reference pixels become uncorrelated. This occurs whenever the complex degree of coherence between the two pixels is equal to zero. Since the complex degree of coherence for the system is proportional to the Bessel function J_1, the optimum separations d correspond to the zeroes of this function. For $G^{(4)}$ the exact position of the optimum is more complicated, due to the fact that the zeroes of J_1 are not uniformly distributed. However, the optimum positions are approximately located at the position where adjacent reference pixels are uncorrelated from their nearest neighbours. This should be expected to some extent as we can see from the following argument: A measurement of $G^{(3)}(x, s_1, s_2)$ contains information about the correlations between the fields at x and s_1, and the fields at x and s_2, therefore if the fields at s_1 and s_2 are highly correlated we would expect to gain very little from obtaining additional information regarding the correlations between x and s_2. However, the more independent s_1 and s_2 become the more we can expect to gain additional information from measurements of the correlations between x and s_2. Tab. 5.3 shows the variance of the estimators for the first three correlation functions as calculated from the CRB and directly measured in the simulations. We see that the measured variance in our estimators closely follows that obtained from the CRB.

Another interesting feature of Fig. 5.6 is the ability for $G^{(4)}$ to outperform $G^{(2)}$, a feature that does not occur for fixed reference pixels. This behaviour is reminiscent of the "magic angles" in Refs. [Thiel et al., 2007, Oppel et al., 2012], where the detectors had to be placed at specific positions (the magic angles) in order to obtain the $(n-1)$-fold increased sinusoidal modulation in the scanning detector.

The exact relation between the variance of our estimators and the correlation order also depends on the intensity distribution of the source. Fig. 5.7 shows the

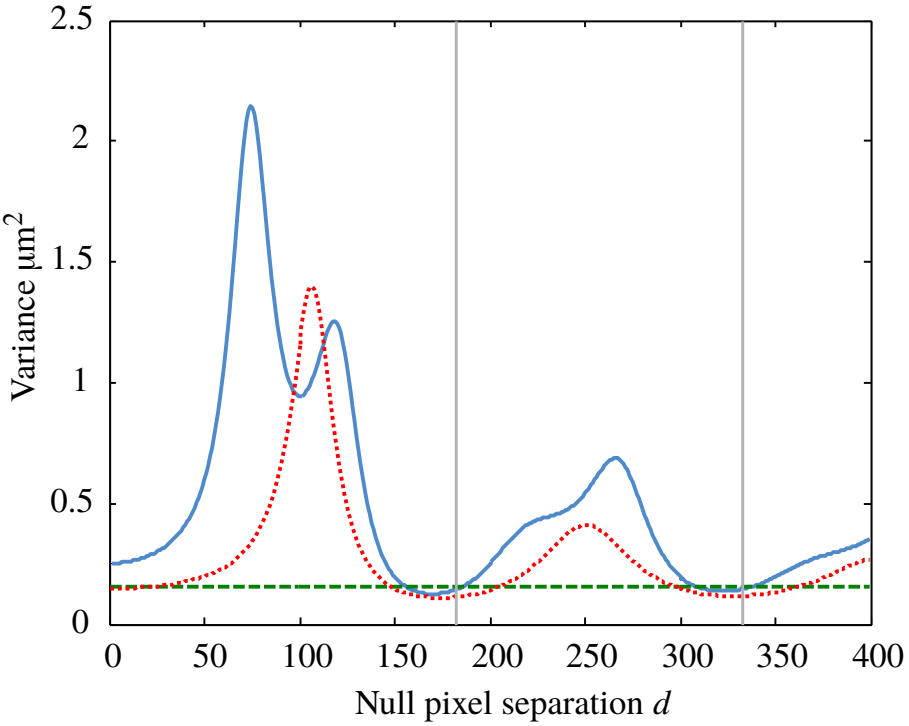

Figure 5.6: Standard deviation of the estimator $\check{a}$ for $G^{(2)}$ (dashed) , $G^{(3)}$ (dotted) and $G^{(4)}$ (solid). The grey vertical lines correspond to the first and second zeroes of the Bessel function J_1.

Table 5.3: Results of the maximum likelihood estimation for the correlation functions $G^{(2)}$ to $G^{(4)}$ and the Cramér-Rao lower bounds, with a reference pixel separation of $d = 182$ that corresponds to the first zero of J_1. The estimation procedure was performed on 1000 simulated data sets.

n	Estimated a (μm)	$(\Delta a)^2_{\text{Sim}}$ (μm^2)	$(\Delta a)^2_{\text{CRB}}$ (μm^2)
2	99.976	0.194	0.175
3	99.980	0.126	0.123
4	99.957	0.169	0.157

dependence of the variance as a function of the reference pixel separation for a slit of width $a = 200$ μm. The complex degree of coherence for such a geometry is given by the sinc function. Since the zeros of the sinc function are uniformly distributed, it is possible to achieve independence for all the reference pixels simultaneously. Fig. 5.7 shows that for the optimal choice of d the estimator for $G^{(4)}$ outperforms $G^{(2)}$ and is about as good as $G^{(3)}$.

5.5 Discussion and conclusions

We have discussed the exact role of higher order intensity correlations in regards to parameter estimation of the intensity distribution of thermal sources, and demonstrated that it is beneficial to post-process the data in such a way as to measure intensity correlations of order $n > 2$. We have also shown how the post processing can be optimised with respect to the placement of the reference pixels, to find the

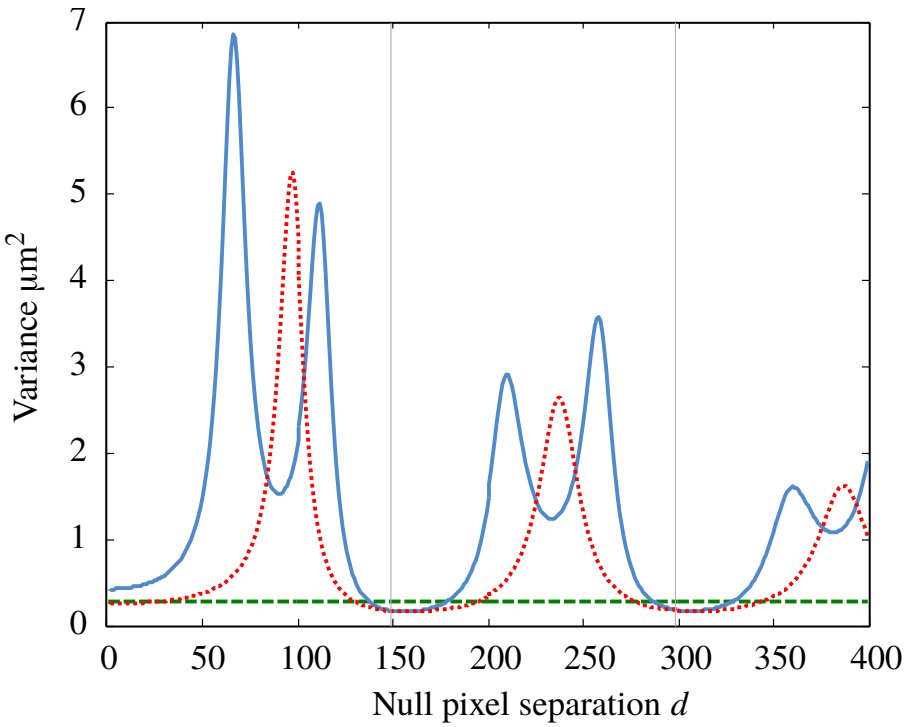

Figure 5.7: Standard deviation of the estimator $\check{a}$ for $G^{(2)}$ (dashed) , $G^{(3)}$ (dotted) and $G^{(4)}$ (solid). The grey vertical lines correspond to the first and second zeroes of the sinc function.

most informative measurements. A major benefit of this method is that it does not require particularly elaborate experimental arrangements, and when we can neglect any additional noise introduced by the detectors it becomes possible to increase the precision simply by taking powers of the measured intensities. Since we explicitly account for correlations between the data points, all the measurements can be made simultaneously, thus reducing the measurement time required to obtain the data. While we have framed the discussion in the context of detector pixels on a CCD camera, the same methods apply to any array of field detectors, including telescopes. By fully determining the PDF for measurements of intensity correlation functions, including the covariance matrix of the correlated data, we are able to determine the Fisher information for such experiments. This allows us to calculate the maximum achievable precision via the Cramér-Rao bound and also to saturate that bound by performing a maximum likelihood estimation.

The techniques presented here can in principle be used to estimate the source dimensions of objects that emit non-thermal light, including quantum emitters. In these cases the Gaussian moment theorem generally does not apply and extra care must be taken in the calculation of the correlation functions $G^{(n)}(x_1, s_2, \ldots, s_n)$. As long as the measured intensity is a random variable then the PDF for the data can be considered as a multivariate normal when averaged over many measurements. Also, the estimated parameters need not be spatial parameters of the source: as long as the PDF depends on the parameters to be estimated in a deterministic way our procedure can be used to perform the estimation. The inclusion of a realistic noise model allows the effect of detection noise to be accounted for and estimated, thus allowing the estimation to proceed in the presence of noise.

The experimental implementation of this procedure is not hugely challenging, but a number of critical requirements must be met. First, when measuring the intensity correlation functions, the integration time of the detectors must be well below the coherence time of the radiation to ensure that each measurement captures a single mode of the radiation allowing us to consider a single mode. In addition, the area of each detector (pixel) must be much smaller than the coherence area of the source such that a single pixel can be considered as measuring a single mode. This ensures that every detection event samples no more than a single mode volume (i.e., the speckle size). Second, our use of the central limit theorem implies that we have a large data set that provides a good signal to noise ratio. It is well-known that a maximum likelihood estimation procedure requires a good signal to noise ratio to produce good estimates [Kay, 1993]. Third, we have assumed uniform intensity across the detectors. In practice this can be challenging if the source is not sufficiently out of focus, or if there are spurious interference effects in the detector itself (e.g., an etalon effect due to the protective glass of a CCD camera). Finally, the detector noise is uncorrelated, which means that the pixel efficiencies are random, and there is no cross-talk between the pixels. The latter two requirements can in principle be included in the modelling of the experiment, but this comes at the cost of a significantly increased complexity of the correlation functions.

We notice from Fig. (5.4) that all of the estimates fall just below the actual value of $a = 100$ μm. Since all estimates are within one standard deviation of the true value this may well be due to statistical fluctuations but it is also possible that there is a small bias in the estimator. Presumably this would be due to any slight deviations away from a normal distribution as was assumed for the PDF for the data. To model the PDF as a multivariate normal requires the use of the central limit theorem, which is only strictly true in the limit of infinite averaging. If there is in fact a bias and it is caused by the non-normal nature of the PDF, we would expect this effect to disappear for larger data sets as the PDF approaches normality.

The ability for higher order correlations to obtain more information can often be a source of confusion. This is most clearly demonstrated when the reference pixels are all the same $s_2 = s_3 = \cdots = s_n$. We can take the output of two photodetectors and simply by taking powers of one of the outputs, we gain a more precise estimate of angular diameter of the source. We can understand this increase in precision by comparing the measurements of the intensity and second order correlation function as in the original HBT experiment. We use the same set of data when measuring the intensity as we use to measure the second order intensity correlation and yet a measurement of the intensity reveals very little information about the source since the intensity of a thermal source in the far field is constant across x, the position of the detector. In contrast, the second order intensity correlation function is highly dependent on x which allows for a much more precise estimate of the angular diameter to be achieved, see Fig. 5.3. When considered as another method of post processing the data, it is no more surprising that higher order correlations outperform the second order correlation than the second order correlation outperforming the intensity.

Our method is rather general, and should apply to sources of light that are not thermal. Future work will focus on (multi-mode) squeezed light and single photon sources. Since all correlation functions can be determined from the same data set, it would be advantageous to combine all of the estimates achieved via different n into a

single estimate. In order to do this properly we would need to know exactly how all of the individual estimates are correlated to determine the appropriate weighting for the combined estimate and its error. The difficulty in determining the correlation is in knowing how the maximum likelihood estimation procedure affects the correlation, if at all. Once this is known, it should be possible to determine the weighting and obtain the final estimate.

In conclusion, we found that higher order correlation functions can substantially improve estimation of the parameters that characterise the geometry of a thermal light source. As long as the source and the detection system are properly modelled, the procedure can be implemented with current technology. Our techniques extend immediately to other forms of Gaussian light, and potentially more general classical and quantum light sources as well.

Chapter 6

Obtaining Spatial Information From Far Field Sources

We often wish to determine the spatial properties of a radiation source from measurements of the radiation in the far field. There are numerous measurements we can make to aid us in determining these spatial properties. In this chapter we focus on the problem of determining the spatial properties of a source of thermal radiation from measurements of the radiation in the far field although our method only relies on the detected radiation being diagonal in the number basis and isotropic (at least across the modes that are measured). We calculate the symmetric logarithmic derivative for such a system and determine under what circumstances the expressions lead to a well behaved quantum Fisher information matrix. We also determine under what circumstances the quantum Fisher information (QFI) is ill-defined and in fact becomes singular. This allows us to demonstrate certain physical restrictions upon the measured radiation and therefore informs us as to how information about the spatial distribution of a source is conveyed to the far field.

At the end of this chapter we examine how, using the method developed here, we might come to obtain a complete calculation of the quantum Fisher information. We also give a brief account of an entirely different approach that may lead to a successful calculation of the QFI and point out what difficulties we anticipate to encounter.

6.1 Blackbody radiation

We start by discussing the properties of blackbody radiation. It is well known that blackbody radiation is described by the density operator [Mandel and Wolf, 1995]

$$\rho = \bigotimes_{\mathbf{k},s} \sum_{n_{\mathbf{k}s}=0}^{\infty} \frac{\langle \hat{n}_{\mathbf{k}s} \rangle^{n_{\mathbf{k}s}}}{(1+\langle \hat{n}_{\mathbf{k}s} \rangle)^{n_{\mathbf{k}s}+1}} |n_{\mathbf{k}s}\rangle \langle n_{\mathbf{k}s}| \tag{6.1}$$

$$\equiv \bigotimes_{\mathbf{k},s} \sum_{n_{\mathbf{k}s}=0}^{\infty} p(n_{\mathbf{k}s}) |n_{\mathbf{k}s}\rangle \langle n_{\mathbf{k}s}|, \tag{6.2}$$

where the product $\bigotimes_{\mathbf{k},s}$ labels all of the occupied modes, $\mathbf{k}$ labels the wave vector and s labels the polarisation. The product should be understood as a tensor product

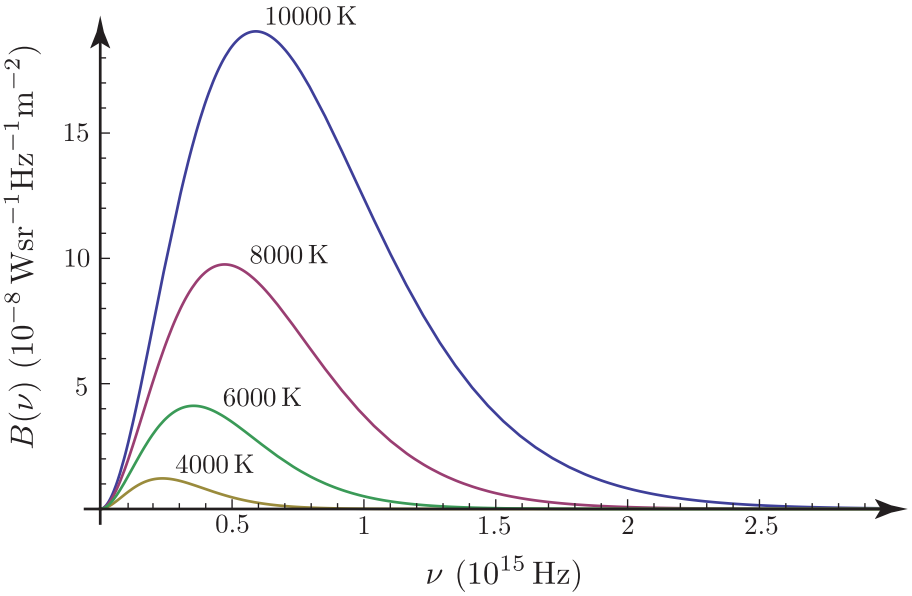

Figure 6.1: The Planck distribution describing the energy spectrum of the radiation emitted by a blackbody. The peak of the distribution is proportional to the temperature of the blackbody T and can be determined via Wien's displacement law.

over the individual density operators describing each mode

$$\rho_{\mathbf{k}s} = \sum_{n_{\mathbf{k}s}} p(n_{\mathbf{k}s}) |n_{\mathbf{k}s}\rangle \langle n_{\mathbf{k}s}|$$
$$\rho = \bigotimes_{\mathbf{k},s} \rho_{\mathbf{k}s}, \tag{6.3}$$

which therefore describe independent states of the field. We can consider the product as split into two parts $\bigotimes_{\mathbf{k}} = \bigotimes_k \bigotimes_{[\mathbf{k}]}$ where the first product is taken over wavevectors of different magnitude and the second product is constrained to wavevectors of magnitude k but specifying unique directions. By noticing that the magnitude of the wave vector is proportional to the frequency ν we can write the product as $\bigotimes_\nu \bigotimes_{[\mathbf{k}]}$.

A blackbody emits isotropically in space meaning that all $\mathbf{k}$ modes are occupied and the energy of the state follows a Planck distribution in frequency (see Fig. 6.1)

$$B(\nu) = \frac{2h\nu^3}{c^2} \frac{1}{\mathrm{e}^{\beta h\nu} - 1}, \tag{6.4}$$

where h is the Planck constant, $\beta = 1/k_B T$ and k_B is the Boltzmann constant. The Planck distribution gives the amount of energy emitted by a blackbody per unit time, per steradian, per Hertz, per square meter of the source, $\mathrm{Wsr}^{-1}\mathrm{Hz}^{-1}\mathrm{m}^{-2}$[1]. To calculate the actual energy output of a blackbody of area A_S, over a frequency range

[1] A steradian is defined in analogous way to the radian in 3 dimensions. As the definition of the radian is the angle subtended at the centre of a circle by an arc that has length equal to the radius of the circle, the steradian is the solid angle subtended at the centre of a sphere by a cone that encompasses an area of the radius squared over the surface of the sphere.

(ν_1, ν_2) and over a solid angle Ω therefore requires evaluation of the integral

$$E = \int_{\nu_1}^{\nu_2} A_S \Omega \tau B(\nu)\, \mathrm{d}\nu \tag{6.5}$$

where τ is the time the blackbody is observed for.

It is natural to ask how many modes a blackbody excites. To answer this we make use of another well know result that states that for a single mode in thermal equilibrium $\rho_{\mathbf{k}s}$ the expectation value of the number operator is

$$\langle \hat{n} \rangle = \frac{1}{\mathrm{e}^{\beta h\nu} - 1}. \tag{6.6}$$

where $\nu = |\mathbf{k}|c/2\pi$. This gives the average number of photons in the mode. By multiplying by the energy per photon and then by the number of modes we arrive at the total energy of the system. For a small frequency range $\Delta\nu$

$$h\nu \frac{1}{\mathrm{e}^{\beta h\nu} - 1} \times \mathbb{N}(\Delta\nu) = A_S \Omega \tau B(\nu) \Delta\nu, \tag{6.7}$$

where $\mathbb{N}(\Delta\nu)$ is the number of modes in the frequency range of width $\Delta\nu$. Eq. (6.7) implies

$$\mathbb{N}(\Delta\nu) = 2 A_S \Omega \tau \frac{\nu^2}{c^2} \Delta\nu. \tag{6.8}$$

This result was first stated by Gabor [Gabor, 1961]. The appearance of time τ in the above equations reflects the fact that, by measuring the state for longer, we measure more temporal modes as the photons become independent at times greater than the coherence time. It was necessary to restrict $\Delta\nu$ to be small such that the variation in $B(\nu)$ is negligible allowing us to approximate the integral by $\int B(\nu)\, \mathrm{d}\nu \approx B(\nu)\Delta\nu$. There is also a practical need in making the preceding approximation. Integration over the frequency implies that the frequency can be divided into arbitrarily small intervals with an energy density proportional to $B(\nu)$. However in any practical measurement this will not be the case. In fact the smallest distinguishable frequency interval is proportional to the inverse of the measurement time τ. We therefore suggest that the state be subdivided into distinguishable frequency modes [2] of width $1/\tau$ each consisting of $2A_S\Omega\nu^2/c^2$ modes differing by the exact direction of their wavevector [Levitin and Toffoli, 2014].

6.2 Frequency mode representation of ρ

We can now express the state ρ in a slightly different form to that of Eq.(6.1). We write the state as a product of independent frequency modes with a mode spacing $1/\tau$ as outlined above

$$\rho = \bigotimes_{\nu} \rho_\nu, \tag{6.9}$$

[2] Each frequency mode is itself a product of all the occupied $\mathbf{k}$ modes satisfying $|\mathbf{k}| = 2\pi\nu/c$.

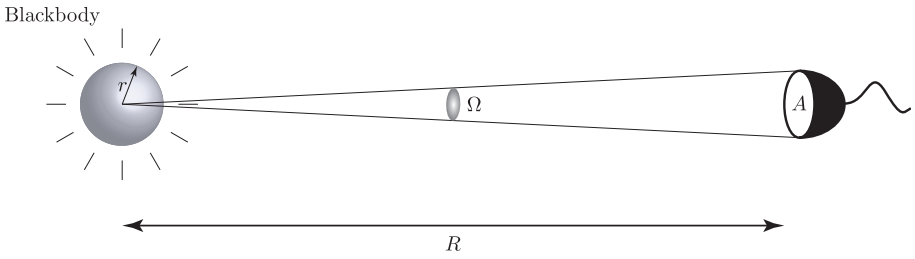

Figure 6.2: Demonstrating the arrangement. The blackbody emits isotropically radiation that follows the Planck distribution. A detector at a distance R in the far field observes the state over the detection time τ and the detection area A creating a solid angle $\Omega = A/R^2$. The area of the source, $A_S = \pi r^2$ is the projection of the source onto a plane.

with

$$\rho_\nu \equiv \sum_{n_\nu=0}^{\infty} p(n_\nu)|n_\nu\rangle\langle n_\nu|, \tag{6.10}$$

where the product $\bigotimes_\nu$ now runs over all the distinguishable frequency modes present and the product over $\mathbf{k}$ modes of the same energy has been absorbed by the kets $|n_\nu\rangle$. Since the density matrices for the individual $\mathbf{k}$ modes which comprise the ν modes are themselves diagonal in the Fock basis, the density matrices for the ν modes are also diagonal in this basis.

Since all of the modes are independent, the total probability of getting n_ν photons of frequency ν is

$$p(n_\nu) = \sum_{\{m\}} p(\{m\})\delta_{m,n_\nu}, \tag{6.11}$$

where the sum over $\{m\}$ is a sum over all $\mathbf{k}$ modes of frequency ν and therefore sums over all the possible configurations of the photons across the allowed $\mathbf{k}$ modes, the delta function picking out only those configurations where the total number of photons is n_ν. The notation $\{m\}$ therefore represents a particular configuration of m total photons occupying the D_ν modes. In other words, $\{m\}$ represents a tuple of length D_ν which specifies a number of photons for each mode, $\{m\} = \{m_1, \ldots, m_{D_\nu}\}$ and the sum is over every possible configuration. The probability $p(\{m\})$ of any particular configuration is given by

$$p(\{m\}) = \prod_{[\mathbf{k}],s} p(n_{\mathbf{k}s}), \tag{6.12}$$

where the product is again restricted to all $\mathbf{k}$ modes of the same energy $k = 2\pi\nu/c$. The probability distribution of the photon number for a thermal mode, $p(n_{\mathbf{k}s})$, is given by the Bose-Einstein distribution

$$p(n_{\mathbf{k}s}) = \frac{\langle\hat{n}_{\mathbf{k}s}\rangle^{n_{\mathbf{k}s}}}{(1+\langle\hat{n}_{\mathbf{k}s}\rangle)^{n_{\mathbf{k}s}+1}} \tag{6.13}$$

where $\langle\hat{n}_{\mathbf{k}s}\rangle$ is the average number of photons in the mode $\mathbf{k}, s$, which is given by Eq. (6.6). We note that this does not depend on the exact value of $\mathbf{k}$ except

through its absolute value $|\mathbf{k}| = k$, we therefore find that each term in the product of Eq. (6.12) gives the same contribution allowing us to write

$$p(\{m\}) = p(n_{\mathbf{k}s})^{D_\nu}, \tag{6.14}$$

where we have defined D_ν to be the number of $\mathbf{k}$ modes at frequency ν.

Substituting from Eq. (6.13) into Eq. (6.14), we find

$$p(\{m\}) = \frac{\langle \hat{n}_{\mathbf{k}s} \rangle^{n_\nu}}{(1 + \langle \hat{n}_{\mathbf{k}s} \rangle)^{n_\nu + D_\nu}}, \tag{6.15}$$

where $n_\nu \equiv D_\nu n_{\mathbf{k}s}$. As this depends only upon the total number of photons n_ν and not upon the actual configuration, each term that contributes to the sum in Eq. (6.11) gives the same result, namely Eq. (6.14). The probability $p(n_\nu)$ is therefore given by $\mathcal{N} p(\{m\})$ where $\mathcal{N}$ is the number of ways of arranging n_ν photons across D_ν modes. We find

$$p(n_\nu) = \frac{(n_\nu + D_\nu - 1)!}{(D_\nu - 1)! n_\nu!} \frac{\langle \hat{n}_{\mathbf{k}s} \rangle^{n_\nu}}{(1 + \langle \hat{n}_{\mathbf{k}s} \rangle)^{n_\nu + D_\nu}}, \tag{6.16}$$

which is the probability of measuring n_ν photons of frequency ν given that there are D_ν unique $\mathbf{k}$ modes in which the photons are distributed where each mode has the same average occupation number $\langle n_{\mathbf{k}s} \rangle$ [Mandel, 1959]. The number of $\mathbf{k}$ modes of a source of area A_S viewed under a solid angle Ω is equal to [Levitin and Toffoli, 2014]

$$D_\nu = 2 \frac{A_S \Omega \nu^2}{c^2}. \tag{6.17}$$

The factor of 2 in the above equation accounts for polarisation allowing us to omit s from the following treatment to simplify notation. In going from Eq. (6.8) we have cancelled the factor $\Delta\nu$ with the $1/\tau$. In doing so we insist that whatever time interval the state is measured over inversely defines the frequency interval $\Delta\nu$, which ensures that whatever time interval is chosen, only one temporal mode is measured.

6.3 The quantum Fisher information for blackbody radiation

To calculate the QFI for the spatial parameters of the source, we need to take the derivative of ρ with respect to the parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_r\}$, which parameterise the spatial properties of the source. The QFI matrix is given by [Helstrom, 1976]

$$[\boldsymbol{\mathcal{I}}_Q]_{ij} = \mathrm{Tr}[(\partial_i \rho) \mathscr{L}_j], \tag{6.18}$$

where $\mathscr{L}_j$ is the symmetric logarithmic derivative with respect to the parameter θ_j and $\partial_i \equiv \frac{\partial}{\partial \theta_i}$.

Calculation of the QFI therefore relies upon calculating the symmetric logarithmic derivative (SLD). In the basis that diagonalises ρ the SLD is given by

$$\mathscr{L}_j = 2 \sum_{l,m} \frac{\langle l | \partial_j \rho | m \rangle}{p(l) + p(m)} |l\rangle \langle m| \tag{6.19}$$

where $p(l)$ and $|l\rangle$ are the eigenvalues and eigenbasis of ρ respectively. In general the eigenvalues and eigenbasis of ρ will change with the parameters $\boldsymbol{\theta}$, therefore when we take the derivatives we must be careful to ensure that we take this into account. However, we will avoid explicitly writing the $\boldsymbol{\theta}$ dependence in both the eigenvalues and eigenbasis for brevity. Since each individual $\mathbf{k}$ mode that constitutes ρ is diagonal in the number basis, ρ itself is diagonal in the basis $|\{l\}\rangle = \bigotimes_{\mathbf{k}} |l_{\mathbf{k}}\rangle$, where again the notation $\{l\}$ stands for a tuple of length D_ν denoting the the number of photons in mode $\mathbf{k}$. The matrix elements $\langle\{l\}|\partial_j\rho|\{k\}\rangle$ are

$$\begin{aligned}[\partial_j\rho]_{\{l\}\{k\}} &= \langle\{l\}| \sum_\nu \partial_j\rho_\nu \bigotimes_{\mu\neq\nu} \rho_\mu |\{k\}\rangle \\ &= \sum_\nu \langle l_\nu|\partial_j\rho_\nu|k_\nu\rangle \bigotimes_{\mu\neq\nu} p(l_\mu)\delta_{l_\mu,k_\mu}. \end{aligned} \tag{6.20}$$

The SLD then becomes

$$\mathscr{L}_j = 2\sum_\nu \sum_{l_\nu,k_\nu} \frac{\langle l_\nu|\partial_j\rho_\nu|k_\nu\rangle}{p(l_\nu)+p(k_\nu)} |l_\nu\rangle\langle k_\nu|. \tag{6.21}$$

Using this form for the SLD and Eq. (6.18), the QFI is

$$[\boldsymbol{\mathcal{I}}_Q]_{ij} = \sum_\nu \mathrm{Tr}\left[2\partial_i\rho_\nu \sum_{l_\nu,k_\nu} \frac{\langle l_\nu|\partial_j\rho_\nu|k_\nu\rangle}{p(l_\nu)+p(k_\nu)} |l_\nu\rangle\langle k_\nu|\right]. \tag{6.22}$$

The QFI is therefore the sum of the QFI for each mode, we will omit the mode label and write the QFI for a single mode ν as

$$[\boldsymbol{\mathcal{I}}_Q]^{(\nu)}_{ij} = \mathrm{Tr}\left[2\partial_i\rho \sum_{l,k} \frac{\langle l|\partial_j\rho|k\rangle}{p_l+p_k} |l\rangle\langle k|\right]. \tag{6.23}$$

Upon inserting Eq. (6.20) into Eq. (6.23) we find

$$[\boldsymbol{\mathcal{I}}_Q]^{(\nu)}_{ij} = 2\sum_{l,k} \frac{\langle k|\partial_j\rho|l\rangle\langle l|\partial_i\rho|k\rangle}{p_l+p_k}, \tag{6.24}$$

where the QFI now depends only on the probabilities p_l and the matrix elements $\langle l|\partial_i\rho|k\rangle$.

Evaluating first the matrix elements $\langle k|\partial_i\rho|k\rangle$, we find

$$\langle l|\partial_i\rho|k\rangle = \sum_n \partial_j p_n \delta_{l,n}\delta_{n,k} + p_n \langle l|\left(\partial_j|n\rangle\langle n|\right)|k\rangle, \tag{6.25}$$

where we are careful to remember that the basis $|n\rangle$ is the basis that diagonalises ρ and is therefore dependent on $\boldsymbol{\theta}$. Remembering that the state $|n\rangle$ is defined by $|n\rangle = \bigotimes_{\mathbf{k}_\nu} |n_{\mathbf{k}}\rangle$ where the product is again constrained to wave vectors that satisfy $|\mathbf{k}| = k = 2\pi\nu c$, which we denote as $\mathbf{k}_\nu$. We are now faced with the issue that the number of $\mathbf{k}$ modes at a particular frequency ν, D_ν, is dependent on the spatial properties of the source. To proceed we calculate $\langle l_{\mathbf{k}}|\left(\partial_i|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|\right)|k_{\mathbf{k}}\rangle$ and find that it is not dependent on $\mathbf{k}$ except through the magnitude $|\mathbf{k}|$, and hence we may write

$$\langle l|\left(\partial_i|n\rangle\langle n|\right)|k\rangle = \langle l_{\mathbf{k}}|^{D_\nu}\left(\partial_i|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|^{D_\nu}\right)|k_{\mathbf{k}}\rangle^{D_\nu}, \tag{6.26}$$

where both D_ν and $|n_\mathbf{k}\rangle\langle n_\mathbf{k}|$ depend on θ_j.

Since θ_i is a spatial property of the source, we find it convenient to express $|n_\mathbf{k}\rangle$ in the spatial basis

$$|n_\mathbf{k}\rangle = \int_A \mathrm{d}\mathbf{x}_1 \ldots \int_A \mathrm{d}\mathbf{x}_n \, \psi_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)|\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle, \tag{6.27}$$

where the limits on the integral are imposed by the area over which we detect the state ρ, $\psi_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is the spatial wavefunction of the n-photon Fock state in mode $\mathbf{k}$ and $|\mathbf{x}_1, \ldots, \mathbf{x}_n\rangle = n!^{-1}\hat{a}^\dagger(\mathbf{x}_1)\ldots\hat{a}^\dagger(\mathbf{x}_n)|0\rangle$. We can relate the spatial coordinates $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in the detection plane to the spatial coordinates $\mathbf{y}_1, \ldots, \mathbf{y}_n$ in the source plane via the appropriate Green's function, $G(\mathbf{y}_i, \mathbf{x}_i)$. The n photon Fock state then becomes

$$\begin{aligned}|n_\mathbf{k}\rangle =& \frac{1}{\sqrt{n!N}} \int_A \mathrm{d}^n\mathbf{x} \int_S \mathrm{d}^n\mathbf{y} \, \phi_n(\mathbf{y}_1, \ldots, \mathbf{y}_n) \\ &\times G(\mathbf{y}_1, \mathbf{x}_1)\ldots G(\mathbf{y}_n, \mathbf{x}_n)\hat{a}^\dagger(\mathbf{y}_1)\ldots\hat{a}^\dagger(\mathbf{y}_n)|0\rangle,\end{aligned} \tag{6.28}$$

where $\sqrt{N}$ is the normalisation factor and the subscript S denotes integration over the source.

Upon expanding $\langle l_\mathbf{k}|\left(\partial_i|n_\mathbf{k}\rangle\langle n_\mathbf{k}|\right)|k_\mathbf{k}\rangle$, we find

$$\langle l_\mathbf{k}|\left(\partial_i|n_\mathbf{k}\rangle\langle n_\mathbf{k}|\right)|k_\mathbf{k}\rangle = \langle l_\mathbf{k}|\partial_i|n_\mathbf{k}\rangle\delta_{n,k} + (\partial_i\langle n_\mathbf{k}|)|k_\mathbf{k}\rangle\delta_{l,n}. \tag{6.29}$$

Using Eq. (6.28) we see that the inner product $\langle l_\mathbf{k}|\partial_i|n_\mathbf{k}\rangle$ is proportional to an anti-normally ordered product of creation and annihilation operators

$$\begin{aligned}\langle l_\mathbf{k}|\partial_i|n_\mathbf{k}\rangle &\propto \langle 0|\hat{a}(\mathbf{y}'_1)\ldots\hat{a}(\mathbf{y}'_l)\hat{a}^\dagger(\mathbf{y}_1)\ldots\hat{a}^\dagger(\mathbf{y}_n)|0\rangle \\ &= \delta_{n,l}\sum_{\sigma\in S_n}\prod_{i=1}^{n}\delta(\mathbf{y}_i - \mathbf{y}'_{\sigma(i)}),\end{aligned} \tag{6.30}$$

where S_n is the symmetric group containing all the permutations of the set $\{1, \ldots, n\}$. Since we integrate over all coordinates $\mathbf{y}_1 \ldots \mathbf{y}_n$ and $\mathbf{y}'_1 \ldots \mathbf{y}'_n$ each of the terms in the sum over S_n gives the same result after an arbitrary re-labelling of variables. Due to the Kronecker delta $\delta_{n,l}$, we need only consider the term $\langle n_\mathbf{k}|\partial_i|n_\mathbf{k}\rangle$:

$$\begin{aligned}\langle n_\mathbf{k}|\partial_i|n_\mathbf{k}\rangle = \frac{1}{\sqrt{N}}\int_A \mathrm{d}^n\mathbf{x}' \, \partial_i\Bigg[\frac{1}{\sqrt{N}}\int_A \mathrm{d}^n\mathbf{x} \int_S \mathrm{d}^n\mathbf{y} \, |\phi_n(\mathbf{y}_1, \ldots, \mathbf{y}_n)|^2 \\ \times \prod_{i=1}^{n} G^*(\mathbf{y}_i, \mathbf{x}'_i)G(\mathbf{y}_i, \mathbf{x}_i)\Bigg],\end{aligned} \tag{6.31}$$

where we have integrated over all n of the $\mathbf{y}'$ coordinates, making use of the Dirac delta functions in Eq. (6.30).

The inner product now becomes

$$\langle n_\mathbf{k}|\partial_i|n_\mathbf{k}\rangle = \frac{1}{\sqrt{N}}\partial_i\left[\frac{\mathcal{J}_n}{\sqrt{N}}\right]. \tag{6.32}$$

where we have defined

$$\mathcal{J}_n = \int_A \mathrm{d}^n\mathbf{x}' \int_A \mathrm{d}^n\mathbf{x} \int_S \mathrm{d}^n\mathbf{y} \, |\phi_n(\mathbf{y}_1, \ldots, \mathbf{y}_n)|^2 G^*(\mathbf{y}_i, \mathbf{x}'_i)G(\mathbf{y}_i, \mathbf{x}_i). \tag{6.33}$$

Now, using the fact that the state $|n_{\mathbf{k}}\rangle$ is normalised, we can write $\langle n_{\mathbf{k}}|n_{\mathbf{k}}\rangle \equiv 1 = \mathcal{J}_n/N$, and therefore

$$\langle n_{\mathbf{k}}|\partial_i|n_{\mathbf{k}}\rangle = \frac{1}{2}\partial_i\left[\ln(\mathcal{J}_n)\right] \tag{6.34}$$

The inner product $(\partial_i\langle n_{\mathbf{k}}|)|n_{\mathbf{k}}\rangle$ in Eq. (6.25) gives the same result[3]. If the Green's function is dependent on $\mathbf{k}$ only through the magnitude $|\mathbf{k}|$, then the inner product does not depend on the exact $\mathbf{k}$ mode we are considering [4]. Therefore, we can write

$$\begin{aligned}(\partial_i|n_\nu\rangle\langle n_\nu|) =&\partial_i|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|^{D_\nu}\\ =&\left(\partial_i D_\nu\right)|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|^{D_\nu}\ln(|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|)\\ &+D_\nu\left(\partial_i|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|\right)|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|^{D_\nu-1}\\ =&D_\nu\left(\partial_i|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|\right)|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|^{D_\nu-1},\end{aligned} \tag{6.35}$$

where we have used,

$$\begin{aligned}&(\partial_i D_\nu)|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|^{D_\nu}\ln(|n_{\mathbf{k}}\rangle\langle n_{\mathbf{k}}|) =\\ &\qquad\lim_{x\uparrow 1}\begin{pmatrix}x^{D_\nu}\ln(x) & 0 & \dots\\ 0 & (1-x)^{D_\nu}\ln(1-x) & \dots\\ \vdots & \vdots & \ddots\end{pmatrix}\\ &\qquad=0,\end{aligned} \tag{6.36}$$

The QFI then becomes

$$[\boldsymbol{\mathcal{I}}_Q]^{(\nu)}_{ij} = \sum_{n_\nu} p(n_\nu)\left[\partial_i\ln(p(n_\nu)\mathcal{J}_{n_\nu})\right]\left[\partial_j\ln(p(n_\nu)\mathcal{J}_{n_\nu})\right]. \tag{6.37}$$

We notice that the form of the QFI as given by Eq. (6.37) is manifestly symmetric in i, j. This is a well understood property if the quantum Fisher information matrix and has its roots in the fact that the QFI matrix is the metric in the probability space of the parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$. Expanding the logarithms of Eq. (6.37) we find

$$\begin{aligned}[\boldsymbol{\mathcal{I}}_Q]^{(\nu)}_{ij} = \sum_{n_\nu} p(n_\nu)\big\{&\left[\partial_i\ln(p(n_\nu))\right]\left[\partial_j\ln(p(n_\nu))\right] + \left[\partial_i\ln(\mathcal{J}_{n_\nu})\right]\left[\partial_j\ln(p(n_\nu))\right]\\ &+\left[\partial_i\ln(p(n_\nu))\right]\left[\partial_j\ln(\mathcal{J}_{n_\nu})\right] + \left[\partial_i\ln(\mathcal{J}_{n_\nu})\right]\left[\partial_j\ln(\mathcal{J}_{n_\nu})\right]\big\}.\end{aligned} \tag{6.38}$$

The first term in Eq. (6.38) has a clear physical interpretation, it is the classical Fisher information associated with the probability distribution $p(n_\nu)$ and as such describes the amount of information we can obtain by measuring the total number of photons (or equivalently the total intensity) across the whole detector. Physically we would expect that total photon number recorded by the detector would be proportional to the radiating area of the source, A_S. Therefore, a radiating source

[3] Technically the inner product $(\partial_i\langle n_{\mathbf{k}}|)|n_{\mathbf{k}}\rangle = (\langle n_{\mathbf{k}}|\partial_i|n_{\mathbf{k}}\rangle)^*$. However, since $\mathcal{J}_n$ is typically a positive, real quantity, we will ignore this technicality from now on, keeping in mind that we can use $\alpha + \alpha^* = 2\,\mathrm{Re}\,\alpha$ later if we are required.

[4] This approximation is valid for far field paraxial observation where the Green's function is approximately given by $(-ik/2\pi R)\mathrm{e}^{ik|\mathbf{y}-\mathbf{x}|}$ [Mandel and Wolf, 1995].

parameterised by more than one parameter, e.g. a rectangular source of sides a and b, should present us with an over-defined problem due to the infinity of solutions to the equation $A_S = ab$. Using the definition of $p(n_\nu)$, Eq. (6.16) and the chain rule we can write $\partial_i p(n_\nu) = \partial_{A_S} p(n_\nu) \partial_i A_S$ and we therefore find

$$\begin{aligned}\sum_{n_\nu} p(n_\nu) \left[\partial_i \ln(p(n_\nu))\right] \left[\partial_j \ln(p(n_\nu))\right] &= \sum_{n_\nu} \frac{1}{p(n_\nu)} \left[\partial_i p(n_\nu)\right] \left[\partial_j p(n_\nu)\right] \\ &= \sum_{n_\nu} \frac{1}{p(n_\nu)} [\partial_{A_S} p(n_\nu)]^2 [\partial_i A_S][\partial_j A_S]. \end{aligned} \tag{6.39}$$

The entire dependence on n_ν has now factored out of the derivatives which leads to a singular Fisher information matrix, which can be seen by writing the expression out for two parameters,

$$P \begin{pmatrix} \partial_1 A_S \partial_1 A_S & \partial_1 A_S \partial_2 A_S \\ \partial_2 A_S \partial_1 A_S & \partial_2 A_S \partial_2 A_S. \end{pmatrix} \tag{6.40}$$

where we have defined $P = \sum_{n_\nu} p(n_\nu)^{-1} [\partial_{A_S} p(n_\nu)]^2$. We see straight away that this matrix is singular since the determinant is zero. Another way of observing this result is that the matrix previously defined can be written as the outer product of the vector $(\partial_1 A_S, \ldots, \partial_r A_S)$ with itself. The rank of a matrix formed by taking an outer product of two vectors is always equal to 1. Since this matrix is then rank deficient, it must be singular. This establishes the fact that only a single parameter, namely the area of the source, can be estimated from measurements of the total photon number or intensity.

Now we consider the cross terms in Eq. (6.38). We can write

$$\begin{aligned}\sum_{n_\nu} p(n_\nu) \left[\partial_i \ln(\mathcal{J}_{n_\nu})\right] \left[\partial_j \ln(p(n_\nu))\right] &= \sum_{n_\nu} \left[\partial_i \ln(\mathcal{J}_{n_\nu})\right] \left[\partial_j p(n_\nu)\right] \\ &= \sum_{n_\nu} \left[\partial_{A_S} p(n_\nu)\right] \left[\partial_i \ln(\mathcal{J}_{n_\nu})\right] \left[\partial_j A_S\right], \end{aligned} \tag{6.41}$$

and we again find that the matrix defined by this term is rank 1 and singular as it can again be formed by a vector outer product. Similarly the matrix defined by the other cross term is also singular. The last term in Eq. (6.38) is the only term for which the n dependence does not generally factorise outside of the derivatives and therefore it is the only term that, taken on its own, does not lead to a singular QFI matrix.

It is instructive to examine in detail under what conditions the last term in Eq. (6.38) leads to a non-singular QFI matrix. We start our investigation by returning to Eq. (6.33), and in particular we examine the term $|\phi_n(\mathbf{y}_1, \ldots, \mathbf{y}_n)|^2$, which is the spatial wavefunction of the n photon state. If we treat the n photons as independent then we can write $|\phi_n(\mathbf{y}_1, \ldots, \mathbf{y}_n)|^2 = |\phi_1(\mathbf{y}_1)|^2 \ldots |\phi_1(\mathbf{y}_n)|^2$, which leads to

$$\mathcal{J}_n = \int_A \mathrm{d}^n \mathbf{x}' \int_A \mathrm{d}^n \mathbf{x} \int_S \mathrm{d}^n \mathbf{y} \prod_{i=1}^{n} |\phi_1(\mathbf{y}_i)|^2 G^*(\mathbf{y}_i, \mathbf{x}_i') G(\mathbf{y}_i, \mathbf{x}_i). \tag{6.42}$$

Since all n $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{y}$ variables are integrated over the same area (A, A and S respectively), we end up with n copies of the same integral, namely

$$\begin{aligned}\mathcal{J}_n &= \left[\int_A \mathrm{d}\mathbf{x}' \int_A \mathrm{d}\mathbf{x} \int_S \mathrm{d}\mathbf{y}\, |\phi_1(\mathbf{y})|^2 G^*(\mathbf{y},\mathbf{x}')G(\mathbf{y},\mathbf{x})\right]^n \\ &\equiv \mathcal{J}_1^n. \end{aligned} \tag{6.43}$$

This ultimately leads to

$$\sum_{n_\nu} p(n_\nu) n_\nu^2 \left[\partial_i \ln(\mathcal{J}_1)\right]\left[\partial_j \ln(\mathcal{J}_1)\right], \tag{6.44}$$

which again leads to a singular QFI matrix when taken on its own. Since the QFI matrix is then the sum of four singular, rank 1 matrices, the QFI matrix itself may be non-singular. Due to the sub-additivity of the matrix rank, we can be sure that the QFI will be singular when its dimension is greater than four. However, we can be more precise. Writing the QFI as

$$[\boldsymbol{\mathcal{I}}_Q]_{ij}^{(\nu)} = \sum_{n_\nu} p(n_\nu)\left[X_{n_\nu}^2 Y_i Y_j + n_\nu X_{n_\nu} Z_i Y_j + n_\nu X_{n_\nu} Y_i Z_j + n_\nu^2 Z_i Z_j\right], \tag{6.45}$$

where $X_{n_\nu} = \partial_{A_S} p(n_\nu)$, $Y_i = \partial_i A_S$ and $Z_i = \partial_i \ln(\mathcal{J}_1)$, we find that $\mathrm{rank}([\boldsymbol{\mathcal{I}}_Q]^{(\nu)}) = 2$. The maximum dimension of the QFI matrix, and therefore the number of parameters we can estimate, is two. Any attempt to estimate more than two parameters will lead to a rank two QFI matrix of dimension > 2, which means that the matrix will be rank deficient and therefore not invertible. Clearly this does not describe a general parameter estimation problem and the limit to the number of parameters suggests that when we treat the photons as statistically independent entities, we cannot perform the most general parameter estimation on the system. This reflects that truly independent photons will remain independent even after propagation, therefore measurements of the photon correlations cannot be used to determine the spatial properties of the source as in HBT measurements. We therefore conclude that the n-photon wavefunction $|\phi_n(\mathbf{y}_1,\ldots,\mathbf{y}_n)|^2$ cannot simply factorise meaning the photons cannot be independent. Physically we should expect that the wavefunction should not factorise since we integrate over all $\mathbf{y}$ variables and therefore when two or more $\mathbf{y}$ variables are close together, within the coherence area of the source, the photons will certainly be correlated. This indicates that the origin of the spatial information of the source in the far field is the result of correlations that exist in the photons at the source.

6.4 Spatially separated sources

In the previous section we witnessed that the factorisation of the multi-photon wavefunction $|\phi_n(\mathbf{y}_1,\ldots,\mathbf{y}_n)|^2$ leads to a singular QFI matrix. Factorisation of the wavefunction implies independence of the photons. However, it is known from the work of Oppel *et al.* [Oppel et al., 2012] that independent photons can be used to estimate the distribution of independent sources. These two results appear to be in disagreement. However, upon examining the problem more closely we find that no such disagreement exists. We take as an example two spatially separated sources S_1

and S_2. Under certain conditions the photons from each source are indistinguishable in the far field and the integral over the source becomes

$$\int_S \mathrm{d}\mathbf{y} \rightarrow \int_{S_1} \mathrm{d}\mathbf{y} + \int_{S_2} \mathrm{d}\mathbf{y}\,. \tag{6.46}$$

Looking at $\mathcal{J}_2$ we find

$$\begin{aligned}\mathcal{J}_2 = \int_A \mathrm{d}^2\mathbf{x} \int_A \mathrm{d}^2\mathbf{x}' \left[\int_{S_1} \mathrm{d}\mathbf{y}_1 + \int_{S_2} \mathrm{d}\mathbf{y}_1\right]\left[\int_{S_1} \mathrm{d}\mathbf{y}_2 + \int_{S_2} \mathrm{d}\mathbf{y}_2\right] \\ \times |\phi_2(\mathbf{y}_1,\mathbf{y}_2)|^2 G(\mathbf{y}_1,\mathbf{x}_1)G^*(\mathbf{y}_1,\mathbf{x}_1')G(\mathbf{y}_2,\mathbf{x}_2)G^*(\mathbf{y}_2,\mathbf{x}_2').\end{aligned} \tag{6.47}$$

If the two-photon wavefunction factorises as in the previous section, we see straight away that the $\mathcal{J}_2$ factorises as before and we end up with a singular QFI matrix. In their paper [Oppel et al., 2012] the authors also show that the n^{th}-order intensity correlation function can be used to measure the spatial distribution of independent, single photon emitting sources. Although up to now we have been considering sources emitting thermal radiation, since single photon emitting sources are also diagonal in the number basis, very little in the present analysis would change other than the probability distribution $p(n)$. We can therefore go straight ahead and calculate $\mathcal{J}_2$ for the single photon emitters. We note that since the sources both emit a single photon there is zero probability that they both originate from the same source. Also, since they are assumed to be statistically independent the two-photon wavefunction automatically factorises. Therefore, we find

$$\begin{aligned}\mathcal{J}_2 = \int_A \mathrm{d}^2\mathbf{x} \int_A \mathrm{d}^2\mathbf{x}' \left[\int_{S_1} \mathrm{d}\mathbf{y}_1 \int_{S_2} \mathrm{d}\mathbf{y}_2 + \int_{S_1} \mathrm{d}\mathbf{y}_2 \int_{S_2} \mathrm{d}\mathbf{y}_1\right] \\ \times |\phi_1(\mathbf{y}_1)|^2 |\phi_1(\mathbf{y}_2)|^2 G(\mathbf{y}_1,\mathbf{x}_1)G^*(\mathbf{y}_1,\mathbf{x}_1')G(\mathbf{y}_2,\mathbf{x}_2)G^*(\mathbf{y}_2,\mathbf{x}_2').\end{aligned} \tag{6.48}$$

Despite the factorisation, the expression (6.48) cannot be written in the same form as expression (6.47) and therefore does not lead to a singular QFI matrix. This corroborates the findings of Oppel *et al.* and also demonstrates an important point; independent photons can convey spatial information about the source except when they originate from the same spatial region. Looking at the result from the context of the original HBT experiments on Sirius, where the source was a single extended circular object, the origins of the spatial information obtained about the object is not from the interference of statistically independent photons but instead comes from the correlations between photons. Since the thermal light emitted from a star is spatially incoherent (to a good approximation [Blomstedt et al., 2013]) correlated photons must originate from the same point. Photons originating from the same point of the stars surface will be correlated assuming they are emitted within the coherence time of the source. As we showed in section 2.6 the mutual coherence function obeys the two wave equations (2.78) (2.79), it is known that this allows temporal correlations to transform into spatial correlations through propagation [Mandel and Wolf, 1995].

6.5 The spatial wavefunction $\phi_n(\mathbf{y}_1,\ldots,\mathbf{y}_n)$

We now examine the term $|\phi_n(\mathbf{y}_1,\ldots,\mathbf{y}_n)|^2$ in Eq. (6.33). We defined this term to be

$$|\phi_n(\mathbf{y}_1,\ldots,\mathbf{y}_n)|^2 = |\langle n|\mathbf{y}_1,\ldots,\mathbf{y}_n\rangle|^2 = \frac{1}{n!}|\langle n|\hat{a}^\dagger(\mathbf{y}_1)\ldots\hat{a}^\dagger(\mathbf{y}_n)|0\rangle|^2. \tag{6.49}$$

Since the inner product $\langle n|\hat{a}^\dagger(\mathbf{y}_1)\ldots\hat{a}^\dagger(\mathbf{y}_n)|\varphi\rangle$ is zero for all states $|\varphi\rangle \neq |0\rangle$ [Brainis, 2010], we can write

$$\begin{aligned}|\phi_n(\mathbf{y}_1, \ldots, \mathbf{y}_n)|^2 &= \frac{1}{n!}\langle n|\hat{a}^\dagger(\mathbf{y}_1)\ldots\hat{a}^\dagger(\mathbf{y}_n)\hat{a}(\mathbf{y}_n)\ldots\hat{a}(\mathbf{y}_1)|n\rangle \\ &\equiv \frac{1}{n!}\langle n|\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_n)|n\rangle \end{aligned} \tag{6.50}$$

where we have defined $\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_n) = \hat{a}^\dagger(\mathbf{y}_1)\ldots\hat{a}^\dagger(\mathbf{y}_n)\hat{a}(\mathbf{y}_n)\ldots\hat{a}(\mathbf{y}_1)$ for ease of notation. We now wish to find a closed form for the inner product $\langle n|\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_n)|n\rangle$. To this end we notice that the expectation value, $\langle\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_n)\rangle$, taken with respect to the state ρ is known since this is simply the n point intensity correlation for the thermal state. Therefore we find

$$\mathrm{Tr}[\rho\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_n)] = \sum_{m=0}^{\infty} p(m)\langle m|\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_n)|m\rangle. \tag{6.51}$$

Again making use of the spatial distribution of the m-photon state, we obtain

$$\begin{aligned}\langle m|\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_n)|m\rangle = &\int_S \mathrm{d}^m\mathbf{y}' \int_S \mathrm{d}^m\mathbf{y}''\, \phi_m^*(\mathbf{y}'_1, \ldots, \mathbf{y}'_m)\phi_m(\mathbf{y}''_1, \ldots, \mathbf{y}''_m) \\ &\times \langle \mathbf{y}'_1, \ldots, \mathbf{y}'_m|\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_n)|\mathbf{y}''_1, \ldots, \mathbf{y}''_m\rangle, \end{aligned} \tag{6.52}$$

which becomes

$$\langle m|\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1, \ldots, \mathbf{y}_n)|m\rangle = \begin{cases} 0 & \text{if } m < n \\ \binom{m}{n}^2|\phi_n(\mathbf{y}_1, \ldots, \mathbf{y}_n)|^2 & \text{if } m \geq n \end{cases}. \tag{6.53}$$

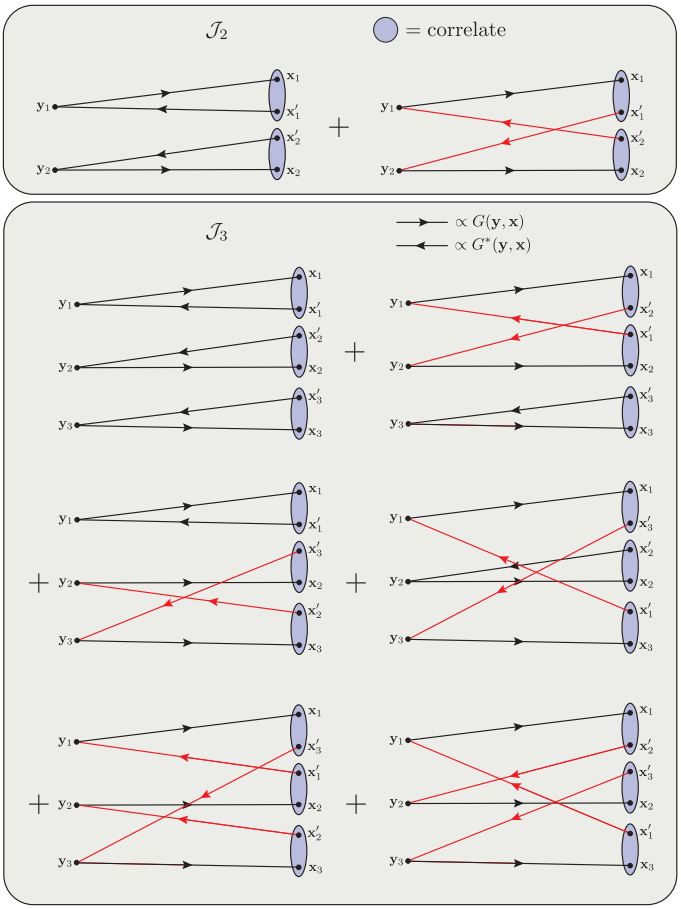

Figure 6.3: This figure gives a graphical representation of the terms involved in the expansion of $\mathcal{J}_2$ and $\mathcal{J}_3$. The first term in each expansion is simply the contribution from n independent photons propagating to the far field. Since the photon fields are only correlated to themselves, the first terms do not give rise to any interference in the far field. The photon fields $\psi(\mathbf{x}_1)$ ($\propto \int_S \mathrm{d}\mathbf{y}\, \phi(\mathbf{y})G(\mathbf{y},\mathbf{x}_1)$) and $\psi^*(\mathbf{x}'_1)$ will be completely coherent across the whole aperture A. When taken alone these terms would lead to singular QFI matrices and therefore cannot describe a general n-photon correlation function $\mathcal{J}_n$.

The n-point intensity correlation with respect to the state ρ can now be written as

$$\mathrm{Tr}[\rho\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1,\ldots,\mathbf{y}_n)] = \sum_{m\geq n} p(m)\binom{m}{n}^2 \langle n|\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1,\ldots,\mathbf{y}_n)|n\rangle, \tag{6.54}$$

and finally

$$|\phi_n(\mathbf{y}_1,\ldots,\mathbf{y}_n)|^2 = \frac{1}{n!}\frac{\mathrm{Tr}[\rho\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1,\ldots,\mathbf{y}_n)]}{\sum_{m\geq n} p(m)\binom{m}{n}^2}, \tag{6.55}$$

which gives the modulus squared of the n-photon wavefunction in terms of the n-point intensity correlation for the thermal state ρ.

The sum in Eq. (6.55) can be evaluated and gives

$$\sum_{m\geq n} p(m)\binom{m}{n}^2 = {}_2F_1\left(n+1, D_\nu+n; 1, \frac{\langle\hat{n}_\mathbf{k}\rangle}{1+\langle\hat{n}_\mathbf{k}\rangle}\right)p(n), \tag{6.56}$$

where ${}_2F_1(a,b;c,z)$ is the hypergeometric function. We can also evaluate the term $\mathrm{Tr}[\rho\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1,\ldots,\mathbf{y}_n)]$ since the n-point intensity correlation for the thermal state is known to be

$$\mathrm{Tr}[\rho\hat{\mathcal{G}}^{(n)}(\mathbf{y}_1,\ldots,\mathbf{y}_n)] = \sum_{\sigma\in S_n}\prod_{i=1}^{n}\langle\hat{a}^\dagger(\mathbf{y}_i)\hat{a}(\mathbf{y}_{\sigma(i)})\rangle. \tag{6.57}$$

Fig. 6.3 shows the contribution to the $\mathcal{J}_2$ and $\mathcal{J}_3$ from the $n!$ terms in the sum $\sum_{\sigma\in S_n}$. In principle these results could be used to fully calculate the QFI for the state ρ and determine via the quantum Cramér-Rao bound the ultimate precision in the spatial dimensions of the source. However the calculation of $\mathcal{J}_n$ for general n is far from simple and is beyond the scope of this thesis. The complications arising in the calculation of $\mathcal{J}_n$ are in determining the additional terms in the expansion of the sum over S_n (see Fig. 6.3), which are also the terms responsible for ensuring $\mathcal{J}_n$ does not factorise as in the previous section leading to a singular QFI matrix.

6.6 Alternative method

Recently there has been much interest in calculating the quantum Fisher information for a general n mode Gaussian state [Gao and Lee, 2014, Monras, 2013]. In both of these papers the authors give expressions for the symmetric logarithmic derivative and quantum Fisher information in terms of the first and second moments of a set of $2n$ operators. In his paper [Monras, 2013] Monras uses the set of $2n$ canonical operators $R^i = (\hat{q}_1,\ldots,\hat{q}_n,\hat{p}_1,\ldots,\hat{p}_n)$ and their first and second moments

$$d^i = \mathrm{Tr}[\rho R^i] \tag{6.58}$$

$$\Gamma^{ij} = \mathrm{Tr}[\rho(\tilde{R}^i\tilde{R}^j + \tilde{R}^j\tilde{R}^i)], \tag{6.59}$$

to express the quantum Fisher information of n bosonic Gaussian modes, where $\tilde{R}^i = R^i - d^i$ are the centred operators. This provides a phase space formalism for the calculation of the quantum Fisher information in terms of the first and second moments of the states distribution in phase space.

In their paper [Gao and Lee, 2014], Gao and Lee use instead the set of $2n$ operators given by the creation and annihilation operators of the n modes. They thus define the $2n$-tuple $a^\mu = (\hat{a}_1,\hat{a}_1^\dagger,\ldots,\hat{a}_n,\hat{a}_n^\dagger)^{\mathrm{T}}$ and their first and second moments

$$\lambda^\mu = \mathrm{Tr}[\rho a^\mu] \tag{6.60}$$

$$\Sigma^{\mu\nu} = \mathrm{Tr}[\rho(\tilde{a}^\mu\tilde{a}^\nu + \tilde{a}^\nu\tilde{a}^\mu)], \tag{6.61}$$

where again the centred operators are defined as $\tilde{a}^\mu = a^\mu - \lambda^\mu$. They show that the quantum Fisher information for a $2n$-dimensional Gaussian state is given by

$$[\boldsymbol{\mathcal{I}}_Q]_{ij} = \frac{1}{2}\mathfrak{M}^{-1}_{\alpha\beta,\mu\nu}\partial_j\Sigma^{\alpha\beta}\partial_i\Sigma^{\mu\nu} + \Sigma^{-1}_{\mu\nu}\partial_j\lambda^\mu\partial_i\lambda^\nu \tag{6.62}$$

where $\mathfrak{M} = \Sigma\otimes\Sigma + \frac{1}{4}\Omega\otimes\Omega$ and Ω is the symplectic matrix constructed from the Pauli y matrix on each mode $\Omega = \oplus_{j=1}^{n} i\sigma_y^{(j)}$. This allows us calculate the quantum Fisher information for any n mode bosonic Gaussian system provided we can take the derivatives of the first and second moments.

We hope to apply these results to the system considered above however it will require careful consideration as to how we construct our modes. We believe that the n modes can be constructed as spatially localised modes but this involves partitioning the measurement area A into pixels of area $\Delta A = A/n$ such that we can associate with each partitioning a localised mode. To truly find the quantum Fisher information in this setting it will be necessary to take the limit $\Delta A \to 0$ so that the spatial degrees of freedom for the state become continuous. This must be done with care since it is known that as the localisation volumes associated with each mode approach the order of the wavelength the commutation of the mode operators must be accounted for [Mandel and Wolf, 1995].

6.7 Discussion and conclusions

In this chapter we have made use of the quantum description of the thermal radiation field to examine how information regarding the spatial distribution of a thermal source is conveyed to the far field. We found that the statistical independence of photons in the source plane is detrimental in the retrieval of spatial information from thermal sources and actually leads to a singular quantum Fisher information matrix when we try to estimate more than two parameters. We found that the quantum Fisher information is purely determined by two fundamental quantities. Firstly, the probability distribution $p(n_\nu)$ that gives the probability of finding n_ν photons of frequency ν in the detection area. And secondly, $\mathcal{J}_{n_\nu}$, which is related to the distribution of the n_ν photons in space. The first term corresponds to the information present in a measurement of the total photon number across the detection area and as such can only be used to estimate the total area of the source. Any attempts to determine more than a single parameter from such a measurement will evidently fail due to the singular nature of the resulting QFI matrix. We therefore conclude that the second term $\mathcal{J}_{n_\nu}$, is where the majority of the information resides. Treating the photons in the source plane as independent leads to the following relation $\mathcal{J}_{n_\nu} = \mathcal{J}_1^{n_\nu}$ and inevitably to a rank two QFI matrix. Therefore, we find that only two parameters can be estimated. The term $\mathcal{J}_1$ is proportional to the integral over the far field complex degree of coherence for the source.

Chapter 7

Summary and Outlook

In this thesis I have applied the theory of parameter estimation to various classical and quantum imaging systems and found a number of interesting results regarding the precision such systems can achieve. In particular, I closely examined the intensity correlation measurements as first performed by Hanbury Brown and Twiss.

In chapter 4 I made the identification of an image with an un-normalised probability distribution to derive resolution limits for various systems including the double slit experiment, classical and quantum lithography. The results of this chapter reproduce a number of well established resolution limits. I also demonstrated how general detector imperfections can be incorporated into the theory, making use of POVMs, and calculated the effect for a few select examples. I then used our method to determine the resolution limits of multi-photon correlation measurements as performed by Oppel *et al.* Our method allows for a relatively simple calculation of imaging resolution and can be applied to a variety of cases not considered here.

In chapter 5 I analyse higher-order intensity correlation measurements in great detail in order to exactly quantify the enhanced resolution that can be achieved through these measurements. In order to do so I first determined the probability distribution functions (PDFs) of data that comes from measurements of this kind. I found that the data can be considered to be normally distributed and explicitly calculate the moments of this distribution. This allows us to determine the statistical performance of estimators that take as input intensity correlation data. I was therefore able to compare the performance of estimators for various correlation orders and found that often the best estimates occur for third order intensity correlations. Full characterisation of the PDF also allowed us to perform a maximum likelihood estimation procedure. I also presented the results of simulations that were performed, simulating the output of intensity correlation measurements and using the simulated data to estimate the input parameters. I found that the results agree very well with the predictions of the theory. I also showed how the estimator performance can be optimised by adjusting the arrangement of the experiment and from this I inferred the conditions which should be sought in an experiment to increase the precision of the estimates obtained. The use of maximum likelihood estimators ensures that the estimates achieve the Cramér-Rao bound. Therefore, the precision quoted in this chapter provides a lower bound on the precision that cannot be beaten by an unbiased estimator. The results of this chapter are quite general and although I use the example of thermal radiation, the only necessity is that the state be a state of uncertain photon number. The results therefore generalise quite simply and can

be used for numerous parameter estimation tasks. By including noise in the model the results should be reproducible with experimental data although this remains to be demonstrated. I anticipate that certain complications may arise in attempting to perform such an experiment. Firstly, the additional parameters that must be estimated alongside the source parameters, are assumed to be constant across the pixel array. These parameters include the expectation of the intensity and any noise parameters. If this is not the case then the model will not be accurate and the estimation procedure is likely to fail. In principle this could be dealt with by including more parameters. However, it is important to avoid this if at all possible to reduce the size of the parameter estimation problem as each new parameter will generally decrease the performance of the procedure. Another foreseen obstacle is the stationarity of the light source. Again since this is assumed by the model it is a prerequisite for the procedure to work.

The experiments of Oppel *et al.* make use of so called magic angles to acquire a pure sinusoidal modulation of the intensity correlation functions in the far field of a regular linear array of thermal sources. Using the results of chapter 5 I was able to demonstrate that the magic angles also minimise the variance in estimates of the inter source spacing d. This is an interesting connection and I would like to explore this further. In particular it would be of great interest to know which of the two is the more fundamental principle, i.e. does maximising the Fisher information imply a sinusoidal modulation or is the relationship the other way round. Also, it is not currently known why the third order intensity correlation outperforms the other intensity correlations in these measurements or indeed if the third order correlation is always the best regardless of the geometry. It has been shown that the signal to noise ratio decreases with the correlation order and yet the visibility increases. Perhaps the optimality of the third order correlation lies in resolving these two competing facets but this remains to be shown.

Chapter 6 uses the quantum mechanical description of the thermal state and the quantum Fisher information to determine how information about the spatial distribution of a thermal source is propagated to the far field. I establish a connection between correlations that exist between photons in the source plane and the ability to perform a general parameter estimation procedure in the far field. I also found that, in agreement with our intuition, measurements of the total intensity in the far field can only reveal information about the total area of the radiating source and cannot give general information about the source distribution. The statistical independence of the photons in the source plane does not necessarily inhibit the acquisition of information in the far field. When the photons come from spatially separated sources and the probability of both photons coming from the same source is zero (e.g. when both photons originate from an individual atom) I again found that the general parameter estimation task is well defined and can proceed without difficulty. The analysis uncovers two key quantities in the retrieval of information in the far field. First the probability distribution of the total number of photons in the detection area, $p(n_\nu)$, and second the term, $\mathcal{J}_{n_\nu}$, which contains information about how the photons are spatially distributed. Together, these two terms contain all the information in the state and therefore the quantum Fisher information is purely given in terms of them. It remains an open question exactly what parameter the correlation term $\mathcal{J}_n$ adds to the quantum Fisher information matrix, taking the rank from one to two and I aim to clarify this in future work.

Appendix A

Quasi-Probability Distributions in Quantum Optics

In quantum optics the use of quasi-probability distributions to calculate quantum expectation values can prove an indispensable tool. Together with the optical equivalence theorem, quasi-probability distributions provide a method of calculating quantum mechanical expectation values in much the same way as expectation values are calculated in classical optics. Sudarshan was the first to point out that by making use of the over-completeness of the coherent states, it is possible to express the density operator ρ in the "diagonal" form

$$\rho = \int \mathrm{d}^2\alpha \, P_\rho(\alpha) |\alpha\rangle \langle\alpha|, \tag{A.1}$$

where the integration takes place over the entire complex plane [Sudarshan, 1963]. As was shown in chapter 2, by writing the density operator in this form the expectation value of normally ordered operators reduces to a classical-like expectation integral with $P_\rho(\alpha)$ playing the role of the probability distribution. The P-representation is always real valued and integrates to 1, however for some states the P-representation can be more singular than a delta function and can even take negative values in regions of phase space [Mandel and Wolf, 1995]. It is for this reason that the P-representation cannot be considered a true probability distribution.

Similarly, for anti-normally ordered operators we can define the Husimi Q-representation. Consider an arbitrary, anti-normally ordered operator, which we can write as a function of the operators $\hat{a}$ and $\hat{a}^\dagger$

$$g(\hat{a}, \hat{a}^\dagger) = \sum_{n,m} b_{nm} \hat{a}^n \hat{a}^{\dagger m}. \tag{A.2}$$

We find that the expectation of any such operator is given by

$$\begin{aligned}\langle g(\hat{a},\hat{a}^\dagger)\rangle &= \left\langle \sum_{n,m} b_{nm}\hat{a}^n\hat{a}^{\dagger m}\right\rangle \\ &= \sum_{n,m} b_{nm}\mathrm{Tr}[\rho\hat{a}^n\hat{a}^{\dagger m}] \\ &= \sum_{n,m}\frac{b_{nm}}{\pi}\int \mathrm{Tr}\left[\rho\hat{a}^n|\alpha\rangle\langle\alpha|\hat{a}^{\dagger m}\right]\mathrm{d}^2\alpha \\ &= \sum_{n,m}\frac{b_{nm}}{\pi}\int \alpha^n\alpha^{*m}\mathrm{Tr}[\rho|\alpha\rangle\langle\alpha|]\,\mathrm{d}^2\alpha \\ &= \frac{1}{\pi}\int g(\alpha,\alpha^*)\langle\alpha|\rho|\alpha\rangle\,\mathrm{d}^2\alpha\,. \end{aligned} \tag{A.3}$$

Again we see that the expectation of a quantum mechanical operator has been reduced to a classical-like expectation integral, this time with the weighting function being $\frac{1}{\pi}\langle\alpha|\rho|\alpha\rangle$, which we define as the Husimi Q-representation, $Q_\rho(\alpha)$.

Finally we state without proof that the expectation of symmetric or Weyl ordered operators can be evaluated with the use of the Wigner function. Symmetric ordering performs the following action [Fujii and Suzuki, 2003]

$$\hat{a}^{\dagger n}\hat{a}^m = \binom{n+m}{n}^{-1} \times (\text{sum of all symmetric products of } n\hat{a}^\dagger \text{ and } m\hat{a}), \tag{A.4}$$

and the expectation value of any symmetrically ordered function of the creation and annihilation operators, $h(\hat{a},\hat{a}^\dagger)$ is [Cahill and Glauber, 1969]

$$\langle h(\hat{a},\hat{a}^\dagger)\rangle = \frac{1}{\pi}\int h(\alpha,\alpha^*)W_\rho(\alpha)\,\mathrm{d}^2\alpha\,. \tag{A.5}$$

The Wigner function is calculated from the state ρ by

$$W_\rho(\alpha) = 2\mathrm{Tr}[\rho D(2\alpha)\exp(i\pi\hat{a}^\dagger\hat{a})], \tag{A.6}$$

where $D(\alpha)\equiv\exp(\alpha\hat{a}^\dagger-\alpha^*\hat{a})$ is the displacement operator.

Appendix B

Moments of the Noise Distribution

Here we evaluate the moments of the noise distribution for the general case and also for the case of a Gaussian noise distribution as considered in chapter 5. First, we need to evaluate the term $\langle \eta(x_i)\eta(s_2)\dots\eta(s_n)\rangle$ which appears in Eq. (5.19). In chapter 5 we calculated this term for the Gaussian noise distribution, here we calculate it for the general case. Denoting $J_1 = \langle\eta(s_2)\rangle\dots\langle\eta(s_n)\rangle$ and $S = \{s_2,\dots,s_n\}$ we find

$$\langle \eta(x_i)\eta(s_2)\dots\eta(s_n)\rangle = \begin{cases} \langle\eta(x_i)\rangle J_1 & \text{if } x_i \notin S \\ \frac{\langle\eta(x_i)^2\rangle}{\langle\eta(x_i)\rangle} J_1 & \text{if } x_i \in S \end{cases}, \tag{B.1}$$

if none of the reference pixels are equal, i.e., $s_2 \neq s_3 \neq \dots \neq s_n$. If instead we use detection scheme 1, where all the reference pixels are the same, we find

$$\langle \eta(x_i)\eta(s_2)^{n-1}\rangle = \begin{cases} \langle\eta(x_i)\rangle J_2 & \text{if } x_i \neq s_2 \\ \frac{\langle\eta(x_i)^n\rangle}{J_2} J_2 & \text{if } x_i = s_2 \end{cases}, \tag{B.2}$$

where $J_2 = \langle\eta(s_2)^{n-1}\rangle$.

Now we evaluate the second term in Eq. (5.21), $\langle \eta(x_i)\eta(x_j)\eta(s_2)^2\dots\eta(s_n)^2\rangle$. Since the noise is treated as uncorrelated between the pixels, the expectation value factorises into $\langle\eta(x_i)\rangle\langle\eta(x_j)\rangle\langle\eta(s_2)^2\rangle\dots\langle\eta(s_n)^2\rangle$ if x_i and x_j are not equal to each other or any of the positions $s_2,\dots,s_n$. However more generally the expression is more complicated. Denoting $J_3 = \langle\eta(s_2)^2\rangle\dots\langle\eta(s_n)^2\rangle$ we obtain

$$\langle \eta(x_i)\eta(x_j)\eta(s_2)^2\dots\eta(s_n)^2\rangle = \begin{cases} \frac{\langle\eta(x_i)^4\rangle}{\langle\eta(x_i)^2\rangle} J_3 = a & \text{if } x_i = x_j \in S \\ \langle\eta(x_i)^2\rangle J_3 = b & \text{if } x_i = x_j \notin S \\ \frac{\langle\eta(x_i)^3\rangle}{\langle\eta(x_i)^2\rangle}\frac{\langle\eta(x_j)^3\rangle}{\langle\eta(x_j)^2\rangle} J_3 = c & \text{if } x_i \neq x_j;\ x_i, x_j \in S \\ \langle\eta(x_i)\rangle\langle\eta(x_j)\rangle J_3 = d & \text{if } x_i \neq x_j;\ x_i, x_j \notin S \\ \frac{\langle\eta(x_i)^3\rangle}{\langle\eta(x_i)^2\rangle}\langle\eta(x_j)\rangle J_3 = e & \text{if } x_i \neq x_j;\ x_i \in S;\ x_j \notin S \\ \langle\eta(x_i)\rangle\frac{\langle\eta(x_j)^3\rangle}{\langle\eta(x_j)^2\rangle} J_3 = f & \text{if } x_i \neq x_j;\ x_i \notin S;\ x_j \in S \end{cases}, \tag{B.3}$$

and for detection scheme 1

$$\langle \eta(x_i)\eta(x_j)\eta(s_2)^{2n-2}\rangle = \begin{cases} \frac{\langle \eta(x_i)^4\rangle}{J_4} J_4 & \text{if } x_i = x_j = s_2 \\ \langle \eta(x_i)^2\rangle J_4 & \text{if } x_i = x_j \neq s_2 \\ \langle \eta(x_i)\rangle\langle \eta(x_j)\rangle J_4 & \text{if } x_i \neq x_j;\ x_i \neq s_2;\ x_j \neq s_2\ , \\ \frac{\langle \eta(x_i)\rangle}{J_4}\langle \eta(x_j)^{2n-1}\rangle J_4 & \text{if } x_i \neq x_j = s_2 \\ \frac{\langle \eta(x_j)\rangle}{J_4}\langle \eta(x_i)^{2n-1}\rangle J_4 & \text{if } x_j \neq x_i = s_2 \end{cases} \tag{B.4}$$

where $J_4 = \langle \eta(s_2)^{2n-2}\rangle$. For a general noise distribution, each term in these piecewise functions can be associated with a parameter to be estimated. The benefit of using a Gaussian noise model is that there are only two additional noise parameters corresponding to the first and second moments of the Gaussian distribution, or a combination of them as in chapter 5, (ν, χ).

To give a more visual understanding we can represent the resulting piecewise function, Eq. (B.3) in matrix form as $\mathbf{M}_{ij} = \langle \eta(x_i)\eta(x_j)\eta(s_2)^2 \dots \eta(s_n)^2\rangle$

$$\mathbf{M} = \begin{pmatrix} b & d & d & e & d & d & d & e & \dots \\ d & b & d & e & d & d & d & e & \dots \\ d & d & b & e & d & d & d & e & \dots \\ e & e & e & a & e & e & e & c & \dots \\ d & d & d & e & b & d & d & e & \dots \\ d & d & d & e & d & b & d & e & \dots \\ d & d & d & e & d & d & b & e & \dots \\ e & e & e & c & e & e & e & a & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{B.5}$$

where we have made use of the fact that $f = e$ since the noise is assumed to be the same for all pixels.

Bibliography

[Aaronson and Arkhipov, 2010] Aaronson, S. and Arkhipov, A. (2010). The Computational Complexity of Linear Optics. page 94, 1011.3245.

[Aaronson and Arkhipov, 2013] Aaronson, S. and Arkhipov, A. (2013). BosonSampling Is Far From Uniform. page 41, 1309.7460.

[Abbe, 1873] Abbe, E. (1873). Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Arch. für Mikroskopische Anat.*, 9(1):413–418.

[Agafonov et al., 2008a] Agafonov, I., Chekhova, M., Iskhakov, T., and Penin, A. (2008a). High-visibility multiphoton interference of Hanbury Brown–Twiss type for classical light. *Phys. Rev. A*, 77(5):053801.

[Agafonov et al., 2008b] Agafonov, I. N., Chekhova, M. V., Iskhakov, T. S., and Penin, A. N. (2008b). High-visibility multiphoton interference of hanbury brown-twiss type for classical light. *Phys. Rev. A*, 77:053801.

[Agafonov et al., 2009a] Agafonov, I. N., Chekhova, M. V., Iskhakov, T. S., and Wu, L.-A. (2009a). High-visibility intensity interference and ghost imaging with pseudo-thermal light. *J. Mod. Opt.*, 56(2-3):422–431.

[Agafonov et al., 2009b] Agafonov, I. N., Chekhova, M. V., and Penin, A. N. (2009b). Multiphoton ghost imaging with classical light. 0911.3718.

[Barnett and Radmore, 1997] Barnett, S. M. and Radmore, P. M. (1997). *Methods in Theoretical Quantum Optics*. Oxford University Press.

[Bennink et al., 2002] Bennink, R., Bentley, S., and Boyd, R. (2002). "Two-Photon" Coincidence Imaging with a Classical Source. *Phys. Rev. Lett.*, 89(11).

[Beran and Parrent, 1964] Beran, M. J. and Parrent, G. B. (1964). *Theory of Partial Coherence*. Prentice-Hall.

[Blomstedt et al., 2013] Blomstedt, K., Setälä, T., Tervo, J., Turunen, J., and Friberg, A. T. (2013). Partial polarization and electromagnetic spatial coherence of blackbody radiation emanating from an aperture. *Phys. Rev. A*, 88(1):013824.

[Born and Wolf, 1980] Born, M. and Wolf, E. (1980). *Principles of Optics*. Cambridge University Press, 6th edition.

[Boto et al., 2000] Boto, A., Kok, P., Abrams, D., Braunstein, S., Williams, C., and Dowling, J. (2000). Quantum Interferometric Optical Lithography: Exploiting Entanglement to Beat the Diffraction Limit. *Phys. Rev. Lett.*, 85(13):2733–2736, 9912052.

[Brainis, 2010] Brainis, E. (2010). Huygens-Fresnel principle for N-photon states of light. page 5, 1011.1402.

[Brannen and Ferguson, 1956] Brannen, E. and Ferguson, H. I. S. (1956). The Question of Correlation between Photons in Coherent Light Rays. *Nature*, 178(4531):481–482.

[Braunstein and Caves, 1994] Braunstein, S. and Caves, C. (1994). Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72(22):3439–3443.

[Braunstein et al., 1996] Braunstein, S. L., Caves, C. M., and Milburn, G. (1996). Generalized Uncertainty Relations: Theory, Examples, and Lorentz Invariance. *Ann. Phys. (N. Y).*, 247(1):135–173.

[Brown and Twiss, 1956a] Brown, R. H. and Twiss, R. Q. (1956a). A test of a new type of stellar interferometer on Sirius. *Nature*, 178:1046–1048.

[Brown and Twiss, 1956b] Brown, R. H. and Twiss, R. Q. (1956b). Correlation between Photons in two Coherent Beams of Light. *Nature*, 177(4497):27–29.

[Brown and Twiss, 1957] Brown, R. H. and Twiss, R. Q. (1957). Interferometry of the Intensity Fluctuations in Light. I. Basic Theory: The Correlation between Photons in Coherent Beams of Radiation. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 242(1230):300–324.

[Brown and Twiss, 1958a] Brown, R. H. and Twiss, R. Q. (1958a). Interferometry of the Intensity Fluctuations in Light II. An Experimental Test of the Theory for Partially Coherent Light. *Proc. R. Soc. Lond. A. Math. Phys. Sci.*, 243(1234):291–319.

[Brown and Twiss, 1958b] Brown, R. H. and Twiss, R. Q. (1958b). Interferometry of the Intensity Fluctuations in Light. III. Applications to Astronomy. *Proc. R. Soc. Lond. A. Math. Phys. Sci.*, 248(1253):199–221.

[Brown and Twiss, 1958c] Brown, R. H. and Twiss, R. Q. (1958c). Interferometry of the Intensity Fluctuations in Light. IV. A Test of an Intensity Interferometer on Sirius A. *Proc. R. Soc. Lond. A. Math. Phys. Sci.*, 248(1253):222–237.

[Cahill and Glauber, 1969] Cahill, K. E. and Glauber, R. J. (1969). Density operators and quasiprobability distributions. *Phys. Rev.*, 177(5):1882–1902.

[Cao et al., 2008] Cao, D., Xiong, J., and Zhang, S. (2008). Enhancing visibility and resolution in¡ equation¿ N¡ font face='verdana'¿ th¡/font¿¡/equation¿-order intensity correlation of thermal light. *Appl. Phys. ....*

[Chan et al., 2009] Chan, K. W. C., O'Sullivan, M. N., and Boyd, R. W. (2009). High-order thermal ghost imaging. *Opt. Lett.*, 34(21):3343–5.

[Chen et al., 2010] Chen, X.-H., Agafonov, I. N., Luo, K.-H., Liu, Q., Xian, R., Chekhova, M. V., and Wu, L.-A. (2010). High-visibility, high-order lensless ghost imaging with thermal light. *Opt. Lett.*, 35(8):1166–8.

[Dirac, 1989] Dirac, P. A. M. (1989). *The Principles of Quantum Mechanics*. Oxford University Press, 4th edition.

[Fano, 1957] Fano, U. (1957). Description of States in Quantum Mechanics by Density Matrix and Operator Techniques. *Rev. Mod. Phys.*, 29(1):74–93.

[Fano, 1961] Fano, U. (1961). Quantum Theory of Interference Effects in the Mixing of Light from Phase-Independent Sources. *Am. J. Phys.*, 29(8):539.

[Ferri et al., 2005] Ferri, F., Magatti, D., Gatti, A., Bache, M., Brambilla, E., and Lugiato, L. A. (2005). High-Resolution Ghost Image and Ghost Diffraction Experiments with Thermal Light. *Phys. Rev. Lett.*, 94(18):183602.

[Ferrie, 2014] Ferrie, C. (2014). Data-processing inequalities for quantum metrology. *Phys. Rev. A*, 90(1):014101.

[Fujii and Suzuki, 2003] Fujii, K. and Suzuki, T. (2003). A New Symmetric Expression of Weyl Ordering. page 15, 0304094.

[Gabor, 1961] Gabor, D. (1961). IV Light and Information. *Prog. Opt.*, 1:109–153.

[Gao and Lee, 2014] Gao, Y. and Lee, H. (2014). Bounds on quantum multiple-parameter estimation with Gaussian state. *Eur. Phys. J. D*, 68(11):347.

[Gard et al., 2014a] Gard, B. T., Motes, K. R., Olson, J. P., Rohde, P. P., and Dowling, J. P. (2014a). An introduction to boson-sampling. 1406.6767.

[Gard et al., 2014b] Gard, B. T., Olson, J. P., Cross, R. M., Kim, M. B., Lee, H., and Dowling, J. P. (2014b). Inefficiency of classically simulating linear optical quantum computing with Fock-state inputs. *Phys. Rev. A*, 89(2):022328.

[Gatti et al., 2006] Gatti, A., Bache, M., Magatti, D., Brambilla, E., Ferri, F., and Lugiato, L. A. (2006). Coherent imaging with pseudo-thermal incoherent light. *J. Mod. Opt.*, 53(5-6):739–760.

[Gatti et al., 2007] Gatti, A., Brambilla, E., Bache, M., and Lugiato, L. A. (2007). Ghost imaging. In Kolobov, M. I., editor, *Quantum Imaging*, chapter 5. Springer.

[Gerry and Knight, 2006] Gerry, C. C. and Knight, P. L. (2006). *Introductory Quantum Optics*. Cambridge University Press.

[Glauber, 1963a] Glauber, R. (1963a). Coherent and Incoherent States of the Radiation Field. *Phys. Rev.*, 131(6):2766–2788.

[Glauber, 1963b] Glauber, R. (1963b). Photon Correlations. *Phys. Rev. Lett.*, 10(3):84–86.

[Glauber, 1963c] Glauber, R. (1963c). The Quantum Theory of Optical Coherence. *Phys. Rev.*, 130(6):2529–2539.

[Glauber, 2007] Glauber, R. J. (2007). *Quantum Theory of Optical Coherence*. Wiley.

[Gnedenko, 1962] Gnedenko, B. V. (1962). *The Theory of Probability*. Chelsea Publishing.

[Gogolin et al., 2013] Gogolin, C., Kliesch, M., Aolita, L., and Eisert, J. (2013). Boson-Sampling in the light of sample complexity. page 22, 1306.3995.

[Hecht and Zajac, 1980] Hecht, E. and Zajac, A. (1980). *Optics*. Addison-Wesley.

[Hell and Wichmann, 1994] Hell, S. W. and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.*, 19(11):780.

[Heller et al., 2013] Heller, I., Sitters, G., Broekmans, O. D., Farge, G., Menges, C., Wende, W., Hell, S. W., Peterman, E. J. G., and Wuite, G. J. L. (2013). STED nanoscopy combined with optical tweezers reveals protein dynamics on densely covered DNA. *Nat. Methods*, 10(9):910–6.

[Helstrom, 1976] Helstrom, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press.

[Hess et al., 2006] Hess, S. T., Girirajan, T. P. K., and Mason, M. D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.*, 91(11):4258–72.

[Holevo, 2011] Holevo, A. (2011). *Probabilistic and Statistical Aspects of Quantum Theory*. Edizioni Della Normale.

[Huh et al., 2015] Huh, J., Guerreschi, G. G., Peropadre, B., McClean, J. R., and Aspuru-Guzik, A. (2015). Boson sampling for molecular vibronic spectra. *Nat. Photonics*, 9(9):615–620.

[Iskhakov et al., 2011] Iskhakov, T., Allevi, A., Kalashnikov, D. A., Sala, V. G., Takeuchi, M., Bondani, M., and Chekhova, M. (2011). Intensity correlations of thermal light. *Eur. Phys. J. Spec. Top.*, 199(1):127–138.

[Jackson, 1999] Jackson, J. D. (1999). *Classical Electrodynamics*. John Wiley and Sons, 3rd edition.

[Jordan, 1964] Jordan, T. (1964). Operators for observables in quantum optics. *Phys. Lett.*, 11(4):289–291.

[Kay, 1993] Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall.

[Klauder and Sudarshan, 1968] Klauder, J. R. and Sudarshan, E. C. G. (1968). *Fundamentals of Quantum Optics*. W. A. Benjamin.

[Klyshko, 1988] Klyshko, D. (1988). Combine EPR and two-slit experiments: Interference of advanced waves. *Phys. Lett. A*, 132(6-7):299–304.

[Kok and Lovett, 2010] Kok, P. and Lovett, B. W. (2010). *Introduction to Optical Quantum Information Processing*. Cambridge University Press.

[Levitin and Toffoli, 2014] Levitin, L. B. and Toffoli, T. (2014). The Capacity of a Channel with an Image as the Information Source. *J. Quantum Inf. Sci.*, 04(02):111–116.

[Lipson et al., 2011] Lipson, A., Lipson, S. G., and Lipson, H. (2011). *Optical Physics*. Cambridge University Press, 4th edition.

[Liu and Shih, 2009] Liu, J. and Shih, Y. (2009). Nth -order coherence of thermal light. *Phys. Rev. A - At. Mol. Opt. Phys.*, 79(2):1–15.

[Liu et al., 2009] Liu, Q., Chen, X.-H., Luo, K.-H., Wu, W., and Wu, L.-A. (2009). Role of multiphoton bunching in high-order ghost imaging with thermal light sources. *Phys. Rev. A*, 79(5):053844.

[Macintosh et al., 2014] Macintosh, B., Graham, J. R., Ingraham, P., Konopacky, Q., Marois, C., Perrin, M., Poyneer, L., Bauman, B., Barman, T., Burrows, a. S., Cardwell, a., Chilcote, J., De Rosa, R. J., Dillon, D., Doyon, R., Dunn, J., Erikson, D., Fitzgerald, M. P., Gavel, D., Goodsell, S., Hartung, M., Hibon, P., Kalas, P., Larkin, J., Maire, J., Marchis, F., Marley, M. S., McBride, J., Millar-Blanchaer, M., Morzinski, K., Norton, a., Oppenheimer, B. R., Palmer, D., Patience, J., Pueyo, L., Rantakyro, F., Sadakuni, N., Saddlemyer, L., Savransky, D., Serio, a., Soummer, R., Sivaramakrishnan, a., Song, I., Thomas, S., Wallace, J. K., Wiktorowicz, S., and Wolff, S. (2014). First light of the Gemini Planet Imager. *Proc. Natl. Acad. Sci.*, 111(35):12661–12666.

[Mandel and Wolf, 1995] Mandel and Wolf (1995). *Optical Coherence and Quantum Optics*. Cambridge University Press.

[Mandel, 1959] Mandel, L. (1959). Fluctuations of Photon Beams: The Distribution of the Photo-Electrons. *Proc. Phys. Soc.*, 74(3):233–243.

[Minc, 1978] Minc, H. (1978). *Permanents*. Addison-Wesley.

[Monras, 2013] Monras, A. (2013). Phase space formalism for quantum estimation of Gaussian states. 1303.3682.

[Monras and Illuminati, 2011] Monras, A. and Illuminati, F. (2011). Measurement of damping and temperature: Precision bounds in Gaussian dissipative channels. *Phys. Rev. A - At. Mol. Opt. Phys.*, 83(1):1–12, 1010.0442.

[Motes et al., 2013] Motes, K. R., Dowling, J. P., and Rohde, P. P. (2013). Spontaneous parametric down-conversion photon sources are scalable in the asymptotic limit for boson sampling. *Phys. Rev. A*, 88(6):063822.

[Nielsen and Chuang, 2000] Nielsen, M. A. and Chuang, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge University Press.

[Oppel, 2012] Oppel, S. (2012). *Multi-Photon Interferences of Independent Light Sources*. PhD thesis, Der Naturwissenschaftlichen Fakultät der Friedrich-Alexander-Universität Erlangen-Nürnberg.

[Oppel et al., 2012] Oppel, S., Büttner, T., Kok, P., and von Zanthier, J. (2012). Superresolving Multiphoton Interferences with Independent Light Sources. *Phys. Rev. Lett.*, 109(23):233603.

[Paris, 2009] Paris, M. (2009). Quantum estimation for quantum technology. *Int. J. Quantum Inf.*

[Pearce et al., 2015] Pearce, M. E., Mehringer, T., von Zanthier, J., and Kok, P. (2015). Precision estimation of source dimensions from intensity correlation functions. In preparation.

[Pendry, 2000] Pendry, J. B. (2000). Negative Refraction Makes a Perfect Lens. *Phys. Rev. Lett.*, 85(18):3966–3969.

[Perelomov, 1986] Perelomov, A. (1986). *Generalized Coherent States and Their Applications*. Springer-Verlag.

[Pérez-Delgado et al., 2012] Pérez-Delgado, C., Pearce, M., and Kok, P. (2012). Fundamental Limits of Classical and Quantum Imaging. *Phys. Rev. Lett.*, 109(12).

[Pittman et al., 1995] Pittman, T. B., Shih, Y. H., Strekalov, D. V., and Sergienko, A. V. (1995). Optical imaging by means of two-photon quantum entanglement. *Phys. Rev. A*, 52(5):R3429–R3432.

[Purcell, 1956] Purcell, E. M. (1956). Question of correlation between photons in coherent light rays. *Nature*, 178:1449–50.

[Rayleigh, 1879] Rayleigh, L. (1879). XXXI. Investigations in optics, with special reference to the spectroscope. *Philos. Mag. Ser. 5*, 8(49):261–274.

[Reed, 1962] Reed, I. (1962). On a moment theorem for complex Gaussian processes. *IRE Trans. Inf. Theory*, 8(3):0–1.

[Rust et al., 2006] Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods*, 3(10):793–5.

[Silva and Freire Jr, 2013] Silva, I. and Freire Jr, O. (2013). The Concept of the Photon in Question: The Controversy Surrounding the HBT Effect circa 1956–1958. *Hist. Stud. Nat. Sci.*, 43(4):453–491.

[Sudarshan, 1963] Sudarshan, E. (1963). Equivalence of Semiclassical and Quantum Mechanical Descriptions of Statistical Light Beams. *Phys. Rev. Lett.*, 10(7):277–279.

[Thiel et al., 2007] Thiel, C., Bastin, T., Martin, J., Solano, E., von Zanthier, J., and Agarwal, G. (2007). Quantum imaging with incoherent photons. *Phys. Rev. Lett*, 99(13):133603.

[Titulaer and Glauber, 1965] Titulaer, U. M. and Glauber, R. J. (1965). Correlation Functions for Coherent Fields. *Phys. Rev.*, 140(3B):B676–B682.

[van Cittert, 1934] van Cittert, P. (1934). Die Wahrscheinliche Schwingungsverteilung in Einer von Einer Lichtquelle Direkt Oder Mittels Einer Linse Beleuchteten Ebene. *Physica*, 1(1-6):201–210.

[van der Vaart, 1998] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[Walls and Milburn, 2008] Walls, D. F. and Milburn, G. J. (2008). *Quantum Optics*. Springer, 2nd edition.

[Wolf, 1955] Wolf, E. (1955). A Macroscopic Theory of Interference and Diffraction of Light from Finite Sources. II. Fields with a Spectral Range of Arbitrary Width. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 230(1181):246–265.

[Wootters, 1981] Wootters, W. (1981). Statistical distance and Hilbert space. *Phys. Rev. D*, 23(2):357–362.

[Zernike, 1938] Zernike, F. (1938). The concept of degree of coherence and its application to optical problems. *Physica*, 5(8):785–795.

[Zhou et al., 2010] Zhou, Y., Simon, J., Liu, J., and Shih, Y. (2010). Third-order correlation function and ghost imaging of chaotic thermal light in the photon counting regime. *Phys. Rev. A*, 81(4).