Super-resolution microscopy with spatial modes of light





José Inácio da Costa Filho

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2021

Abstract

The resolution of imaging devices is ultimately limited by the diffraction of light. This limit can be circumvented by exploiting near-field phenomena, non-linearities of the sample or specific illumination schemes, which led to the recent explosion of optical super-resolution techniques. However, it would be desirable to reach the nanoscale in the most general way possible, with far-field linear imaging of uniformly illuminated samples.

In the last few years, the answer to that came with the field of quantum-inspired super-resolution (QISR), where quantum metrological techniques were used to find novel measurement schemes that, beyond the conventional direct imaging of light, would achieve higher precisions. Interestingly, these ideas would always in some way or another involve transverse spatial modes of light. This thesis presents our work on the subject, with novel theoretical and experimental results.

Firstly, we develop the traditional theory of QISR, with has mostly focused on either the estimation of separation of two point sources or on geometric moments of samples. Then, we introduce our own results on optical imaging of general distributions of point sources, in a step to take QISR from toy examples to real imaging scenarios.

Secondly, we apply all the concepts studied on a experimental, proof-of-principle implementation of a QISR scheme dubbed Hermite-Gaussian Microscopy (HGM), which uses heterodyne detection with several Hermite-Gaussian modes as local oscillator in order to reconstruct the object sample and achieve super-resolution. We report the techniques necessary for the manipulation of spatial modes of light, as well as our results, where we demonstrated 2D imaging with twofold resolution enhancement beyond the diffraction limit.

Acknowledgements

First and foremost, I want to thank my supervisor Alex Lvovsky for everything he provided me during these four years of my DPhil. He gave me the unique opportunities of studying at both the University of Calgary and the University of Oxford, and of working at his excellent research group. He also made me more of a physicist and less of mathematician, making me think a lot more about the intuition behind things, and provided me the right balance of freedom and guidance so I could pursue and develop my own research ideas without getting lost in the process. Without him, my life and career would have been in totally different places, and I will be eternally thankful for that.

Secondly, I want to thank my main laboratory partners Nastya and Giorgio. I am definitely sure that without your help I wouldn't be able to take our main experiment to fruition, specially as we went through these disruptive pandemic times. You were my role models of what an experimental physicist should be, serving as a reference frame which guided me on what I should study and improve. You were also great friends with which I shared great times together, both in Calgary and Oxford. Finally, without you by now I would be either unemployed, as Giorgio referred me to an amazing postdoctoral position which I will undertake right after I submit this text, or stranded in Czech Republic, as Nastya lent me money there when my bank account unexpectedly reached the ground state. I also thank the important help of our other laboratory partners, Russell Bate, Bohan Li, Parth Patel and Jernej Frank. Parth and Jernej, the future of our experimental works lies in your hands now!

During these four years I wouldn't have maintained my sanity without my beloved friends. In Calgary, the Brazilian gang of Camila, Marta, Rapha, Ju and Evan (honorary Brazilian) kept my heart warm during these eight months of winter. I will never forget all the amazing Rocky Mountains hiking trips with Eugene and Arina, specially when I saw snow for the first time in my life (and took 800 pictures of it). In Oxford I had the company of my friends from the Brazilian Society, Darlan, Mari, Isa, Licio, Sergio and Bia, with which we organized awesome events... until the pandemic hit. From my College, I could share endless dining room meals (and a few restaurant ones) with Ferhat, Peter, Wangzhe, Jan, Farah, Oana and Vlad, and I had the coolest and tidiest housemates with Pete, George, Govind, Ana and

Emily. I am also glad to have shared the office, where I spent most of my time in Oxford, with David, Tom, James, Xianxin, Aleksei, Jiawei and Mert.

I wouldn't have gone anywhere without my family. Catarina and José Inácio, being a son of university professors is one of the greatest privileges one can have. All I achieved in life stems from the unending support, care and love you unconditionally gave me, and I thank you for making me the happy and accomplished person I am today. Carina and Tião, which were more like a second set of parents due to our age differences, are the best siblings one can ask for. I also gratefully thank my lovely sister-in-law Isabella, my beloved nephews Vitor (who managed to start Uni before I finished mine) and artist-in-the-making João, and my brilliant niece Valentina.

Finally, as we save dessert for the end of dinner, I would like to thank my wife Lorena Coelho. Although we only physically met after the middle of my DPhil, you were spiritually, or "whatsappaly", there since before it started. In these two crazy years we went from friends in separate continents to a married couple living together with two dogs and two cats, and already planning a move to another continent. You are the person who understands me the most in the Universe and I want to be forever at your side. I thank you for staying by mine during this never-ending DPhil, when every week I would promise that next week would be the last. Even now, as I finish this Acknowledgements at 1 am so that I can submit my thesis and go to sleep, you are still here, watching Netflix besides me, and giving me support. You also brought to my life Joquinha (the closest a dog can get to a human being), Lilica, Didica and Estrela, the first members of a family that will only expand in time with love as its main tenet.

I gratefully acknowledge funding from the Department of Physics, University of Oxford.

Contents

List of Figures **viii**

1 Introduction **1**
1.1 Preface . . . 1
1.2 List of publications . . . 4
1.3 Thesis outline . . . 6

2 Parameter estimation theory **7**
2.1 General setting . . . 7
2.2 Cramér-Rao bound . . . 9
2.3 Quantum Cramér-Rao bound . . . 12
2.3.1 Quantum-optimal measurements . . . 13
2.3.2 Calculation techniques . . . 15

3 Classical optical imaging theory **23**
3.1 Imaging systems . . . 23
3.1.1 Fourier optics . . . 23
3.1.2 Point-spread function . . . 25
3.1.3 Spatial correlations . . . 26
3.2 Incoherent light . . . 27
3.3 Coherent light . . . 31
3.4 Measurements . . . 33

4 Statistical analysis of incoherent light imaging **37**
4.1 Two-point resolution . . . 37
4.1.1 Rayleigh's curse . . . 38
4.1.2 Mode sorting . . . 41
4.1.3 Maximum likelihood estimation . . . 47
4.1.4 Strong thermal light . . . 50
4.1.5 Further studies . . . 54
4.2 Intensity moments . . . 55
4.2.1 Quantum limit . . . 56

4.2.2 Mode sorting . . . 58
4.3 Imaging of point sources . . . 61
4.3.1 Quantum limit . . . 61
4.3.2 Mode sorting . . . 65
4.3.3 Maximum likelihood estimation . . . 73
4.3.4 Quadrature measurements . . . 75
4.3.5 Circular aperture . . . 76
4.3.6 Conclusion . . . 78

5 Statistical analysis of coherent light imaging **81**
5.1 Quantum Fisher information . . . 81
5.2 Fisher information . . . 84
5.3 Imaging scenarios . . . 86
5.3.1 Two-point resolution . . . 86
5.3.2 Sources' parameters . . . 90
5.4 Conclusion . . . 93

6 Spatial modulation of light **95**
6.1 Phase-only modulators . . . 95
6.1.1 Diffraction gratings . . . 95
6.1.2 Phase and amplitude modulation . . . 97
6.2 Modelling and calibration of LCoS SLMs . . . 104
6.2.1 Operating principle of LCoS SLMs . . . 104
6.2.2 Imperfections . . . 105
6.2.3 Experiment . . . 109
6.2.4 Results . . . 111
6.2.5 Compensation . . . 115
6.3 Conclusion . . . 118

7 Microscopy experiment **119**
7.1 Hermite-Gaussian microscopy . . . 120
7.1.1 Coherent light . . . 120
7.1.2 Shot-noise analysis . . . 122
7.1.3 Incoherent light . . . 125
7.2 Experiment . . . 127
7.2.1 Setup . . . 128
7.2.2 Procedures . . . 134
7.2.3 Neural network . . . 137
7.3 Results . . . 141
7.3.1 Qualitative results . . . 142

7.3.2 Quantitative results . . . 143
7.3.3 Comparison with direct imaging and deconvolution . . . 145
7.3.4 Shot noise behaviour . . . 146
7.4 Discussion . . . 148
7.4.1 What about Fisher information? . . . 148
7.4.2 What about incoherent sources? . . . 149
7.4.3 What about microscopic samples? . . . 149
7.4.4 Other adaptations . . . 150

8 Conclusion **151**
8.1 Final words . . . 151
8.2 My vision for the future . . . 151

Appendices

A Density matrices review **154**

B Vectorization of a matrix **156**

C Coherent states and displacement operators **158**

D Quantum theory of imaging **163**
D.1 Image plane operators . . . 163
D.2 P-representation . . . 165
D.3 Quantum measurements . . . 169
D.4 Incoherent light . . . 173
D.5 Coherent light . . . 175

E Weak thermal light **177**

F Gaussian aperture properties **181**

G Circular aperture properties **185**

H Single mode Fisher information **191**

I Numerical evaluation of the QFI **195**

J An optical Eratosthenes' sieve for large prime numbers **198**

List of Figures

2.1 Quantum metrology pipeline 8

3.1 Propagation of light between apertures 24
3.2 Imaging system schematics 26
3.3 Airy and Gaussian point-spread functions 28
3.4 Two incoherent sources above, at, and below the Rayleigh limit . . 29
3.5 Two coherent sources at the Rayleigh limit 33
3.6 Direct imaging and mode sorting 34
3.7 Homodyne and heterodyne detection 35

4.1 Two-point resolution problem 38
4.2 QFI, direct imaging and binary SPADE 40
4.3 Hermite-Gaussian modes 43
4.4 Bessel modes 45
4.5 Effect of PAD basis mismatch 46
4.6 PAD basis saturation 46
4.7 Bias and mean squared error for the two sources direct imaging MLE 48
4.8 Bias and mean squared error for the two sources mode sorting MLE 49
4.9 Direct imaging versus mode sorting MLEs 49
4.10 Two-point resolution for strong incoherent light 51
4.11 Example of sample objects 63
4.12 Resolution of QFI and direct imaging 64
4.13 Quantum and direct imaging total precision for estimation of intensities or positions only 65
4.14 2D Hermite-Gaussian modes 66
4.15 Ghost images 67
4.16 Intensity and covered area of HG modes 67
4.17 Resolution of QFI, direct imaging and mode sorting 68
4.18 Total precision versus modes' centroid displacement 69
4.19 Optimal mode displacement versus mode order 70
4.20 Resolution of QFI, direct imaging, and regular and displaced mode sorting 70

4.21 Total precision of scanning mode sorting as a function of the number of scans . . . 71
4.22 Resolution of QFI, direct imaging, and regular, displaced and scanning mode sorting . . . 72
4.23 iPAD basis . . . 72
4.24 Resolution of QFI, direct imaging, and regular, displaced, scanning and interferometric mode sorting . . . 73
4.25 MLE of point sources . . . 74
4.26 Resolution of QFI, direct imaging, and regular, displaced, scanning, interferometric and interferometric with quadratures mode sorting . 75
4.27 Effect of changing the number of modes on resolution . . . 76
4.28 Fourier-Zernike modes . . . 77
4.29 Resolution of QFI and measurement schemes for circular aperture . 78

5.1 Average photon number at the image plane for coherent light . . . 87
5.2 Comparison of per-photon coherent QFI definitions . . . 89
5.3 Per-photon coherent light two sources QFI and FI for direct imaging 91
5.4 Total precision for estimation of positions, for zero and non-zero relative phases . . . 92

6.1 Blazed gratings and their phase-unwrapped versions . . . 96
6.2 Diffraction orders for various grating amplitudes . . . 98
6.3 Intensity of the first diffraction orders as a function of grating amplitude 99
6.4 Amplitude and phase modulation scheme . . . 100
6.5 Step-wise function approximation . . . 102
6.6 Example of SLM hologram . . . 103
6.7 SLM layered structure . . . 106
6.8 Reflection and transmission at SLM layers . . . 107
6.9 Crosstalk correction algorithm flowchart . . . 109
6.10 Diffraction orders near and far field for different grating amplitudes 112
6.11 SLM phase map from fit and wavefront sensor . . . 113
6.12 Fit quality for first and second orders . . . 114
6.13 Near and far field corrected profiles . . . 116
6.14 Simulated $HG_{12,12}$ fidelity . . . 117
6.15 Simulated power efficiency as a function of crosstalk PSF width . . 117

7.1 Resolvable distance as a function of mode order for coherent light . 123
7.2 Oxford logo ground truth . . . 124
7.3 Oxford logo reconstruction under different shot noise values . . . 124
7.4 Resolvable distance as a function of mode order for coherent light . 127

7.5 Schematic of the experiment . . . 128
7.6 LCoS SLM. . . . 129
7.7 Iris and metal ruler. . . . 130
7.8 Heterodyne setup . . . 131
7.9 SLM holograms for various HG modes . . . 132
7.10 Digital micromirror device. . . . 133
7.11 Electronics setup. . . . 134
7.12 Examples of DMD frames used in the experiment . . . 136
7.13 Neural network . . . 139
7.14 Original, reconstructed and direct images of the Oxford logo and coat-of-arms . . . 142
7.15 Original, reconstructed and direct images of alphabet symbols . . . 143
7.16 Original, reconstructed and direct images of pairs of lines . . . 143
7.17 Qualitative assessment of the resolution improvements of HGM . . . 144
7.18 Resolution improvement for Oxford logo, Oxford coat of arms and and the text symbols test set . . . 145
7.19 RL deconvolution results . . . 146
7.20 Simulated NN reconstructions and direct imaging with different shot noise values . . . 147

D.1 Object and image plane operators . . . 164

J.1 Light intensity profiles at near and far fields . . . 200
J.2 Optical sieve diffraction mask (left) in the case of $\{n\} = \{2, 3, 5, 7\}$. 202
J.3 Experimental implementation of the optical prime sieve . . . 203
J.4 Experimental results . . . 204
J.5 Effect of the SLM's spatial resolution on the produced far field intensity profile . . . 205
J.6 Number of misclassified primes and composites . . . 207

1 | Introduction

"Forgive me if what has seemed little to you, to me is all."

José Saramago

1.1 Preface

The history of imaging systems is a tale of exploration. Just as vehicles enabled us to overpower our feet and reach undiscovered territories, microscopes and telescopes, as extensions of our eyes, allowed us to see deep into the uncharted microscopic and celestial realms.

This exploration was powered by lenses, which could collect and focus light beams in ways that our eyes weren't capable of. With lens-making technology and design improving over the centuries, we could venture further and further into the unseen, bringing along many scientific advances as we progressed.

That quest would eventually hit a brick wall. As noted and formalized by Ernst Abbe [1, 2] and Lord Rayleigh [3] in the nineteenth century, an imaging system, regardless of its intricacies, has its resolution ultimately bounded by its angular aperture and the optical wavelength used. This, dubbed the *Abbe-Rayleigh* or *diffraction limit*, is a fundamental law of nature and a direct consequence of the wave aspect of light. Therefore, in order to keep progressing, only two options were left: either increase the aperture or decrease the wavelength.

For telescopes, which image far-away objects, the operating wavelength is determined and therefore constrained by the source, but their apertures can be increased with bigger lenses or larger baselines, albeit with *astronomically* increasing

costs. For microscopes, which already worked close to the angular aperture limit since their objects sit very close to their lenses, the way out was via shorter wavelengths, which eventually led to the powerful electron microscope. However, as stronger radiation damages most biological samples (ruling out *in vivo* imaging), among other limitations, optical microscopes were still required for many imaging scenarios. Scientists tried, even had some success, but could not get rid of the diffraction limit.

With the advent of cameras, photodetectors and the use of computational processing, scientists realized that, by careful statistical analysis of the photons hitting the image plane, much higher imaging resolutions, beyond what was deemed possible by the Rayleigh-Abbe criteria, could be achieved. Deconvolution, denoising and other image processing technique seemed to take optical microscopy to the next level but, was the diffraction limit finally broken? Unfortunately, no. Indeed, optical resolution was now limited by the stochastic behaviour of photons — shot noise — as a consequence of the particle nature of light, an aspect which the traditional, heuristic resolution criteria failed to account for.

In order to incorporate the wave-particle nature of light, the diffraction limit had to be reinterpreted in the rigorous framework of statistical inference [4, 5]. There, resolution is defined as the precision with which one can estimate the light source's distribution parameters with a given number of detected photons, and the limit is manifested in the fact that this precision sharply deteriorates as the sources' separations decrease past some point (which happens to be somehow close to the traditional Rayleigh-Abbe resolvable distance). Under this new light, the diffraction limit remained undefeated, and this statistical version of it was dubbed *Rayleigh's curse* [6].

No wall is large enough that it cannot be circumvented. The validity of the Abbe-Rayleigh limit or Rayleigh's curse depends on a set of assumptions: that the incoming light is uniform and incoherent, processed in its far-field, and only linear-optical imaging processes are involved [7]. In the last decades, the realization that breaking at least one of such hypotheses could make the diffraction limit irrelevant led to an explosion of new techniques, which now constitute the well-established field of

super-resolution microscopy [8]. These methods rely, for example, on using coherent shorter-wavelength radiation [9], near-field probing [10], nonclassical [11] or nonlinear [8, 12] optical properties of the object or switching the object's emission on and off [13–15]. With these advances, optical microscopy could finally reach the nanoscale.

Since there is no free lunch in nature, each of these techniques came with its own host of limitations, in addition to the increased cost and complexity. Near-field probing requires extremely small working distances, nonclassical light is difficult to generate, non-linear effects are dependent on the nature of the sample, and so forth. Furthermore, many of these methods were unfeasible for astronomy, where far-field and passive imaging are non-negotiable. This led to a situation where there is almost one different technique for every object to be imaged, but many imaging scenarios are still not covered. Scientists are still happy, but they could be happier.

In 2015, a major breakthrough[1] shook the whole landscape of optical imaging. Tsang *et al.* [6] showed, using tools of quantum statistical inference, that it was possible to achieve higher resolutions than what Rayleigh's curse allowed in in far-field, linear, passive incoherent light imaging . The catch? Not restricting oneself to direct photon intensity measurements only. Indeed, they also showed that this new resolution limit could be achieved if the far-field light was demultiplexed into an orthonormal basis of transverse spatial modes prior to detection. Furthermore, in the simple case of estimation of the separation of two point sources, the achieved precision was independent of scale: Rayleigh's curse was finally lifted!

Subsequent theoretical research [16–86] has extended the results of Ref. [6] to several imaging scenarios beyond the simple case of two point sources and the usual incoherent light. Unfortunately (again!), entirely lifting Rayleigh's curse was only possible for a few simple, mostly idealized cases. However, in a plethora of situations, resolutions beyond what was possible for conventional direct imaging methods were shown to be achievable. This new class of schemes, all of which involve some coherent processing of the far-field light and which were discovered using quantum metrological tools, was dubbed *quantum-inspired super-resolution* (QISR) [28].

[1] Although — at the time of writing, and in the author's opinion — not as well recognized as it should be.

On the experimental front, this new theory was implemented by various groups [87–107]. However, to date, these experiments were restricted to simple imaging scenarios, such as one or two point sources. Nevertheless, they have confirmed that these quantum-inspired methods are superior, practical, and have a place in the vast arsenal of super-resolution techniques already well-established in science and industry.

So, which further turns will this imaging exploration take? Where will it lead us to? After centuries of steady progress, it reached an insurmountable barrier put by nature. In order to keep going, it had to branch into a multitude of different paths, all of which found their own means of dealing with hurdles along the way. Quantum-inspired super-resolution is just another one of these paths, but one that holds great promise. However, it is a very recent discovery, and not enough of it is known, both theoretically and experimentally. The purpose of this thesis, and the works which it refers to, is to shine some light on important aspects of QISR that the scientific literature hasn't been able to clarify yet. Namely, we report the first experimental realization of QISR for a full 2D imaging task with coherent light, and the theory behind it, for both incoherent and coherent light. This work is a small effort towards taking QISR from its small niche inside the physics community to the broader scientific public, and we hope to convince you, if you're not already, that this road has a long way forward.

1.2 List of publications

I was involved in the following publications and preprints during my DPhil, listed here in chronological order.

- A. A. Pushkina, **J. I. Costa-Filho**, G. Maltese, A. I. Lvovsky. "Comprehensive model and performance optimization of phase-only spatial light modulators". *Measurement Science and Technology 31 12* (2020).

This work was a consequence of our quest for improving accuracy in our spatial light modulators. I was the one who noticed the cavity effects, which were

causing unexplainable amplitude modulations in our signals, and developed the compensation scheme. Besides that, I helped with the optical setup and the data analysis. It is reported in Chapter 6.

- B. Li, G. Maltese, **J. I. Costa-Filho**, A. A. Pushkina, A. I. Lvovsky. "Optical Erastothenes' sieve for large prime numbers". *Optics Express 28 8* (2020).

This paper was a summer project of student Bohan Li. I came up with the idea of doing the experiment, after presenting its theoretical proposal on our Journal Club, and assisted in the data analysis and writing of the paper. This "spin-off" work, which is a nice application of our spatial light modulation skills, is reported in Appendix J.

- A. A. Pushkina, G. Maltese, **J. I. Costa-Filho**, P. Patel, A. I. Lvovsky. "Super-resolution linear optical imaging in the far field". *ArXiv preprint 2105.01743* (2021). Accepted for publication at *Physical Review Letters*.

This is our main work, reported in Chapter 7. I helped with the experimental setup and worked on the data analysis, specially with respect to the post-processing with neural networks. Finally, I was assigned with writing the paper.

- **J. I. Costa-Filho**, A. I. Lvovsky. "Quantum precision bounds of coherent light imaging". *Manuscript in preparation*.

- **J. I. Costa-Filho**, A. I. Lvovsky. "Quantum-inspired super-resolution of incoherent point sources". *Manuscript in preparation*.

These two final theoretical manuscripts were fully conceived and developed by me, under the careful guidance of Alex Lvovsky.

1.3 Thesis outline

This thesis is roughly divided into three parts, and structured as follows. The first part consists of the background theory that supports the whole text. Chapter 2 introduces quantum metrology and the all-important classical and quantum Cramér-Rao bounds, the central theorems that guide our search for new measurement schemes. It also discusses several techniques to calculate the quantum Fisher information. Chapter 3 reviews the classical theory of light imaging, which is complemented by Appendix D, it's quantum counterpart.

The second part contains our theoretical results. Chapter 4 starts with a review of the advances in the new field of quantum-inspired super-resolution for incoherent (thermal) light, in order to introduce the main concepts and put things into perspective, and ends (Section 4.3) with our own results on imaging of a general 2D distribution of point sources [108]. Then, Chapter 5 does the same, but this time for coherent light imaging. This chapter expands on our work in Ref. [109].

The final part of the thesis concerns our experimental results. Chapter 6 introduces the general theory of spatial light modulators (SLMs), with special emphasis on phase-only SLMs. Then, we report our results [110] on the modelling and calibration of a liquid crystal on silicon SLM. This work resulted from our efforts to increase the accuracy of our SLMs, in order to use them for our final microscopy project. In Chapter 7, we outline the theory of Hermite-Gaussian microscopy (HGM), which uses mode sorting photocurrents in order to reconstruct objects, and report in extensive detail our microscopy experiment [111]. Finally, in Chapter 8 we provide an outlook on the future directions of this research area.

2 | Parameter estimation theory

In this chapter, we introduce the main concepts and results from quantum parameter estimation, or *quantum metrology*, that will be necessary for the remaining of this thesis. It is assumed the reader is familiar with basic quantum mechanics and the density matrix formalism, which is briefly reviewed in Appendix A. Quantum metrology provides the framework for all our theoretical results of chapters 4 and 5.

2.1 General setting

Suppose one is interested in estimating a set of parameters which are encoded in a certain quantum state. How, and how well, can it be done? In the paradigm of quantum metrology [11, 112–116], the usual procedure is to measure several copies of that state and use the collected data to estimate the parameters. Our goal, then, is to find the combination of measurements and estimates that lead to the smallest error.

In what follows, we assume some constraints over this setup. Firstly, we restrict ourselves to independent measurements over single copies of the state. Joint or *collective* measurements, i.e., measurements which are performed over multiple copies of the probe state at once, although very promising for imaging [56, 92], are of difficult experimental implementation and the theory behind them is outside the scope of this text. Secondly, we presuppose the parameters, although unknown, are deterministically defined: they are fixed and can be exactly determined. Working with random parameters, which will also not be covered here, is the subject of *Bayesian estimation theory* [117, 118].

In physics notation, our quantum metrology paradigm, schematised in Fig. 2.1, can be stated as follows. A vector of real parameters $\boldsymbol{\theta}$, also known as *ground truth*, is

encoded in a quantum state $\rho(\boldsymbol{\theta})$. After performing a general quantum measurement or *POVM* (See Appendix A) $\Pi = \{\Pi_m\}$ we obtain the outcome m with probability

$$\Pr(m|\boldsymbol{\theta}) = \operatorname{tr}\{\Pi(m)\rho(\boldsymbol{\theta})\}. \tag{2.1}$$

After repeating the measurement N times over single copies of $\rho(\boldsymbol{\theta})$, we are left with a dataset $\{m_j\}_{j=1}^N$, that happens with probability

$$\Pr(\{m_j\}|\boldsymbol{\theta}) = \prod_{j=1}^{N} \Pr(m_j|\boldsymbol{\theta}). \tag{2.2}$$

With the data, we can build an *estimator* $\hat{\boldsymbol{\theta}}(\{m_j\}_{j=1}^N)$ of the ground truth parameters. An estimator is simply a function $\hat{\boldsymbol{\theta}}(\cdot)$ from the collected data $\{m_j\}_{j=1}^N$ to the parameter space, and the whole purpose of quantum metrology is to find an estimator that takes the data we measured to a set of parameters that lies as close as possible to the ground truth.

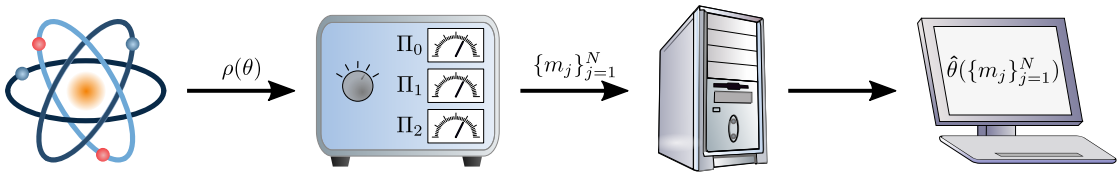


Figure 2.1: Depiction of the quantum metrology pipeline.

How to benchmark an estimation performance? we usually want our predictions to be accurate, i.e., close to the ground truth, and precise, i.e., close to one another. Naively, we would quantify accuracy via the *estimator error* $\mathcal{E}(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. Since data are random, however, it is more practical to look at its average value, known as *bias*,

$$B(\hat{\boldsymbol{\theta}}) = \left\langle \mathcal{E}(\hat{\boldsymbol{\theta}}) \right\rangle = \langle \hat{\boldsymbol{\theta}} \rangle - \boldsymbol{\theta}. \tag{2.3}$$

An estimator with zero bias is said to be *unbiased*. For precision, we use the *error covariance matrix*

$$\operatorname{Cov}(\hat{\boldsymbol{\theta}}) = \left\langle \left(\hat{\boldsymbol{\theta}} - \langle \hat{\boldsymbol{\theta}} \rangle\right) \left(\hat{\boldsymbol{\theta}} - \langle \hat{\boldsymbol{\theta}} \rangle\right)^{\mathsf{T}} \right\rangle. \tag{2.4}$$

Its diagonal elements are the variances of each component of $\hat{\boldsymbol{\theta}}$,

$$\left[\operatorname{Cov}(\hat{\boldsymbol{\theta}})\right]_{ii} = \left\langle \left(\hat{\theta}_i - \langle \hat{\theta}_i \rangle\right)^2 \right\rangle \equiv \operatorname{Var}(\hat{\theta}_i), \tag{2.5}$$

which tell how much the predictions are spread around the average value, while its off-diagonal elements govern how the estimates of θ_i and θ_j correlate with one another. Eventually, we might want to condense the covariance matrix into a single number. That is usually done with its trace, the *total variance*, which captures the variances of individual parameters, or its determinant, the *generalized variance*, which focuses on the covariances. Finally, both the bias and the covariance can be summed in a single quantity, the *mean squared error* (MSE)

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}) = \left\langle \left\| \mathcal{E}(\hat{\boldsymbol{\theta}}) \right\|^2 \right\rangle = \mathrm{Tr}\,\mathrm{Cov}(\hat{\boldsymbol{\theta}}) + \left\| B(\hat{\boldsymbol{\theta}}) \right\|^2, \tag{2.6}$$

where $\|\boldsymbol{v}\| = \sqrt{\sum_i |v_i|^2}$ denotes the L^2 norm [119]. The MSE directly measures the spread of $\hat{\boldsymbol{\theta}}$ around the ground truth $\boldsymbol{\theta}$.

Minimizing the mean squared error of an estimator should be the end goal of our estimation task[1]. Its explicit dependence on the ground truth, however, makes it difficult to use in practice [120]. Therefore, the most common procedure is to restrict oneself to unbiased estimators only and, among them, look for the ones with smallest covariance[2] .

2.2 Cramér-Rao bound

For a given dataset, the estimation "landscape" is encapsulated by the *log-likelihood function*

$$\ell(\boldsymbol{\theta}) = \ln \Pr(\{m_j\}|\boldsymbol{\theta}) = \sum_j \ln \Pr(m_j|\boldsymbol{\theta}), \tag{2.7}$$

which tells us how likely the parameters $\boldsymbol{\theta}$ are of having generated the present data. Seeing these parameters as coordinates, $\ell(\boldsymbol{\theta})$ defines a surface in parameter space [121], whose geometry governs how estimators behave and what they can achieve. For example, its gradient, the *score* vector

$$\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}), \tag{2.8}$$

[1]The MSE is not the only one possible loss function to be minimised in an estimation task [120], but it is the one suitable for our purposes in this thesis.

[2]Although there might be biased estimators with smaller MSE [120], the unbiased results are, in most scenarios, still good enough.

indicates how sensitive the estimation is with respect to infinitesimal changes of the parameters.

With the score function we can find a lower bound on the covariance of any unbiased estimator, the celebrated *Cramér-Rao bound* (CRB) [117, 118]:

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \leq \langle \boldsymbol{\mathcal{L}}(\boldsymbol{\theta})\boldsymbol{\mathcal{L}}(\boldsymbol{\theta})^{\intercal}\rangle^{-1}, \tag{2.9}$$

where the matrix inequality $A \leq B$ means that $A - B$ is positive semidefinite. Although the bound can be exactly achieved only in a few restricted scenarios [118], it can always be saturated *asymptotically*, i.e., in the limit of a infinite number of measurements, by the *maximum likelihood estimator* (MLE),

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}), \tag{2.10}$$

i.e., the estimator that maximizes the log-likelihood function. An estimator that saturates the CRB is said to be *efficient*.

The Cramér-Rao bound is the holy grail of unbiased estimation, as it tells us the best estimator performance we can possibly get. It is a central theorem in this text, as it is in terms of it that the resolution of imaging systems is defined. For a given quantum measurement, the CRB returns the best attainable precision, and it is through it that different measurement schemes are compared.

Since finding an efficient estimator is a matter of art more than science, we usually make use of the MLE, which is guaranteed to be asymptotically efficient. However, it is important to note that, for a finite number of measurements, the MLE is not guaranteed to be efficient or even unbiased. For these reasons, the Cramér-Rao bound is mostly studied in the domain of *asymptotic statistics*, and for non-asymptotic regimes other bounds can yield better results [122].

The quantity $\langle \boldsymbol{\mathcal{L}}(\boldsymbol{\theta})\boldsymbol{\mathcal{L}}(\boldsymbol{\theta})^{\intercal}\rangle$, which appears on the right-hand side of the CRB, is called the *Fisher information* (FI) matrix. It is usually known in its expanded forms

$$\begin{aligned}[\mathcal{F}(\boldsymbol{\theta})]_{ij} &= \sum_{\{m_j\}} \left(\frac{\partial}{\partial\theta_i}\ln \Pr(m_j|\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_j}\ln \Pr(m_j|\boldsymbol{\theta})\right)\Pr(m_j|\boldsymbol{\theta}) \\ &= \sum_{\{m_j\}} \left(\frac{\partial}{\partial\theta_i}\Pr(m_j|\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_j}\Pr(m_j|\boldsymbol{\theta})\right)\Pr(m_j|\boldsymbol{\theta})^{-1},\end{aligned} \tag{2.11}$$

where the sums are replaced by integrals when the sample space is continuous. Its main properties are (the $\boldsymbol{\theta}$ dependence will be omitted where there is no risk of ambiguity):

- Real-valuedness and symmetry: $\mathcal{F}_{ij} = \mathcal{F}_{ji} \in \mathbb{R}$.
- Positive-semidefiniteness: $\mathcal{F} \geq 0$.
- Additivity: the FI for independent measurements is the sum of the individual FIs.
- Reparametrization: if $\boldsymbol{\eta}$ is a function of $\boldsymbol{\theta}$, then $\mathcal{F}(\boldsymbol{\theta}) = J^{\intercal}\mathcal{F}(\boldsymbol{\eta}(\boldsymbol{\theta}))J$, where $J_{ij} = \partial\eta_i/\partial\theta_j$ is the Jacobian matrix.

Note that, since we are only dealing with single-copy measurements, the additivity property allows us to re-cast the Cramér-Rao bound as

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \leq [N\mathcal{F}(\boldsymbol{\theta})]^{-1}\,, \tag{2.12}$$

where N is the number of measurements and $\mathcal{F}(\boldsymbol{\theta})$ is the FI of a single measurement.

When $\mathcal{F}_{ij} = 0$ for some i and j, we say that θ_i and θ_j are *orthogonal parameters*, since they can be independently evaluated as per the CRB and the MLE. Furthermore, if the FI matrix is diagonal, then all parameters obey

$$\mathrm{Var}(\hat{\theta}_i) \leq N/\mathcal{F}_{ii}. \tag{2.13}$$

An FI with zero eigenvalues is said to be *singular*, which means that some parameters cannot be estimated with finite variance. If it is not ($\mathcal{F} > 0$), then it is said to be *regular*, and then it obeys $1/\mathcal{F}_{ii} \leq [\mathcal{F}^{-1}]_{ii}$ for any i, with equality only when $\mathcal{F}$ is diagonal. This implies that, due to the CRB, the precision in estimating independent parameters is always higher than if they were correlated.

If the number of parameters to estimate (the dimension of $\boldsymbol{\theta}$) is too large, it might not be useful to keep track of the covariance error of each parameter [123]. In this case, a simpler figure of merit, such as the aforementioned total

and generalized variance, might be used. These quantities also obey relations similar to the Cramér-Rao bound:

$$\operatorname{Tr}\operatorname{cov}(\hat{\boldsymbol{\theta}}) \leq \operatorname{Tr}\left[N\mathcal{F}(\boldsymbol{\theta})^{-1}\right], \tag{2.14}$$

$$\det\operatorname{cov}(\hat{\boldsymbol{\theta}}) \leq 1/\det\left[N\mathcal{F}(\boldsymbol{\theta})\right]. \tag{2.15}$$

Eqs. (2.14) and (2.15) easily follow from the CRB after using the linearity of the trace function and the Minkowski determinant theorem [124], respectively.

2.3 Quantum Cramér-Rao bound

So far we have only optimized our estimators, keeping fixed the measurement that generated our data. In that regard, we have basically recovered the classical parameter estimation scenario. The full potential of quantum metrology comes into play as we are able to optimize the Cramér-Rao bound over all possible measurements and find the ultimate precision limit that can be achieved by any protocol over the state $\rho(\boldsymbol{\theta})$ and the parameters $\boldsymbol{\theta}$. This new limit is given by the *quantum Cramér-Rao* bound (QCRB) [112, 113]

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}) \leq \left[N\mathcal{F}_{\Pi}(\boldsymbol{\theta})\right]^{-1} \leq \left[N\mathcal{K}(\boldsymbol{\theta})\right]^{-1}, \tag{2.16}$$

where the subscript in the FI matrix emphasizes its dependence on a specific POVM Π, and $\mathcal{K}(\boldsymbol{\theta})$ is the *quantum Fisher information* (QFI) matrix, the most important quantity in this thesis.

The QFI matrix is defined as (dropping the $\boldsymbol{\theta}$-dependence)

$$[\mathcal{K}]_{ij} = \frac{1}{2}\left\langle\{\mathcal{L}_i, \mathcal{L}_j\}\right\rangle, \tag{2.17}$$

where $\langle\,\cdot\,\rangle = \operatorname{Tr}(\rho\,\cdot\,)$ is the quantum expected value, $\{A, B\} = AB + BA$ is the anti-commutator and $\mathcal{L}_i$ is the *quantum score* or *symmetric logarithmic derivative* (SLD), implicitly defined via the *Lyapunov equation* [123]

$$\partial_{\theta_i}\rho = \frac{1}{2}\{\rho, \mathcal{L}_i\}. \tag{2.18}$$

If ρ is diagonal, it corresponds to a classical probability distribution

$$\rho = \sum_m p_m \Pi_m, \tag{2.19}$$

which is equivalent to measuring the quantum state in its eigenbasis. In this case, it is easy to see that the SLD reduces to the classical score function: we have that

$$\mathcal{L}_i = \rho^{-1} \partial_{\theta_i} \rho = \partial_{\theta_i} \ln p_m. \tag{2.20}$$

As a corollary, the QFI becomes the FI with respect to the eigenbasis measurement. Therefore, the QFI can be seen as a quantum, non-commutative generalization of the FI[3] .

Similarly to its classical counterpart, the QFI is real, symmetric, positive-semidefinite and additive: if $\rho = \bigotimes_n \rho_n$, then $\mathcal{K}(\rho) = \sum_n \mathcal{K}(\rho_n)$. Furthermore, it is also invariant with respect to a $\boldsymbol{\theta}$-independent unitary operation U, $\mathcal{K}(\rho) = \mathcal{K}(U\rho U^\dagger)$. As for the SLD, they are Hermitian (as a direct consequence of the Hermiticity of ρ) and have zero expectation value, $\langle \mathcal{L}_i \rangle = 0$, since $\operatorname{Tr} \partial_{\theta_i} \rho = 0$.

2.3.1 Quantum-optimal measurements

Now, the million-dollar question is: which measurements, if any, are *quantum-optimal*, i.e., can saturate the QCRB? In order to answer that, we will start with the simple one-parameter case [126]. Here, the QFI is simply $\mathcal{K} = \operatorname{Tr}[\rho \mathcal{L}^2]$, where $\mathcal{L}$ is defined by $\partial_\theta \rho = \rho \mathcal{L}$, while the FI for a certain POVM Π is

$$\mathcal{F}_\Pi = \sum_m \left(\operatorname{Tr}\left[\partial_\theta \rho \Pi_m\right]\right)^2 / \operatorname{Tr}\left[\rho \Pi_m\right]. \tag{2.21}$$

It is straightforward to show that a measurement in the eigenbasis of $\mathcal{L}$ saturates the QCRB: if the eigendecomposition of $\mathcal{L}$ is $\mathcal{L} = \sum_m l_m \Pi_m$, then $\operatorname{Tr}[\partial_\theta \rho \Pi_m] = l_m \operatorname{Tr}[\rho \Pi_m]$ and

$$\mathcal{F}_\Pi = \sum_m l_m^2 \operatorname{Tr}[\rho \Pi_m] = \operatorname{Tr}\left[\rho \mathcal{L}^2\right] = \mathcal{K}. \tag{2.22}$$

[3]The SLD-based QFI is not the only quantum generalization of the FI [125], but it is the only one which will be covered here.

We might want to take this idea to the multiparameter case by optimally estimating each parameter via its own SLD measurement. It turns out, however, that distinct SLD measurements might be incompatible with one another, due to the non-commutativity of quantum mechanics [123]. Therefore, the QCRB is not attainable in general.

Then, under which conditions can we have quantum-optimal measurements? A first thought is to impose that the aforementioned limitation doesn't happen and all the SLD measurements are compatible with one another. This condition is satisfied when

$$[\mathcal{L}_i, \mathcal{L}_j] = 0 \tag{2.23}$$

for any i and j, i.e., all the SLD matrices commute. In this case, they can be all simultaneously diagonalized, and the optimal measurement is simply their common eigenbasis [123]. This happens, for example, when the QFI matrix is diagonal and all parameters can be independently estimated. More generally, a necessary condition for the achievability of the QCRB is that the *compatibility matrix* [127, 128]

$$\Gamma_{ij} = \frac{1}{2i} \langle [\mathcal{L}_i, \mathcal{L}_j] \rangle \tag{2.24}$$

is identically zero. This condition, which is a weaker version of Eq. (2.23), becomes sufficient too if we allow collective measurements to be made [129]. As a side note, the QFI and compatibility matrices can be seen as the real and imaginary parts, respectively, of a combined matrix

$$\mathcal{H}_{ij} = \langle \mathcal{L}_i \mathcal{L}_j \rangle \,. \tag{2.25}$$

Our general procedure for checking if the QCRB can be saturated is as follows. The first step is to calculate the commutators of the SLD matrices, Eq. (2.23). If they are all zero, then the eigenbasis of the SLDs is our desired quantum-optimal measurement. If not, we proceed with evaluating the compatibility matrix. If it is not identically zero, the QCRB cannot be saturated. If it is, then, we are guaranteed that some measurement is quantum-optimal. However, we don't have

a constructive method to find it and we don't know if it can be implemented by measurements over single copies of the quantum state, as collective measurements might be necessary. In this case, we have to resort to other methods in order to find the optimal measurements.

An astute reader might have noted that, since the SLD is a function of the ground truth parameters, in order to implement its eigenbasis measurement we might need prior knowledge of the very parameters we are trying to estimate. Apparently paradoxical, this situation can be sorted out with *adaptive estimation*, where some feedback mechanism ensures that both the estimates and the POVMs are updated after every measurement [130, 131]. For example, a two-stage process can be implemented where during a fraction of the measurements a rough estimate of the ground truth is done and, for the rest of the protocol, the desired POVM, evaluated on that estimate, is used. This simple implementation allows quantum-optimal measurements, even when dependent on the ground truth, to achieve the QCRB in the asymptotic limit [132]. In practice, more elaborate and complex schemes are used, incorporating Bayesian concepts or even machine learning techniques [133, 134].

2.3.2 Calculation techniques

While calculating the Fisher information matrix is usually straightforward, evaluating its quantum counterpart can be a very involved process. This is largely due to solving the Lyapunov equation, Eq.(2.18), which implicitly defines the SLD. In this section, we outline several techniques that can be used for this task. These techniques will be used throughout the text.

Pure states

We start with the special case of a pure quantum state,

$$\rho = |\psi\rangle \langle\psi| , \tag{2.26}$$

for which the QFI can be easily evaluated. Owing to the purity equation $\rho^2 = \rho$, we have that

$$\partial_i \rho = \partial_i \rho^2 = (\partial_i \rho)\rho + \rho(\partial_i \rho). \tag{2.27}$$

A direct comparison with Eq. (2.18) shows that $\mathcal{L}_i = 2\partial_i\rho$. This implies that the $\mathcal{H}$ matrix is simply

$$\mathcal{H}_{ij} = 4\left\langle \partial_i\rho\partial_j\rho \right\rangle . \tag{2.28}$$

In order to finish the calculation, we need to expand the derivatives with Eq. (2.26),

$$\partial_i\rho = \left|\partial_i\psi\right\rangle\left\langle\psi\right| + \left|\psi\right\rangle\left\langle\partial_i\psi\right| . \tag{2.29}$$

We also need the relation $\langle\partial_i\psi|\psi\rangle = -\langle\psi|\partial_i\psi\rangle$, which is a direct consequence of $\langle\mathcal{L}_i\rangle = 0$. With it, simple algebra leads to the final form

$$\mathcal{H}_{ij} = 4\left[\langle\partial_i\psi|\partial_j\psi\rangle - \langle\partial_i\psi|\psi\rangle\langle\psi|\partial_j\psi\rangle\right] . \tag{2.30}$$

From there, the QFI and compatibility matrices are $\mathcal{K} = \mathrm{Re}\{\mathcal{H}\}$ and $\Gamma = \mathrm{Im}\{\mathcal{H}\}$. The latter can be further simplified to

$$\Gamma_{ij} = 4\,\mathrm{Im}\left\langle\partial_i\psi|\partial_j\psi\right\rangle . \tag{2.31}$$

This technique will be used when dealing with coherent light sources, where the quantum state of light is pure.

Diagonalization method

For the general case, we can calculate the QFI matrix by diagonalization of the quantum state [123]. The idea is to express the SLDs in the state's eigenbasis and have a "coordinate-dependent" formula for the QFI. In order for that to be possible, ρ **must be full rank in the Hilbert space that contains the SLDs**, which means that its eigenbasis must span not only the support of ρ but also of all the derivatives $\partial_i\rho$ [135] or, equivalently, that the derivatives' support is contained in the state's support. The full rank assumption implies that ρ is positive-definite, so it can be diagonalized as

$$\rho = \sum_{a=1}^{\mathrm{rank}\rho} \lambda_a \left|\lambda_a\right\rangle\left\langle\lambda_a\right| , \tag{2.32}$$

where $\lambda_a > 0$ and both eigenvalues and eigenstates are functions of the ground truth $\boldsymbol{\theta}$. The SLDs can be written as

$$\mathcal{L}_i = \sum_{a,b=1}^{\text{rank}\rho} L_{ab}^{(i)} \left|\lambda_a\right\rangle \left\langle\lambda_b\right|, \tag{2.33}$$

where the coefficients $L_{ab}^{(i)}$ are, via Eq. (2.18),

$$L_{ab}^{(i)} = 2\frac{\left\langle\lambda_a\right| \partial_i \rho \left|\lambda_b\right\rangle}{\lambda_a + \lambda_b}. \tag{2.34}$$

By simply plugging this result in the corresponding formulas, we find that the QFI and compatibility matrices are

$$\mathcal{K}_{ij} = 2\sum_{a,b=1}^{\text{rank}\rho} \frac{1}{\lambda_a + \lambda_b} \text{Re}\{\left\langle\lambda_a\right| \partial_i \rho \left|\lambda_b\right\rangle \left\langle\lambda_b\right| \partial_j \rho \left|\lambda_a\right\rangle\}, \tag{2.35}$$

$$\Gamma_{ij} = -2i\sum_{a,b=1}^{\text{rank}\rho} \frac{(\lambda_a - \lambda_b)}{(\lambda_a + \lambda_b)^2} \text{Im}\{\left\langle\lambda_a\right| \partial_i \rho \left|\lambda_b\right\rangle \left\langle\lambda_b\right| \partial_j \rho \left|\lambda_a\right\rangle\}. \tag{2.36}$$

These formulas can be easily adapted for the situation where ρ is not full rank. In that case, we simply drop the diverging summation terms, i.e., the ones for which $\lambda_a + \lambda_b = 0$.

The diagonalization method is the most commonly used in the literature. However, it becomes very cumbersome as the dimension of the system grows. Recently, several other techniques have been devised in order to circumvent the diagonalization step, which is the main bottleneck of the method.

Matrix inversion method

As discovered by Ref. [136], the diagonalization of the quantum state can be avoided by making use of the *vectorization* operation, which turns a matrix M into a vector $\left|M\right\rangle\rangle$ which consists of its columns "stacked" on top of each other. The properties of vectorization are outlined in Appendix B, and here we will need the following:

$$\langle\langle M|N\rangle\rangle = \text{Tr}\, M^\dagger N, \tag{2.37}$$

$$\left|MNO\right\rangle\rangle = (O^\intercal \otimes M)\left|N\right\rangle\rangle. \tag{2.38}$$

To start, we vectorize the Lyapunov equation, Eq. (2.18), and use Eq. (2.38) to reach

$$\begin{aligned} |\,\partial_i\rho\rangle\rangle &= \frac{1}{2}\left(|\,\rho\mathcal{L}_i\rangle\rangle + |\,\mathcal{L}_i\rho\rangle\rangle\right) \\ &= \frac{1}{2}\left(\rho^{\intercal}\otimes I + I\otimes\rho\right)|\,\mathcal{L}_i\rangle\rangle, \end{aligned} \tag{2.39}$$

We then define the matrix

$$\mathcal{M} = \frac{1}{2}\left(\rho^{\intercal}\otimes I + I\otimes\rho\right), \tag{2.40}$$

with which we express the vectorization of the SLDs as

$$|\,\mathcal{L}_i\rangle\rangle = \mathcal{M}^{-1}|\,\partial_i\rho\rangle\rangle. \tag{2.41}$$

Finally, note that the QFI matrix can be rewritten as

$$\begin{aligned} \mathcal{K}_{ij} &= \frac{1}{2}\,\mathrm{Tr}\left[\rho\mathcal{L}_i\mathcal{L}_j + \rho\mathcal{L}_j\mathcal{L}_i\right] \\ &= \mathrm{Tr}\left[\frac{1}{2}\left(\rho\mathcal{L}_i + \mathcal{L}_i\rho\right)\mathcal{L}_j\right] = \mathrm{Tr}\left[\partial_i\rho\mathcal{L}_j\right]. \end{aligned} \tag{2.42}$$

Then, by using Eq. (2.37), we come to the final form of the QFI,

$$\mathcal{K}_{ij} = \langle\langle\partial_i\rho|\mathcal{L}_j\rangle\rangle = \langle\langle\partial_i\rho\,|\mathcal{M}^{-1}|\,\partial_j\rho\rangle\rangle. \tag{2.43}$$

Analogously, we can find the compatibility matrix as well,

$$\Gamma_{ij} = \frac{1}{2i}\langle\langle\partial_i\rho\,|\mathcal{M}^{-1}\left(\rho^{\intercal}\otimes I - I\otimes\rho\right)\mathcal{M}^{-1}|\,\partial_j\rho\rangle\rangle. \tag{2.44}$$

With this method, we have shifted the calculation burden from a matrix diagonalization to a matrix inversion, which might be more convenient depending on the problem in hand. It also has the advantage of being basis independent. However, it has no advantage in terms of computational complexity over the previous method [136], and **it still requires the quantum state to be full rank**.

Non-orthogonal bases

Another way of avoiding the diagonalization of the quantum state is by making use of non-orthogonal bases [135]. In this method, we expand ρ in some basis $\{\psi_a\}$ which is not necessarily orthogonal,

$$\rho = \sum_{a,b} P_{ab} \left|\psi_a\right\rangle \left\langle\psi_b\right| . \tag{2.45}$$

This approach is extremely convenient, as many models are already formulated in terms of non-orthogonal bases. It also does not require that the quantum state is full rank. Our only requirement is that $\{\psi_a\}$ is also a basis for the full system, i.e., its elements are linearly independent and they cover the full support of ρ and its derivatives $\partial_i\rho$. The former condition is satisfied if the determinant of the *Gramian matrix*

$$\Upsilon_{ab} = \langle\psi_a|\psi_b\rangle \tag{2.46}$$

is nonzero. For the latter, the suggested procedure is to start with a basis for the support of ρ, which is usually already given in the problem, and compute all the nonzero derivatives of this basis elements. Then, one by one, add them to the basis and check if the resulting set is linearly independent. If not, simply remove that element and proceed to the next one. By the end, one will be left with the desired basis [135].

In sequence, we expand the derivatives and the SLDs in this basis,

$$\partial_i\rho = \sum_{a,b} D^{(i)}_{ab} \left|\psi_a\right\rangle \left\langle\psi_b\right| , \tag{2.47}$$

$$\mathcal{L}_i = \sum_{a,b} L^{(i)}_{ab} \left|\psi_a\right\rangle \left\langle\psi_b\right| , \tag{2.48}$$

defining the matrices $D^{(i)}$ and $L^{(i)}$. Direct substitution into Eq. (2.18) leads to

$$2D^{(i)} = L^{(i)}\Upsilon P + P\Upsilon L^{(i)}, \tag{2.49}$$

with which the $L^{(i)}$ matrices are determined. With them, we can calculate the quantity in Eq. (2.25),

$$\mathcal{H}_{ij} = \mathrm{Tr}\left[P\Upsilon L^{(i)}\Upsilon L^{(j)}\Upsilon\right] , \tag{2.50}$$

whose real and imaginary parts yield the QFI and compatibility matrices, respectively.

In analogy with Eq. (2.39), we can use vectorization to evaluate Eq. (2.49),

$$\left|D^{(i)}\right\rangle\!\!\right\rangle = \frac{1}{2}\left[(\Upsilon P)^{\intercal}\otimes I + I\otimes P\Upsilon\right]\left|L^{(i)}\right\rangle\!\!\right\rangle \equiv \mathbb{M}\left|L^{(i)}\right\rangle\!\!\right\rangle. \tag{2.51}$$

Via similar steps, we find the QFI matrix to be

$$\mathcal{K}_{ij} = \left\langle\!\!\left\langle D^{(i)}\right|\left(\Upsilon^{\intercal}\otimes\Upsilon\right)\mathbb{M}^{-1}\left|D^{(j)}\right\rangle\!\!\right\rangle, \tag{2.52}$$

while the compatibility matrix is

$$\Gamma_{ij} = \frac{1}{2i}\left\langle\!\!\left\langle D^{(i)}\right|\mathbb{M}^{-1}\left((P\Upsilon)^{\intercal}\otimes I - I\otimes\Upsilon P\right)\left(\Upsilon^{\intercal}\otimes\Upsilon\right)\mathbb{M}^{-1}\left|D^{(j)}\right\rangle\!\!\right\rangle. \tag{2.53}$$

Note, however, that if ρ is not full rank, then the matrix $\mathbb{M}$ is not invertible, and this method does not work. However, it is shown to work [136] if we replace the inverse of $\mathbb{M}$ with its pseudo-inverse, which is well-defined even for non-invertible matrices.

The method of non-orthogonal bases is what we use for calculating the QFI for weak incoherent light sources in Appendix I, as it is both practical and suitable for non full rank states.

Geometrical method

This final technique is a highly analytical approach for QFI evaluations. In this text, we only use it once, in Chapter 4, as a shorter and simpler alternative to the usual QFI calculation for separation estimation of two incoherent point sources. Although its use cases are limited, it can provide us deep insights into the hidden geometrical structure of quantum metrology. For this reason, it is briefly showcased here and it can be safely skipped by a uninterested reader.

Akin to the log-likelihood function $\ell(\boldsymbol{\theta})$ in classical statistics, the family of density matrices $\rho(\boldsymbol{\theta})$ defines a surface whose geometry is closely tied to the quantum Fisher information. Although the study of the geometric properties of quantum metrology [125] will not be pursued here, we will use a famous result that is useful for QFI calculations.

The *Bures distance* [125] between operators ρ and σ is

$$D_B(\rho,\sigma) = \sqrt{2(1-F(\rho,\sigma))}, \tag{2.54}$$

where

$$F(\rho,\sigma) = \mathrm{Tr}\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}} \tag{2.55}$$

is the *quantum fidelity* between the states. It can be equivalently rewritten as [137]

$$F(\rho,\sigma) = \max_{|\psi_\rho\rangle,|\psi_\sigma\rangle} |\langle\psi_\rho|\psi_\sigma\rangle|\,, \tag{2.56}$$

where $|\psi_\rho\rangle$ and $|\psi_\sigma\rangle$ are *purifications* (see Appendix A, Eq. (A.3)) of ρ and σ, respectively, and the maximization is over all such purifications. Since any two purifications $|\psi_\rho\rangle$ and $|\phi_\rho\rangle$ of a quantum state are related by an isometry[4] over their ancilla spaces (see Appendix A, Eq. (A.4)),

$$|\psi_\rho\rangle = (I\otimes V)\,|\phi_\rho\rangle\,, \tag{2.57}$$

the fidelity can also be expressed by a maximization over all isometries [51],

$$F(\rho,\sigma) = \max_V |\langle\psi_\rho|\,(I\otimes V)\,|\psi_\sigma\rangle|\,, \tag{2.58}$$

where $|\psi_\rho\rangle$ and $|\psi_\sigma\rangle$ are now *any* two purifications of ρ and σ, respectively.

Now consider two infinitesimally separated parameter values $\boldsymbol{\theta}$ and $\boldsymbol{\theta}+\mathrm{d}\boldsymbol{\theta}$. Assuming that the rank of ρ is unchanged due to this small displacement, the infinitesimal (squared) Bures distance between $\rho(\boldsymbol{\theta})$ and $\rho(\boldsymbol{\theta}+\mathrm{d}\boldsymbol{\theta})$ can be expressed as [123, 138]

$$D_B^2(\rho(\boldsymbol{\theta}),\rho(\boldsymbol{\theta}+\mathrm{d}\boldsymbol{\theta})) = \frac{1}{4}\sum_{ij}\mathcal{K}_{ij}\mathrm{d}\theta_i\mathrm{d}\theta_j, \tag{2.59}$$

which is a famous result that will not be derived here. Eq. (2.59) is a powerful connection between the geometry of the quantum statistical manifold and the QFI matrix, and is an alternative way to evaluate the latter. For example, in the single parameter case, we have that

$$\mathcal{K}_\theta = \lim_{\mathrm{d}\theta\to 0} 8\frac{1-F(\rho(\theta),\rho(\theta+\mathrm{d}\theta))}{\mathrm{d}\theta^2}, \tag{2.60}$$

[4] An isometry V is characterized by the property $V^\dagger V = I$.

which can also be expressed as [138]

$$\mathcal{K}_\theta = -4\frac{\partial^2 F(\rho(\theta_1), \rho(\theta_2))}{\partial \theta_2^2}\Big|_{\theta_1=\theta_2=\theta}. \tag{2.61}$$

The expression for the quantum fidelity can be further simplified. Consider a decomposition of the purification $|\psi(\boldsymbol{\theta})\rangle$ of $\rho(\boldsymbol{\theta})$,

$$|\psi(\boldsymbol{\theta})\rangle = \sum_{j,s} c_{js}(\boldsymbol{\theta}) \, |j\rangle \otimes |s\rangle \, , \tag{2.62}$$

where $\{|j\rangle\}$ and $\{|s\rangle\}$ are orthonormal bases of the system and ancilla spaces, respectively. By expanding the arbitrary unitary of Eq. (2.58) in the ancilla basis, with matrix elements $V_{st} = \langle s| \, V \, |t\rangle$, we have that

$$\begin{aligned} F(\rho(\boldsymbol{\theta}), \rho(\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta})) &= \max_V \left| \sum_{st} V_{st} \sum_j c_{js}^*(\boldsymbol{\theta}) c_{jt}(\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta}) \right| \\ &\equiv \max_V \left| \mathrm{Tr}\Big(V^T M\Big) \right| = \|M\|_1, \end{aligned} \tag{2.63}$$

where

$$M_{st} = \sum_j c_{js}^*(\boldsymbol{\theta}) c_{jt}(\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta}) \tag{2.64}$$

is the *quantum metrology matrix* [51, 138] and the last equality in Eq. (2.63) is a property of the trace norm $\|M\|_1 = \mathrm{Tr}\sqrt{M^\dagger M}$ [137].

3 | Classical optical imaging theory

In this chapter we review some aspects of the classical theory of optical imaging, with emphasis on fully coherent and incoherent (thermal) sources. Our goal is to provide the conceptual framework behind the rest of the thesis, introducing important notions such as point-spread function, numerical aperture, spatial coherence and correlations.

The theory of this chapter is rewritten in quantum parlance in Appendix D, which provides the quantitative framework of our work. The quantum theory of imaging was moved out of the main body of the thesis since it is rather cumbersome and only its main results will be used.

3.1 Imaging systems

3.1.1 Fourier optics

Consider a quasi-monochromatic, paraxial and scalar light field in the configuration of Fig. 3.1, where it propagates along the z direction from the $\mathcal{A}_0$ to the $\mathcal{A}_z$ apertures. This general case encapsulates a plethora of imaging scenarios. For example, light from an object plane being captured by a lens can be modelled with $\mathcal{A}_0 = \mathbb{R}^2$ and $\mathcal{A}_z$ a disk function centred at the origin, representing the lens' circular aperture.

If the wavelengths are much smaller than the relevant physical dimensions, then we are in the *near field* regime and the Huygens-Fresnel diffraction integral [139,

140] can be used to relate the fields at the aperture and observation planes,

$$E_z(\vec{r},t) = \frac{e^{i\omega z}}{i\lambda z}\int_{\mathcal{A}_0} \mathrm{d}\vec{r'}\; e^{i\pi|\vec{r}-\vec{r'}|^2/\lambda z}\; E_0(\vec{r'},t-z/c), \tag{3.1}$$

where ω is the field's central frequency, $\lambda = 2\pi c/\omega$ the corresponding wavelength and $\vec{r}$ and $\vec{r'}$ transverse vectors in the aperture and observation planes, respectively. In what follows we will ignore global phases and overall magnitudes, so that Eq. (3.1) can be simplified to

$$E_z(\vec{r},t) \propto \int \mathrm{d}\vec{r'}\; e^{i\pi|\vec{r}-\vec{r'}|^2/\lambda z}\mathcal{A}_0(\vec{r'})E_0(\vec{r'},t-z/c), \tag{3.2}$$

where the integral is now over the entire input plane.

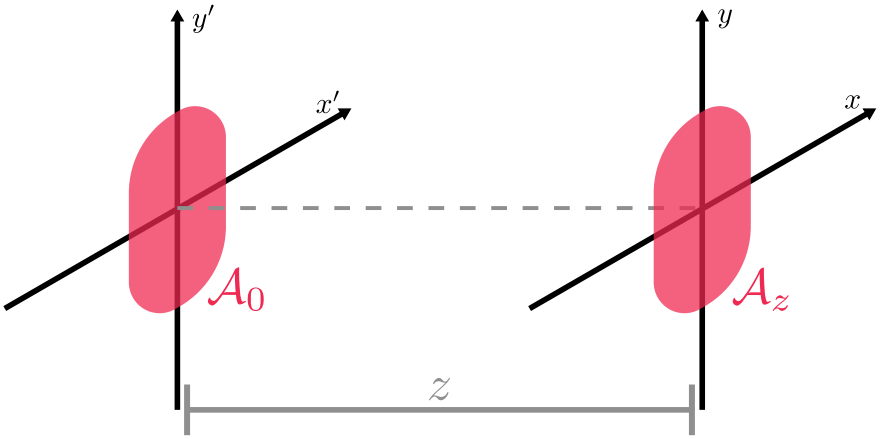


Figure 3.1: Propagation of light from aperture $\mathcal{A}_0$ to aperture $\mathcal{A}_z$ along the z direction.

If the distance to the observation plane is much larger than the aperture $\mathcal{A}_0$ width W,

$$W/\lambda z \ll 1, \tag{3.3}$$

then we are in the *Fraunhofer* or *far field* regime, and we can use the approximation [141]

$$e^{i\pi|\vec{r}-\vec{r'}|^2/\lambda z} \sim e^{-i2\pi\vec{r}\cdot\vec{r'}/\lambda z}. \tag{3.4}$$

Eq. (3.2) can be further simplified to

$$\begin{aligned} E_z(\vec{r},t) &\propto \int \mathrm{d}\vec{r'}\; e^{-i2\pi\vec{r}\cdot\vec{r'}/\lambda z}\mathcal{A}_0(\vec{r'})E_0(\vec{r'},t-z/c) \\ &= \mathcal{F}\left[\mathcal{A}_0(\vec{r'})E_0(\vec{r'},t-z/c)\right]\left(\frac{\vec{r}}{\lambda z}\right), \end{aligned} \tag{3.5}$$

where

$$\mathcal{F}\left[f(\vec{r'})\right](\vec{r}) = \int \mathrm{d}\vec{r'} e^{-i2\pi\vec{r}\cdot\vec{r'}} f(\vec{r'}) \tag{3.6}$$

is the two-dimensional Fourier transform. Alternatively, if a parabolic lens is placed in the aperture plane, then we end up with an extra quadratic phase term [141]

$$E_z(\vec{r},t) \propto \int \mathrm{d}\vec{r'}\ e^{i\pi|\vec{r}-\vec{r'}|^2/\lambda z} e^{-i\pi|\vec{r'}|^2/\lambda f} \mathcal{A}_0(\vec{r'}) E_0(\vec{r'}, t - z/c), \tag{3.7}$$

where f is the focal length of the lens. If the observation plane sits at the focal plane of the lens, i.e., $z = f$, we end up with Eq. (3.5) as well. Therefore, the effect of the lens is, in this case, identical to far field propagation.

3.1.2 Point-spread function

Now consider the imaging system depicted in Fig. 3.2. Despite its apparent simplicity, it can describe almost any passive linear imaging system operating on the far field. It consists of an aperture $\mathcal{A}$ in the far field of the object to be imaged and a lens that focuses its image at a finite distance after the aperture, and the whole setup corresponds to a sequence of two imaging scenarios as in Fig. 3.1. The field at the image plane is related to the one at the object plane via

$$E_I(\vec{r},t) = \mathcal{F}\left[\mathcal{A}(\vec{r_A})\ \mathcal{F}\left[E_O(\vec{r_O}, t - (z+f)/c)\right]\left(\frac{\vec{r_A}}{\lambda z}\right)\right]\left(\frac{\vec{r}}{\lambda f}\right), \tag{3.8}$$

where $\mathcal{A}$ is the lens aperture function and $\vec{r_A}$ the coordinates at that plane, as depicted in Fig. 3.2. If we rescale the aperture and image plane coordinates by

$$\left(\frac{\vec{r_A}}{\lambda z}\right) \to \vec{r_A},\ \left(\frac{\vec{r}}{\lambda f}\right) \to \vec{r}, \tag{3.9}$$

Eq. 3.8 becomes

$$E_I(\vec{r},t) = \psi(\vec{r}) * E_O(\vec{r},t'), \tag{3.10}$$

where $t' = t - (z+f)/c$, $*$ denotes the deconvolution operation and

$$\psi(\vec{r}) = \mathcal{F}\left[\mathcal{A}(\vec{r_A})\right](\vec{r}) \tag{3.11}$$

is the *point-spread function* (PSF) of the imaging system. It is commonly assumed that the PSF is L^2 normalized, i.e.,

$$\int \mathrm{d}\vec{r} |\psi(\vec{r})|^2 = 1. \tag{3.12}$$

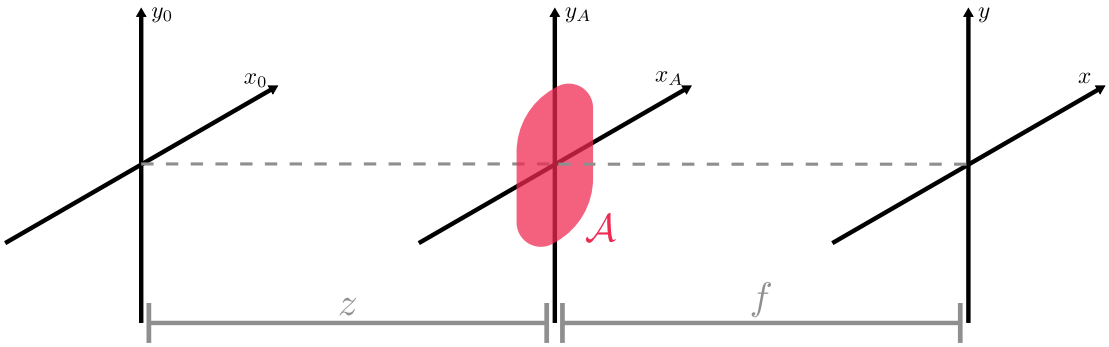


Figure 3.2: Far field propation of light and subsequent imaging by a lens of aperture $\mathcal{A}$.

3.1.3 Spatial correlations

The ability of a light wave to interfere depends on how its points are temporally and spatially correlated to one another. In this work, as the title may have hinted, we are interested in the spatial properties of light. Thus, we will always assume the light fields we are working with in a certain measurement are temporally coherent, i.e., they are inside one *coherence length* interval [6, 142], and only worry about their (transversal) spatial behaviour.

The (first order) *spatial coherence* or *mutual intensity* function of a field at points $\vec{r}$ and $\vec{r'}$ is [141]

$$\Gamma(\vec{r},\vec{r'}) = \left\langle E(\vec{r},t)E^*(\vec{r'},t)\right\rangle, \tag{3.13}$$

where the angular brackets denote a temporal average over multiple coherence length intervals. Due to ergodicity [142], the time-dependence of the right-hand side of Eq. (3.13) doesn't affect the left hand side, and could in principle be dropped. The spatial coherence, when normalized, becomes the *degree of spatial coherence*

$$\mu(\vec{r},\vec{r'}) = \frac{\left\langle E(\vec{r},t)E^*(\vec{r'},t)\right\rangle}{\sqrt{\langle|E(\vec{r},t)|^2\rangle\left\langle|E(\vec{r'},t)|^2\right\rangle}}. \tag{3.14}$$

When $|\mu(\vec{r},\vec{r'})| = 1$, the fields at these two points can undergo full interference when combined, while when $\mu(\vec{r},\vec{r'}) = 0$ they are completely incoherent and will not interfere at all. Photodetectors and cameras measure the "diagonal" elements of the spatial coherence, the average intensity

$$\mathcal{I}(\vec{r}) \equiv \mu(\vec{r},\vec{r}) = \left\langle|E(\vec{r},t)|^2\right\rangle. \tag{3.15}$$

In order to measure the "off-diagonal" terms, the light fields must be interfered in some way prior to (intensity) detection. The spatial coherences at image and object planes can be easily related by Eqs. (3.10) and (3.13):

$$\Gamma_I(\vec{r},\vec{r'}) = \iint \mathrm{d}\vec{\rho}\mathrm{d}\vec{\rho'}\ \Gamma_O(\vec{\rho},\vec{\rho'})\psi(\vec{\rho}-\vec{r})\psi^*(\vec{\rho'}-\vec{r'}). \tag{3.16}$$

3.2 Incoherent light

Most optical sources in nature emit light spontaneously from their excited atoms [142]. This highly random light is usually called *thermal*, due to its most common origin. In terms of its statistical properties, it obeys

$$\mu(\vec{r},\vec{r'}) = \delta^2(\vec{r}-\vec{r'}), \tag{3.17}$$

and for this reason it is also called (first order) spatially *incoherent* light. This means that the sources at the object plane are uncorrelated with one another and their phase relationships are random [142]. The spatial coherence can be written as $\Gamma_O(\vec{\rho},\vec{\rho'}) = \mathcal{I}_O(\vec{r})\delta^2(\vec{r}-\vec{r'})$, where $\mathcal{I}_O(\vec{r})$ is the average intensity at the object plane. As light propagates to the image plane, the spatial coherence, as per Eq. (3.16), becomes

$$\Gamma_I(\vec{r},\vec{r'}) = \int \mathrm{d}\vec{\rho}\ \mathcal{I}_O(\vec{\rho})\psi(\vec{\rho}-\vec{r})\psi^*(\vec{\rho}-\vec{r'}), \tag{3.18}$$

and, in particular, the image plane intensity is

$$\mathcal{I}_I(\vec{r}) = \mathcal{I}_O(\vec{r}) * |\psi(\vec{r})|^2. \tag{3.19}$$

Eq. (3.19) shows the importance of the PSF for incoherent light imaging. It linearly determines the image plane intensity from the object plane one, and has the effect of blurring, smoothing or low-pass filtering the original distribution. It also corresponds to the image of a point source.

The most common aperture for an imaging system is a circular one,

$$\mathcal{A}(r,\phi) = \frac{2}{\sqrt{\pi D^2}}\mathrm{circ}\left(\frac{2r}{D}\right), \tag{3.20}$$

where D is its diameter and $\mathrm{circ}(r)$ is unity for $r \leq 1$ and zero elsewhere. For convenience we assume it is L^2 normalized,

$$\int \mathrm{d}\vec{r}|\mathcal{A}(\vec{r})|^2 = 1. \tag{3.21}$$

The corresponding PSF is

$$\psi(r,\phi) = \frac{1}{\sqrt{\pi}}\frac{J_1(\pi D r)}{r}, \tag{3.22}$$

where $J_n(r)$ is the n-th order Bessel function of the first kind. Eq. (3.22) is known as the *Airy function* PSF and its modulus squared can be seen in Fig. 3.3. It is characterized by a central lobe, followed by smaller radial side lobes.

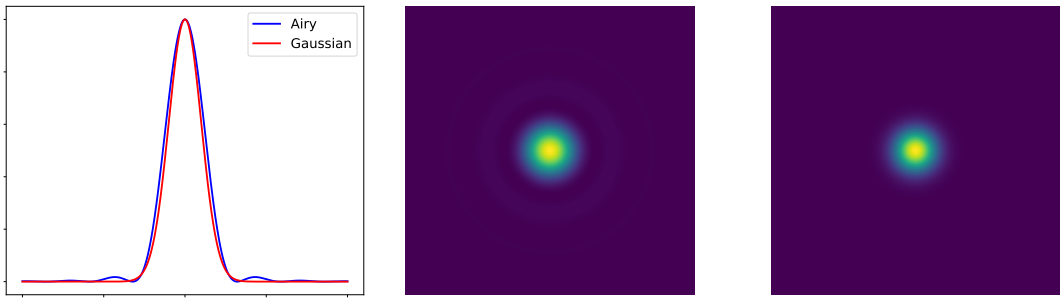


Figure 3.3: Profile of the normalized squared Airy (solid blue curve) and Gaussian (solid red curve) point-spread functions (left); 2D plot of the squared Airy (middle) and Gaussian (right) PSFs.

Now consider two equal (unit) intensity point sources separated by a distance θ[1] . For simplicity they can be seen as lying along the x direction (so the problem becomes effectively one-dimensional) and centred around the origin. Then, the object plane intensity distribution is

$$\mathcal{I}_O(x) = \delta(x-\theta/2) + \delta(x+\theta/2), \tag{3.23}$$

and the image plane one is

$$\mathcal{I}_I(x) = |\psi(x-\theta/2)|^2 + |\psi(x+\theta/2)|^2, \tag{3.24}$$

[1]We use the greek letter θ in order to maintain consistency with the parameter estimation notation. In the imaging literature, however, the angular separation between the sources is also denoted as θ. It is important to remember that we are talking about linear, and not angular separation.

which is plotted in Fig. 3.4. As the separation θ becomes smaller, the two peaks at the image plane become closer and harder to distinguish. At which point can we no longer tell them apart? Lord Rayleigh heuristically defined the smallest resolvable separation as the one for which the central maximum of one source's Airy disk coincides with the other sources' Airy disk's first minimum [7]. Since the first zero of $J_1(x)$ occurs at $x \approx 3.8317$, and keeping track of the rescaling of Eq. (3.9), we find that the smallest resolvable source separation is

$$\theta_R \approx 1.22 \frac{\lambda z}{D}, \tag{3.25}$$

the celebrated *Rayleigh limit* [3, 7].

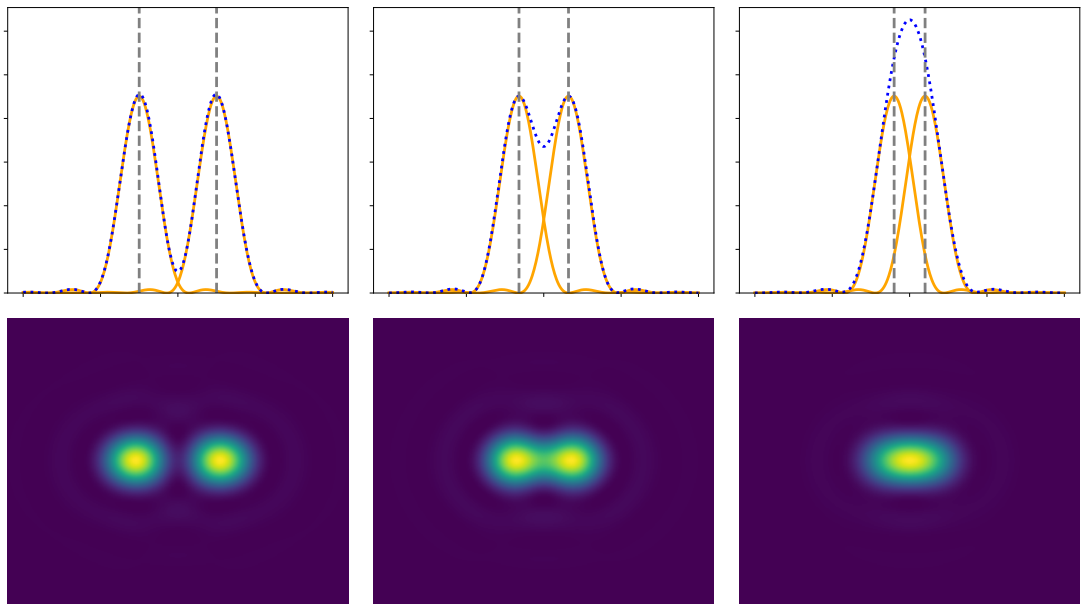


Figure 3.4: Profile (upper column) and 2D distribution (lower column) of the image plane intensity of two incoherent sources which are well resolved (left), just resolved (middle) and not resolved (right).

In microscopy (and fiber optics), the *numerical aperture*[2] is the range of angles of light that the system can receive. It is defined for circular apertures, such as the one seen in Fig. 3.2, and it is simply

$$\mathrm{NA} = n \sin \alpha \approx nD/2z, \tag{3.26}$$

[2]Although the NA is not commonly used for telescopes, in this text it will be generally used when we refer to the resolution of imaging systems.

where n is the index of refraction of the medium. The maximum achievable NA, due to geometrical considerations, is around $0.95n$, and state-of-the-art optical microscopes have an NA of 1.4, where the lens is immersed in an oil with higher refraction index than air [7]. For this thesis we will consider either vacuum ($n = 1$) or air ($n \approx 1$) propagation media, and therefore n will be set to unity from now on. In this case, the Rayleigh limit can be restated as

$$\theta_R \approx 1.22\frac{\lambda}{2\mathrm{NA}}, \tag{3.27}$$

and the Airy PSF can also be rewritten as

$$\psi(r,\phi) = \frac{1}{\sqrt{\pi}}\frac{J_1(2\pi\mathrm{NA}r/\lambda)}{r} = \frac{1}{\sqrt{\pi}}\frac{J_1(k\mathrm{NA}r)}{r}, \tag{3.28}$$

where $k = 2\pi/\lambda$ is the corresponding wavevector. Considering the coordinate rescaling, we can express the aperture diameter as $D = 2\mathrm{NA}/\lambda$.

For theoretical studies, sometimes we consider the simpler case where the plane transverse to propagation is one dimensional. In this case, the circular aperture becomes a top-hat aperture

$$\mathcal{A}(x) = \frac{1}{\sqrt{D}}\,\mathrm{rect}\left(\frac{x}{D}\right), \tag{3.29}$$

where D is again the diameter of the aperture and $\mathrm{rect}(x)$ is unity for $|x| \leq 1/2$ and zero elsewhere. The corresponding PSF is

$$\psi(x) = \sqrt{D}\,\mathrm{sinc}(Dx) = \sqrt{D}\,\mathrm{sinc}\left(\frac{2}{\pi}k\mathrm{NA}x\right), \tag{3.30}$$

where $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$ is the normalised sinc function.

If the Airy PSF is apodized, i.e., its sidelobes are blocked (for example, with an iris), the remaining central peak can be very well approximated by a Gaussian function

$$\psi(\vec{r}) = \frac{1}{(2\pi\sigma^2)^{1/2}}e^{-|\vec{r}|^2/4\sigma^2}, \tag{3.31}$$

where $\sigma = 0.42\lambda/2\mathrm{NA}$ [143], as can be seen in Fig. 3.3. This is the so-called *soft aperture* approximation, since the aperture that gives rise to this PSF, which is also Gaussian (with $\sigma' = 1/4\pi\sigma$), is a "smoothed" version of the *hard aperture*

of Eq. (3.20). Gaussian PSFs are widely used in the literature due to the simple analytical expressions they give rise to. It is important to remember that numerical aperture is only defined for circular apertures, and its use with other aperture shapes is an slight extrapolation of its original meaning.

Finally, note from Eq. (3.18) that the spatial coherence function at the image plane $\Gamma_I(\vec{r}, \vec{r'})$ is nonzero for $\vec{r} \neq \vec{r'}$. This is (part of) the *Zernike-van Cittert theorem* [144, 145]: even completely incoherent light ends up picking up some partial coherence as it propagates. This phenomenon is a double edged sword. On the one hand, it means that sources at the image plane become less distinguishable, which is bad for resolution. This, however, we already know. On the other hand, it means that there is some information in the image plane spatial correlations, which might be useful for estimating some parameters of the original distribution. Since intensity measurements always discard that information, we might have been missing something all this time...

3.3 Coherent light

While thermal light is based on spontaneous emission from atoms, laser light depends on the occurrence of *stimulated emission* from the laser medium [142]. This kind of radiation is highly ordered and has exquisite coherence properties: its spatial coherence (Eq. (3.14))

$$|\mu(\vec{r}, \vec{r'})| = 1, \tag{3.32}$$

and it is said to be fully spatially *coherent*. This means that it has a well-defined phase along all its extent, and that any two points undergo complete interference. Its spatial coherence can be written as $\Gamma_O(\vec{\rho}, \vec{\rho'}) = E(\vec{\rho})E^*(\vec{\rho'})$, where $E(\vec{\rho})$ is the field's complex amplitude [142].

The spatial coherence propagates as

$$\Gamma_I(\vec{r}, \vec{r'}) = (E * \psi)(\vec{r})(E * \psi)^*(\vec{r'}), \tag{3.33}$$

and the image plane intensity is simply

$$\mathcal{I}_I(\vec{r}) = |(E * \psi)(\vec{r})|^2. \tag{3.34}$$

As is expected, coherent light remains coherent as it propagates. Coherent light has a well-defined amplitude, which means that the spatial coherence, which is a mere function of it, becomes a redundant concept. All the information is contained in the field's complex amplitude, which can be fully recovered via a phase resolved measurement such as digital holography or heterodyne detection.

Now consider again two equally strong sources separated by θ. As light is now coherent, the sources have a well-defined relative phase Φ between them. The object plane amplitude distribution can be written as

$$E(x, \Phi) = \delta(x - \theta/2) + e^{i\Phi}\delta(x + \theta/2). \tag{3.35}$$

For simplicity, consider a real-valued point-spread function. Then, the image plane intensity is

$$\mathcal{I}_I(x, \Phi) = |\psi(x - \theta/2)|^2 + |\psi(x + \theta/2)|^2 + 2\cos\Phi\ \psi(x - \theta/2)\psi(x - \theta/2). \tag{3.36}$$

The relative phase, which controls the interference term in the right hand side of Eq. (3.36), is determinant for the distinguishability of the two sources, as illustrated in Fig. 3.5. If $\Phi = 0$, then

$$\mathcal{I}_I(x, 0) = |\psi(x - \theta/2) + \psi(x + \theta/2)|^2. \tag{3.37}$$

The sources are less distinguishable than the incoherent case, as there is some coherence between them, and, by using the same Rayleigh criterion, the resolution limit is now

$$\theta_R \approx 1.64\frac{\lambda}{2\mathrm{NA}} \tag{3.38}$$

for circular apertures. If $\Phi = \pi/2$, then

$$\mathcal{I}_I(x, \pi/2) = |\psi(x - \theta/2)|^2 + |\psi(x + \theta/2)|^2, \tag{3.39}$$

and the intensity distribution is identical to the incoherent one [141]. Finally, if the sources are anti-correlated, $\Phi = \pi$, then

$$\mathcal{I}_I(x, \pi) = |\psi(x - \theta/2) - \psi(x + \theta/2)|^2, \tag{3.40}$$

and the anti-correlation makes them more distinguishable than if they were incoherent.

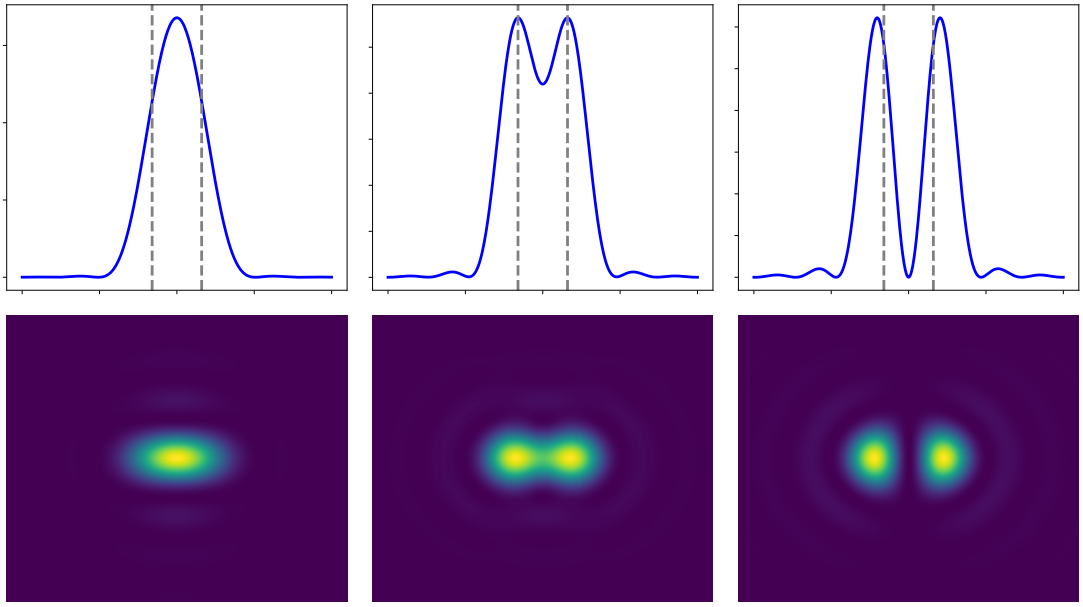


Figure 3.5: Profile (upper column) and 2D distribution (lower column) of the image plane intensity of two coherent sources at the (incoherent) Rayleigh limit, for a relative phase of $\phi = 0$ (left), $\phi = \pi/2$ (middle) and $\phi = \pi$ (right).

3.4 Measurements

The usual way of measuring the image plane light is via direct intensity detection, which can be done with cameras or spatially-resolved photodetectors. They return the average intensity at the image plane, $\Gamma_I(\vec{r}, \vec{r})$, at every position $\vec{r}$, and are depicted in Fig. 3.6. This method is ideal for incoherent light, as their intensity is the only quantity of interest. For coherent light, the measured intensity will display interference effects, and it might be hard to recover the underlying complex amplitude distribution directly from that.

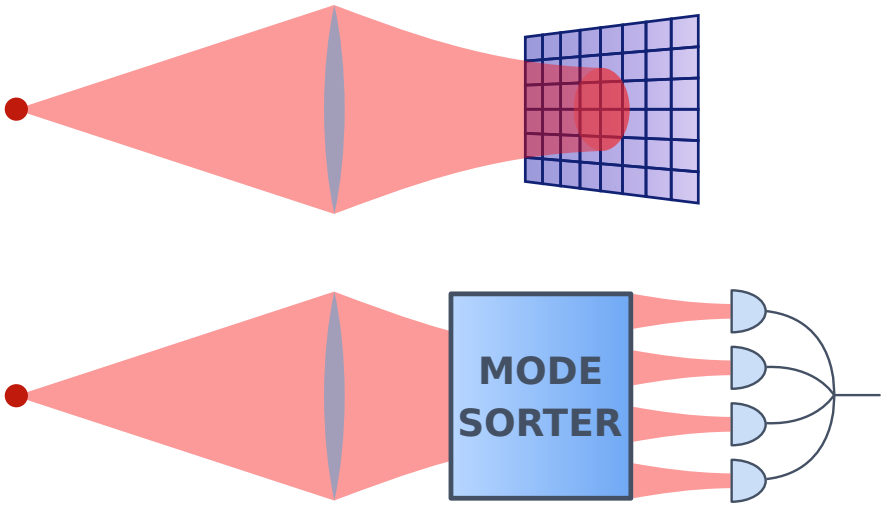


Figure 3.6: depiction of direct imaging (top) and mode sorting (bottom) measurements.

An alternative to direct imaging is to perform a *spatial mode sorting* or *spatial demultiplexing* of the image plane field prior to intensity detection, as schematised in Fig. 3.6. We start by choosing an orthonormal basis $\{\phi_m(\vec{r})\}$ of functions in the image plane ($\mathbb{R}^2$). Orthonormality means that

$$\int \mathrm{d}\vec{r}\phi_m^*(\vec{r})\ \phi_{m'}^*(\vec{r}) = \delta_{mm'}, \tag{3.41}$$

and as a basis it must be complete [146],

$$\sum_{m=0}^{\infty} \phi_m(\vec{r})\phi_m^*(\vec{r'}) = \delta^2(\vec{r}-\vec{r'}). \tag{3.42}$$

Mode sorting consists of decomposing the image plane light into that basis of modes and redirecting each of these components to a different intensity detector. The detector for mode m will return a power proportional to

$$P_m = \iint \mathrm{d}\vec{r}\ \mathrm{d}\vec{r'}\phi_m^*(\vec{r})\Gamma_I(\vec{r},\vec{r'})\phi_m(\vec{r'}). \tag{3.43}$$

Using Eq. (3.16), this can be rewritten as

$$P_m = \iint \mathrm{d}\vec{\rho}\ \mathrm{d}\vec{\rho'}R_m(\vec{\rho})\Gamma_O(\vec{\rho},\vec{\rho'})R_m^*(\vec{\rho'}), \tag{3.44}$$

where

$$R_m(\vec{\rho}) = \int \mathrm{d}\vec{r}\phi_m^*(\vec{r})\psi(\vec{\rho}-\vec{r}) \tag{3.45}$$

is the overlap between each mode and the PSF. For coherent light, we have that

$$P_m = |E(\vec{\rho}) * R_m(\vec{\rho})|^2. \tag{3.46}$$

Although we are still measuring intensities, for a suitable choice of mode basis it might be possible to recover the phase information of light. For incoherent light,

$$P_m = \mathcal{I}_O(\vec{\rho}) * |R_m(\vec{\rho})|^2. \tag{3.47}$$

This time, with a suitable basis we might recover the off-diagonal spacial coherence terms

Another alternative measurement is to perform homodyne or heterodyne detection of the image plane light field [25, 147]. They consist of mixing the image field (called here *signal field*) with a *local oscillator* (LO) field in a beamsplitter (BS), perfoming photodetection of both BS outputs and measuring the difference of the detected photocurrents, as depicted in Fig. 3.7. If the LO is in a sufficiently strong coherent state and it's frequency is equal to ("homodyne") or different ("heterodyne") by a fixed offset than the signal's, then by varying the LO's phase we can recover the light's phase information. This makes it specially useful for coherent light imaging.

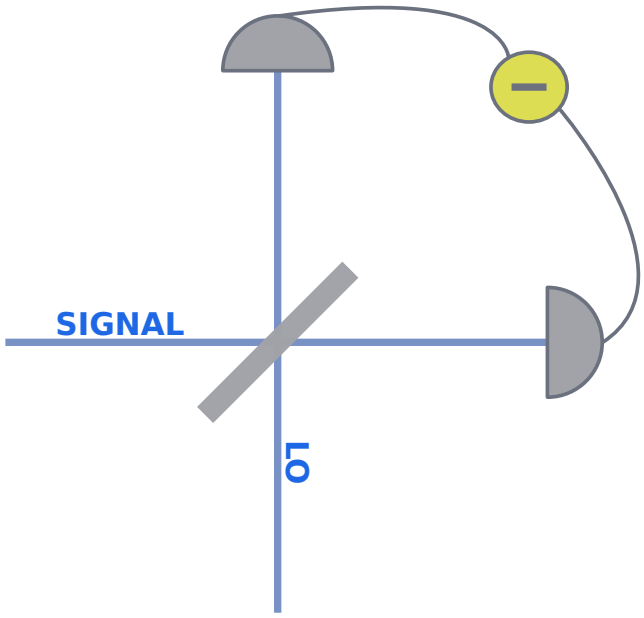


Figure 3.7: Scheme of homodyne and heterodyne detection.

Furthermore, since the photocurrents are proportional to the overlap between LO and signal modes, "dyne" detection also "unintentionally" performs mode sorting [147]. If we sequentially prepare each mode in a mode basis as the LO and do dyne detection with it, we will at the end have done mode sorting. However, since at each measurement we discard the light from the components orthogonal to the current LO, there is a higher photon overhead associated with this measurement.

4 | Statistical analysis of incoherent light imaging

In this chapter we use the statistical tools showcased in Chapter 2 in order to study the far-field imaging of incoherent light sources. The quantum framework underlying the mathematical derivations of the chapter is developed in Appendix D.

In Section 4.1 we review the theory of quantum-inspired super-resolution for the case of imaging two point sources. This scenario is the most studied in the literature, and most of the ideas in QISR are introduced here. In Section 4.2 we review the attempts to extend the theory to full imaging by geometric moments estimation. In this section, we only briefly expose the main results, since its goal is simply to be a bridge between the previous and next sections. Finally, in Section 4.3 we report our results [108] on the theory of full imaging of point sources with QISR.

4.1 Two-point resolution

One of the simplest imaging scenarios consists in estimating the separation θ of two equal-intensity sources. Despite its simplicity, many insightful conclusions can be drawn from this case, as already seen in Chapter 3.

We further consider that the sources lie on an one dimensional axis and are centred around that axis' origin, as in Fig. 4.1. While these two conditions appear to entail no loss of generality, in a geometrical sense, they imply that there is *prior information*, i.e., that the axis of separation and the centroid location are known to us, and the only parameter left to estimate is the sources' separation.

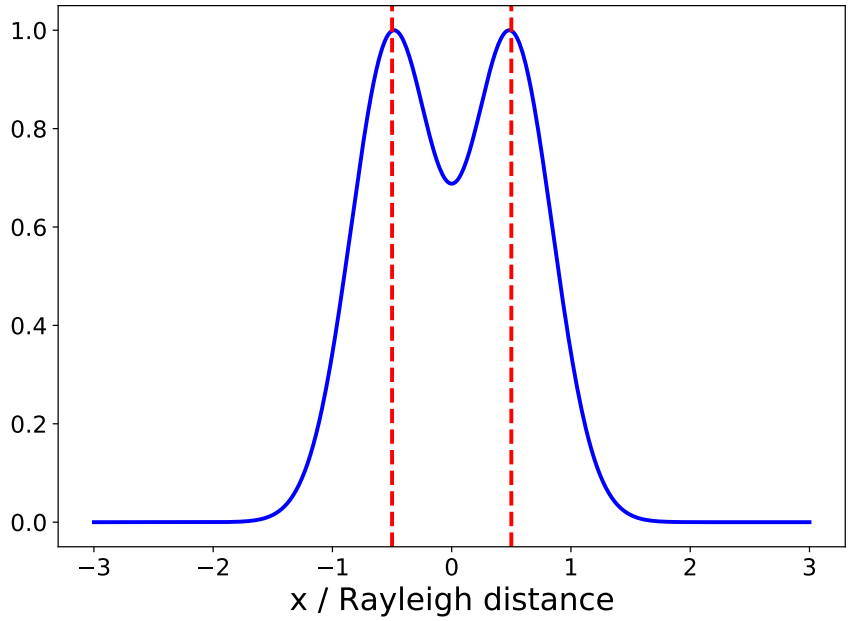


Figure 4.1: Position of the two sources (dashed red lines) and their (normalised) incoherent image plane intensity (solid blue curve) along the one dimensional axis, in units of Rayleigh distance, for a Gaussian point-spread function and unit sources' separation.

4.1.1 Rayleigh's curse

At optical frequencies, the average photon number of most thermal sources is very small [112, 142]. For example, a typical fluorophore emits less than 10^7 photons per coherence time (< 50 fs), which leads to less than 10^6 photons for a single source [6]. In line with the literature on QISR [6], we call this kind of light *weak incoherent light*, and its properties are developed in Appendix E.

In this section we consider weak incoherent light sources in the object place. Their resulting quantum state of light at the image plane can be well approximated by (Proposition 5)

$$\rho \approx (1-\epsilon)\rho_0 + \epsilon\rho_1, \tag{4.1}$$

where ρ_n is a n-photon Fock state and ϵ the average photon number at the image plane (for weak incoherent sources, $\epsilon \ll 1$). Since the vacuum state ρ_0 carries no information, we focus on the one photon state [6]

$$\rho_1 = \frac{1}{2}\left(|\psi_+\rangle\langle\psi_+| + |\psi_-\rangle\langle\psi_-|\right), \tag{4.2}$$

where

$$|\psi_{\pm}\rangle = \int_{-\infty}^{\infty} \mathrm{d}x\ \psi(x \pm \theta/2)\,|x\rangle \tag{4.3}$$

are the quantum states of each source. We perform individual measurements over M copies of ρ_1, using a total of $N = M\epsilon$ photons.

The direct imaging probabilities are (Eq. (E.28))

$$\Pr(x|\theta) = \frac{1}{2}\left(|\psi(x+\theta/2)|^2 + |\psi(x-\theta/2)|^2\right), \tag{4.4}$$

and the associated Fisher information is (Eq. (2.11))

$$\mathcal{F}(\theta) = N\int_{-\infty}^{+\infty} \mathrm{d}x\ \left(\frac{\partial}{\partial\theta}\ln\Pr(x|\theta)\right)^2 \Pr(x|\theta), \tag{4.5}$$

whose integral can be easily evaluated numerically. The per-photon FI $\mathcal{F}(\theta)/N$ is plotted in Fig. (4.2) for Gaussian and top-hat apertures (one dimensional version of the circular aperture). As we enter the subdiffraction limit (separations smaller than the Rayleigh limit), the FI starts deteriorating very fast: this phenomenon has been recently dubbed *Rayleigh's curse* [6], and it's the statistical manifestation of the diffraction limit. Correspondingly, the Cramér-Rao bound tends to infinity: the variance in the (unbiased) estimation of very small separations blows up.

The quantum Fisher information can be quickly evaluated via the geometrical method introduced in Section 2.3.2 . A purification of Eq. (4.2) is

$$|\psi(\theta)\rangle = \frac{1}{\sqrt{2}}\sum_{i=+,-}|\psi_i, i\rangle = \frac{1}{\sqrt{2}}\sum_{i=+,-}\int_{-\infty}^{+\infty}\mathrm{d}x\ \psi(x+i\theta/2)\,|x,i\rangle\,, \tag{4.6}$$

so the quantum metrology matrix is

$$M(\theta_1,\theta_2)) = \frac{1}{2}\begin{bmatrix}\Delta((\theta_1-\theta_2)/2) & \Delta((\theta_1+\theta_2)/2)\\ \Delta^*((\theta_1+\theta_2)/2) & \Delta^*((\theta_1-\theta_2)/2)\end{bmatrix}. \tag{4.7}$$

Its trace norm is [51] $\|M(\theta_1,\theta_2))\|_1 = \Delta((\theta_1-\theta_2)/2)$, so the QFI is given by(Eq. (2.61))

$$\begin{aligned}\mathcal{K}(\theta) &= -4\int_{-\infty}^{+\infty}\mathrm{d}x\frac{\partial^2\psi^*\left(x-((\theta_1-\theta_2)/2)\right)}{\partial\theta_2^2}\psi(x)\Big|_{\theta_1=\theta_2=\theta}\\ &= -\int_{-\infty}^{+\infty}\mathrm{d}x\frac{\partial^2\psi^*(x)}{\partial x^2}\psi(x)\\ &= \int_{-\infty}^{+\infty}\mathrm{d}x\left|\frac{\partial\psi(x)}{\partial x}\right|^2 \equiv \tau(0),\end{aligned} \tag{4.8}$$

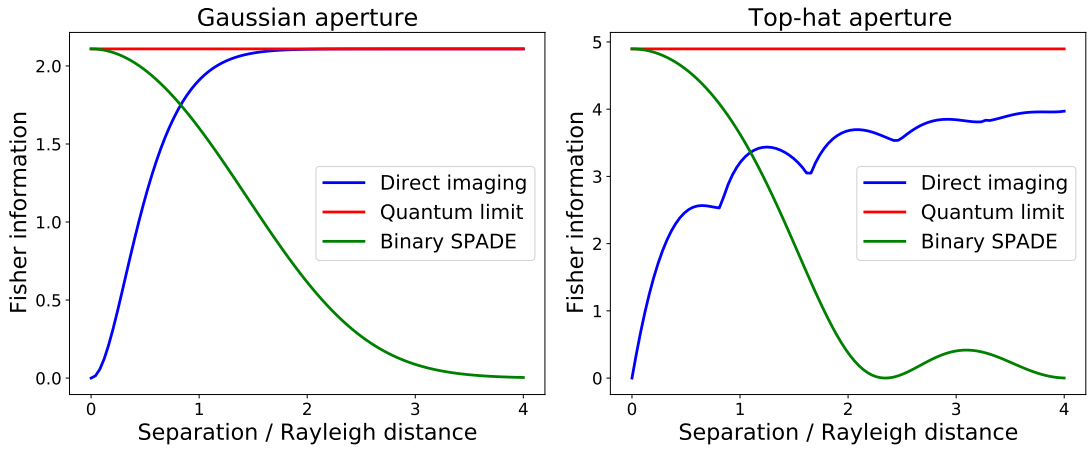


Figure 4.2: Per photon direct imaging (solid blue line) and binary SPADE (solid green line) Fisher information, and quantum Fisher information (solid red line) for the estimation of the separation of two point sources, for Gaussian (left) and top-hat (right) apertures, as a function of the separation. The separation is in units of Rayleigh distance, which is set to unity, so that the widths of the Gaussian and top-hat apertures are equal to $\sigma \approx 0.344$ and $D = 1.22$, respectively.

where we used change of variables and integration by parts. The per-photon QFI $\mathcal{K}(\theta)/N$, which is plotted in Fig. 4.2, is a θ-independent constant. Outside the sub-Rayleigh regime, it is equal to the FI for direct imaging, which means that this measurement is quantum-optimal. Inside the sub-Rayleigh regime, however, it remains constant while the direct imaging FI rapidly decays to zero. This means that it is possible (since the one-parameter QFI is always saturable) to "lift" Rayleigh's curse: there exists a measurement which attains finite variance on the estimation of θ, regardless of the value of the latter.

The term $\tau(0)$ is the spatial-frequency variance of the PSF [6], as can be seen by its alternate form

$$\tau(0) = 4\pi^2 \int_{-\infty}^{\infty} \mathrm{d}k \; k^2 |\mathcal{A}(k)|^2 \equiv 4\pi^2 \langle k^2 \rangle, \tag{4.9}$$

where $\mathcal{A}(u)$ is the aperture function, which is the inverse Fourier transform of the PSF. For Gaussian apertures (see Appendix F),

$$\tau_{\mathrm{Gauss}}(0) = \frac{1}{4\sigma^2} \approx 1.42 \left(\frac{2\mathrm{NA}}{\lambda}\right)^2, \tag{4.10}$$

while for a top-hat aperture (see Appendix G)

$$\tau_{\text{top-hat}}(0) = \frac{\pi^2 D^2}{3} \approx 3.23 \left(\frac{2\text{NA}}{\lambda}\right)^2 . \tag{4.11}$$

In both cases, a wider aperture leads to a larger QFI, as intuitively expected.

4.1.2 Mode sorting

The Cramér-Rao bound tells us which precision can be achieved, but it doesn't tell us how to. Although measuring the system along the eigenbasis of the symmetric logarithmic derivative (SLD, Eq. (2.18)) is known to do the job for the single parameter case (see Section 2.3.1), this measurement is usually unphysical and/or depends on the actual parameter values, making necessary the use of adaptive schemes. Luckily for us, Tsang *et al.* [6] have found a relatively simple scheme to achieve the qCRB.

To start, consider we have a inversion-symmetric aperture [19], i.e., such that $\mathcal{A}(-\vec{r}) = \mathcal{A}^*(\vec{r})$, which is the case for the circular and Gaussian apertures. This implies that the PSF is also inversion-symmetric, $\psi(-\vec{r}) = \psi^*(\vec{r})$. Furthermore, consider that the PSF has a constant $\vec{r}$-independent phase. This feature, which is not restricting since it can be easily implemented by a two-lens system [6, 141], means that the phase is irrelevant and therefore the PSF can be considered real.

The inversion-symmetry and reality of the PSF make the two centred sources imaging scenario highly symmetrical. Therefore, a measurement that can separate the field's symmetric and anti-symmetric parts can achieve high resolution. In order to see that, consider a mode sorting measurement, which has probabilities (Eq. (E.29))

$$\Pr(m|\theta) = \frac{1}{2}\left(|R_m(\theta/2)|^2 + |R_m(-\theta/2)|^2\right), \tag{4.12}$$

where

$$R_m(x) = \int_{-\infty}^{+\infty} \mathrm{d}x' \; \phi_m^*(x')\psi(x'-x) \tag{4.13}$$

are the mode projections. If the modes are real and have a well-defined parity [34], i.e.,

$$\phi_m(-x) = \pm\phi_m(x), \tag{4.14}$$

then the projections also have well-defined parity, $R_m(-x) = \pm R_m(x)$, and the probabilities become $\Pr(m) = |R_m(\theta/2)|^2$, i.e., as if there was simply one source at position $\theta/2$. This means that the task of localising two sources now effectively became the task of finding a single source [34], a procedure which does not suffer from Rayleigh's curse. The Fisher information (Eq. (2.11)),

$$\mathcal{F}(\theta) = N \sum_{m=0}^{\infty} \left(\frac{\partial}{\partial \theta} \Pr(m|\theta) \right)^2 \Pr^{-1}(m|\theta), \tag{4.15}$$

becomes

$$\begin{aligned} \mathcal{F}(\theta) &= 4N \sum_{m=0}^{\infty} \left(\frac{\partial R_m(\theta/2)}{\partial \theta} \right)^2 \\ &= 4N \iint_{-\infty}^{+\infty} \mathrm{d}x \mathrm{d}x' \frac{\partial \psi(x-\theta/2)}{\partial \theta} \frac{\partial \psi(x'-\theta/2)}{\partial \theta} \sum_{m=0}^{\infty} \phi_m(x)\phi_m(x'). \end{aligned} \tag{4.16}$$

Using the completeness of the mode basis leads to

$$\begin{aligned} \mathcal{F}(\theta) &= 4N \int_{-\infty}^{+\infty} \mathrm{d}x \left(\frac{\partial \psi(x-\theta/2)}{\partial \theta} \right)^2 \\ &= N \int_{-\infty}^{+\infty} \mathrm{d}x \left(\frac{\partial \psi(x)}{\partial x} \right) = \mathcal{K}(\theta). \end{aligned} \tag{4.17}$$

Therefore, mode sorting in any basis with well-defined parity is quantum-optimal [34].

SPADE

The first proposal of mode sorting for super-resolution imaging was presented with Hermite-Gaussian (HG) modes, in a scheme dubbed *spatial-mode demultiplexing* (SPADE) [6]. HG modes are given by

$$\phi_m(x) = \frac{1}{\sqrt{2^m m!}} H_m\left(\frac{x}{\sqrt{2}\sigma} \right) \psi(x), \tag{4.18}$$

where $H_m(x)$ are the Hermite polynomials and $\psi(x)$ the Gaussian PSF. Their parity is well-defined as

$$\phi_m(-x) = (-1)^m \phi_m(x), \tag{4.19}$$

which means that odd (even) order modes are odd (even). Some HG modes can be visualised in Fig, 4.3, and their properties are outlined in Appendix F.

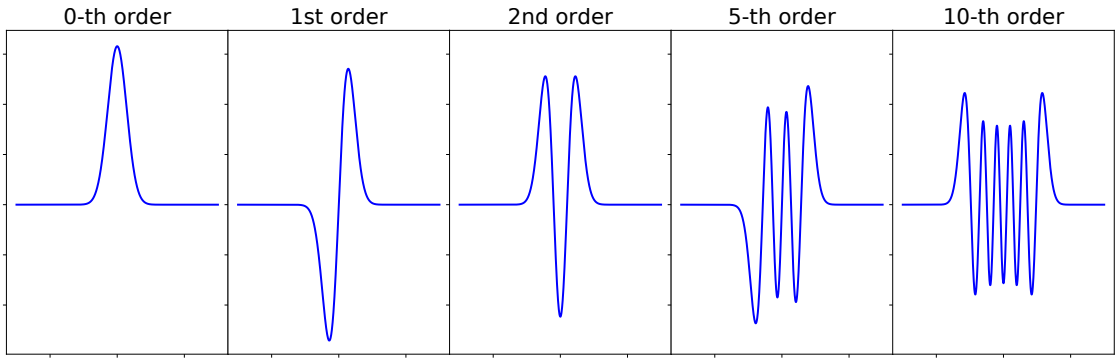


Figure 4.3: From left to right, 0-th, 1st, 2nd, 5-th and 10-th order Hermite-Gaussian modes as in Eq. (4.18).

For a Gaussian PSF, the mode projections are

$$R_m(x) = \frac{1}{\sqrt{m!}} \left(\frac{x}{2\sigma}\right)^m e^{-x^2/8\sigma^2}, \tag{4.20}$$

and the probabilities in the two-source scenario are

$$\Pr(m|\theta) = \frac{1}{m!} \left(\frac{\theta}{4\sigma}\right)^{2m} e^{-\theta^2/16\sigma^2} \equiv \frac{1}{m!} Q^m e^{-Q}, \tag{4.21}$$

where $Q = \theta^2/16\sigma^2$. With the special case of SPADE we can see why mode sorting works. For very small separations, we have $e^{-\theta^2/16\sigma^2} \approx 1$ and the probabilities become $\Pr(m|\theta) \propto (\theta^2)^m$. The (symmetric) 0-th order mode carries no information about the source's separation, while the (anti-symmetric) 1st order measurement is highly sensitive due to its θ^2 dependence. In direct imaging, each measurement is "contaminated" by the 0-th order mode, which just contributes with background noise. In SPADE, however, we can filter out this noisy mode and detect the 1st and higher order modes separately, dramatically increasing the signal-to-noise ratio (SNR) for the estimation procedure. For other mode sorting schemes, the reasoning is the similar.

For very small separations, it is even possible to achieve the qCRB with only two mode projections. Dubbed *binary SPADE* [6], this scheme only discriminates the 0-th order mode, with $\Pr(0|\theta) = e^{-Q}$, and the rest of the modes, for which $\Pr(m > 0|\theta) = 1 - e^{-Q}$. The FI is then

$$\mathcal{F}(\theta) = \frac{N}{4\sigma^2} \frac{Qe^{-Q}}{1 - e^{-Q}}, \tag{4.22}$$

and is plotted in Fig. 4.2. Although the scheme performs poorly for normal separations, it really shines and approaches quantum-optimality in the deep sub-Rayleigh regime.

Binary SPADE can be easily extended to apertures other than Gaussian ones. The idea is that the 0-th order mode of the basis is equal to the PSF, $\phi_0(x) = \psi(x)$, while the other modes can be anything, as long as the orthonormality, completeness and well-defined parity properties hold. In this case, $R_0(x) = \psi^*(x) * \psi(x)$ and the Fisher information is simply

$$\mathcal{F}(\theta) = N \left(\frac{\partial R_0(\theta/2)}{\partial \theta} \right)^2 \frac{4R_0^2(\theta/2)}{1 - R_0^2(\theta/2)}. \tag{4.23}$$

For a top-hat aperture, for example, the FI is

$$\mathcal{F}(\theta) = N(\pi D)^2 \frac{(\cos(Q) - \sin(Q)/Q)^2}{Q^2 - \sin^2(Q)}, \tag{4.24}$$

where $Q = \pi D\theta/2$. It is plotted in Fig. 4.2. The main mechanism of binary SPADE is to filter out the 0-th order mode component, which only brings background noise to the measurement.

PAD bases

All mode bases with well-defined parity are equal, but some are more equal than others. Indeed, while any such basis reaches the qCRB, this happens in the theoretical and unrealistic situation of measuring an infinite number of modes. For a finite number of modes, some bases approach the quantum limit faster, with less modes.

Inspired by SPADE and HG modes, Rehacek *et al.*[34] defined the *point-spread-function-adapted (PAD) bases* [22] which, for a given PSF, are supposed to quickly saturate the qCRB with a small number of modes. They are defined in the aperture plane as

$$\tilde{\phi}_m(\vec{r}_A) = i^m g_m(\vec{r}_A) \mathcal{A}(\vec{r}_A), \tag{4.25}$$

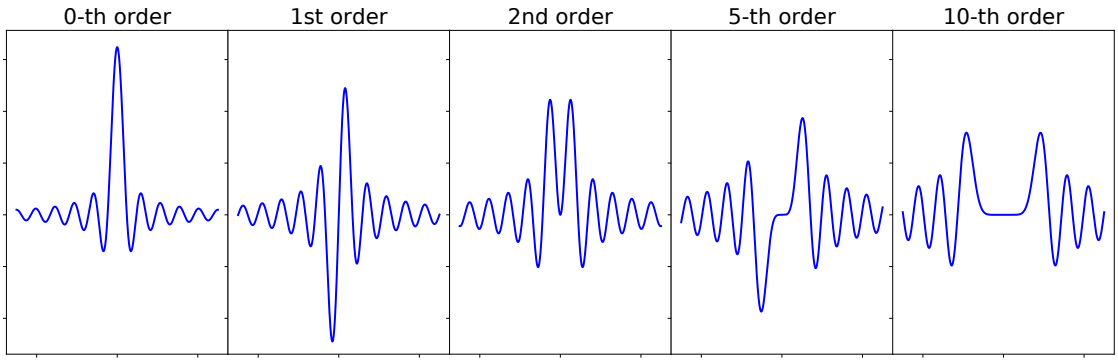


Figure 4.4: From left to right, 0-th, 1st, 2nd, 5-th and 10-th order Bessel modes as in Eq. (4.30).

where $\{g_m(\vec{r}_A)\}$ is a basis of polynomials which are orthonormal with respect to the weight $|\mathcal{A}(\vec{r}_A)|^2$,

$$\int \mathrm{d}\vec{r}_A \; g_m^*(\vec{r}_A) g_n(\vec{r}_A) |\mathcal{A}(\vec{r}_A)|^2 = \delta_{mn}, \tag{4.26}$$

and complete as in

$$\sum_{m=0}^{\infty} \left[g_m(\vec{r}_A)\mathcal{A}(\vec{r}_A)\right]^* \left[g_m(\vec{r'}_A)\mathcal{A}(\vec{r'}_A)\right] = \delta^2(\vec{r} - \vec{r'}). \tag{4.27}$$

Then, the PAD basis is the Fourier transform

$$\phi_m(\vec{r}) = \mathcal{F}\left[\tilde{\phi}_m(\vec{r}_A)\right](\vec{r}). \tag{4.28}$$

The orthonormality and completeness of the PAD basis follow straightforwardly from Eqs. (4.26) and (4.27). Likewise, the symmetry under inversion property of the aperture implies that the $\tilde{\phi}_m(\vec{r}_A)$ and, consequently, the $\phi_m(\vec{r})$, have a well-defined parity as [34]

$$\phi_m(-\vec{r}) = (-1)^m \phi_m(\vec{r}), \tag{4.29}$$

which shows that the basis is quantum-optimal. Finally, the PAD basis is unique (up to reordering of modes) for its corresponding PSF [34].

The HG basis, as in the original proposal of SPADE, is a PAD basis for the Gaussian aperture (see Appendix F). For the top-hat aperture, a possible optimal basis is the Bessel basis (see appendix G) [34]

$$\phi_m(x) = \sqrt{m + 1/2}\frac{J_{n+1/2}(\pi D x)}{\sqrt{x}}, \tag{4.30}$$

where $J_\alpha(x)$ is the α-th order Bessel function of the first kind. A few modes of this basis are shown in Fig. 4.4.

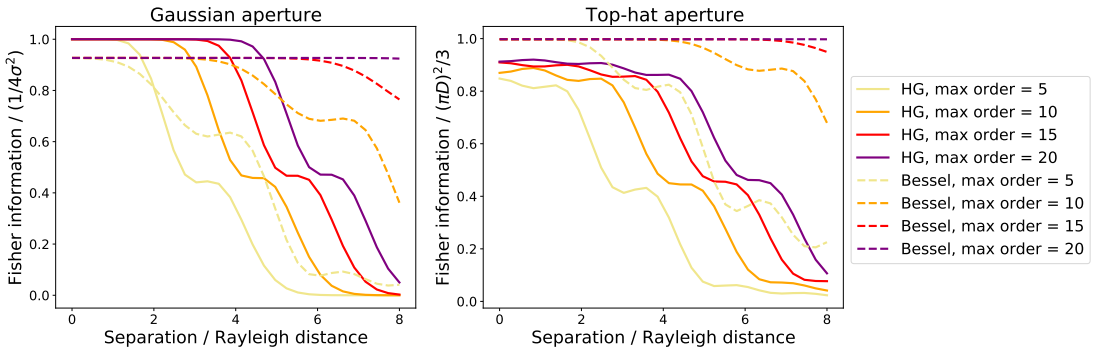


Figure 4.5: Per photon Fisher information for Hermite-Gaussian (solid lines) and Bessel (dashed lines) basis mode sorting measurements, up to 5-th (yellow), 10-th (orange), 15-th (red) and 20-th (purple) order modes, for Gaussian (left) and top-hat (right) apertures, plotted as a function of separation (normalised by Rayleigh distance). We used $\sigma \approx 0.344$ and $D = 1.22$.

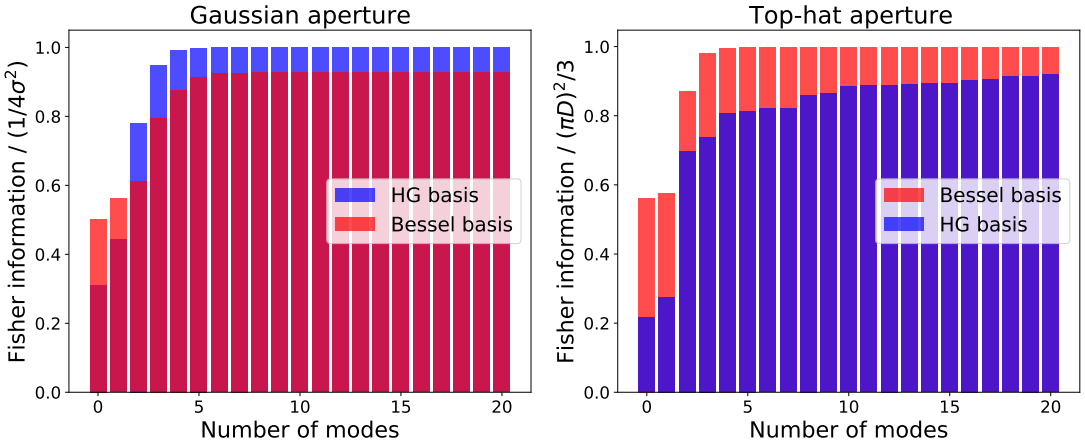


Figure 4.6: Per photon Fisher information for Gaussian (left) and top-hat (right) apertures, normalised by the corresponding quantum Fisher information, as a function of the number of modes in the mode sorting measurement, for Hermite-Gaussian (blue bars) and Bessel (red bars) bases. We used $\sigma \approx 0.344$ and $D = 1.22$.

Each PAD basis is optimal for its intended PSF. In order to see that, consider a Gaussian aperture and both Hermite-Gaussian and Bessel basis measurements. Although both are quantum-optimal for a infinite number of measured components, for a finite number the HG basis is clearly superior, as shown on the left plot of

Fig. 4.5, where the finite-order Fisher informations,

$$\mathcal{F}(M,\theta) = N\sum_{m=0}^{M}\left(\frac{\partial}{\partial\theta}\Pr(m|\theta)\right)^2 \Pr^{-1}(m|\theta), \tag{4.31}$$

are plotted for $M = 5, 10, 15$ and 20. The HG basis FI is given by

$$\mathcal{F}_{\mathrm{HG}}(M,\theta) = \frac{N}{4\sigma^2}\sum_{m=0}^{M}(m-Q)^2\frac{Q^{m-1}}{m!}e^{-Q}, \tag{4.32}$$

where $Q = \theta^2/16\sigma^2$, and the Bessel basis FI is evaluated numerically. Analogously, for a top-hat aperture it's now the Bessel basis which is superior, as can be checked on the right of Fig. 4.5. The HG basis FI is found numerically and the Bessel basis FI is

$$\mathcal{F}_{\mathrm{Bessel}}(M,\theta) = \frac{N\pi^2 D}{(2m+1)\theta}\sum_{m=0}^{M}\left[mJ_{m-1/2}\left(\frac{\pi D\theta}{2}\right) - (m+1)J_{m+3/2}\left(\frac{\pi D\theta}{2}\right)\right]^2. \tag{4.33}$$

The measurements' efficiencies as a function of the number of modes can be seen for a fixed separation, as in Fig. 4.3 where we used $\theta = 1$, i.e., the sources are separated by the Rayleigh distance.

In both Figs. 4.5 and 4.6 the "matched" mode bases (i.e., HG basis for Gaussian aperture and Bessel basis for top-hat aperture) achieve the qCRB with a small number of modes. The "mismatched" bases, however, although quickly reaching around 90% of the qCRB, take a extremely large number of modes in order to become closer to the quantum limit. They converge so slowly that for a small number of modes appear constant, as in the left plots of Figs. 4.5 and 4.6.

4.1.3 Maximum likelihood estimation

The Fisher information analysis so far tells us what precision can be achieved with each measurement. It does not tell us, however, what to do with the measured data in order to achieve that precision, i.e., which *estimators* achieve the CRB. Finding good estimators is a whole research area in itself, but in theoretical studies we can always use the *maximum likelihood estimator* (MLE), Eq. (2.10), which is guaranteed to saturate the Fisher information asymptotically, i.e., in the limit of infinite measurements. In this section we study the MLE for the direct imaging and mode sorting schemes of previous sections.

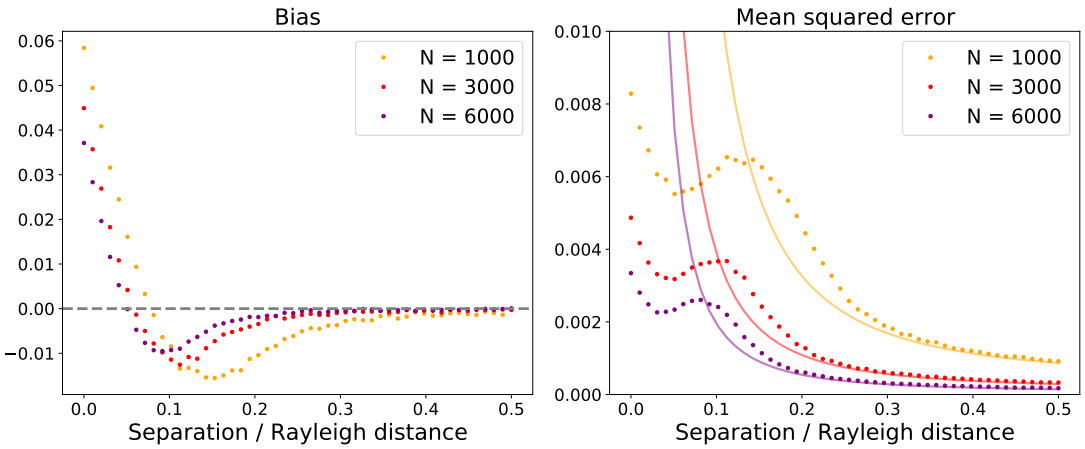


Figure 4.7: Bias (left) and mean squared error (right) of the maximum likelihood estimator for direct imaging as a function of separation for $N = 1000$ (orange dots), $N = 3000$ (red dots) and $N = 6000$ (purple dots) photons. The grey dashed line represents the zero bias reference and the solid lines the corresponding CRBs. We used a Gaussian PSF ($\sigma \approx 0.344$) and the separation is normalised by the Rayleigh distance. The statistics of the MLE for each separation are obtained with 10^5 samples.

In Fig. 4.7 we plot the bias and mean squared error (MSE) of the maximum likelihood estimator for a direct imaging measurement with Gaussian PSF, as a function of the separation θ, for $N = 1000, 3000$ and 6000 photons. Note that the MSE follows the theoretical CRB curve except for very small separations, where it converges to a finite value instead of the predicted divergence to infinity. This happens because the MLE becomes heavily biased in that regime, as can be seen on the left of Fig. 4.7. The MLE is only guaranteed to be unbiased in the asymptotic limit, and in fact it is more usually than not biased for finite samples [120]. The CRB can be easily adapted to account for a fixed bias term: for the single parameter case, for example,

$$\mathrm{cov}(\hat{\theta}) \leq \frac{N}{\mathcal{F}(\theta)} \left(1 + \frac{\partial B(\hat{\theta})}{\partial \theta}\right), \tag{4.34}$$

where $B(\hat{\theta})$ is the bias of the estimator. The diverging term containing the bias derivative compensates the diverging inverse FI and leads to the finite variance of the MLE [90]. That effect is stronger for higher N, as illustrated by Figs. 4.7 and 4.8.

In Fig. 4.8 we plot the bias and mean squared error (MSE) of the maximum likelihood estimator for a mode sorting intensity measurement (SPADE) with 21

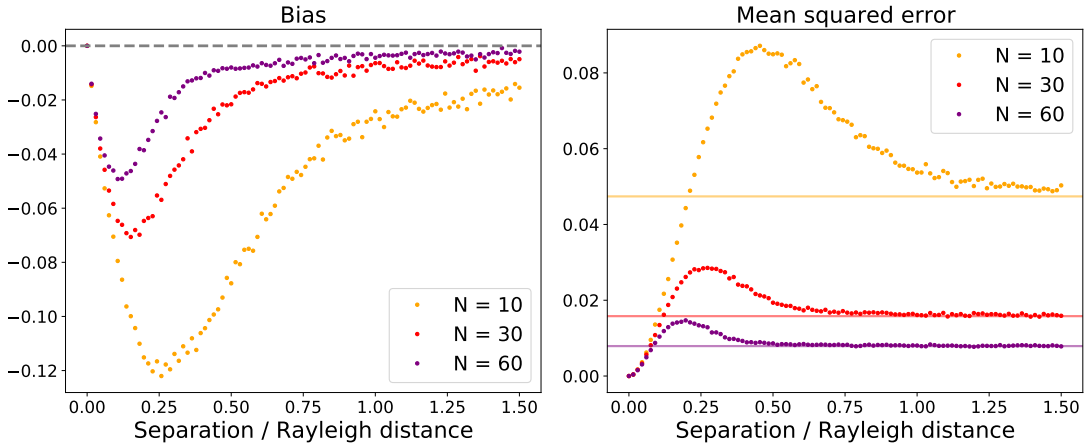


Figure 4.8: Bias (left) and mean squared error (right) of the maximum likelihood estimator for SPADE as a function of separation for $N = 10$ (orange dots), $N = 30$ (red dots) and $N = 60$ (purple dots) photons. The grey dashed line represents the zero bias reference and the solid lines the corresponding qCRBs. We used a Gaussian PSF ($\sigma \approx 0.344$) and the separation is normalised by the Rayleigh distance. The statistics of the MLE for each separation are obtained with 10^5 samples.

modes (up to $\phi_{20}(x)$) and Gaussian PSF, as a function of separation θ, for $N = 10, 30$ and 60. Here the MSE goes temporarily above the CRB and then drops substantially below it for very small separations. This is again caused by the bias but, in this case, the effect is not significant as it disappears for large samples. It is known in the statistics literature as *superefficiency* [6] and it is simply a finite sample artefact.

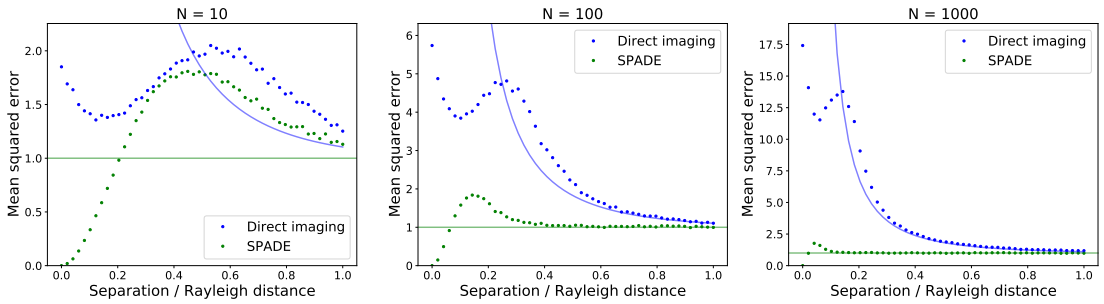


Figure 4.9: Mean squared error of the maximum likelihood estimator for direct imaging (blue) and SPADE (green) as a function of separation for $N = 10$ (left), $N = 100$ (centre) and $N = 1000$ (right) photons. The solid lines represent the corresponding CRBs. We used a Gaussian PSF ($\sigma \approx 0.344$) and the separation is normalised by the Rayleigh distance. The statistics of the MLE for each separation are obtained with 10^4 samples.

In Fig. 4.9 we plot the mean squared error (MSE) of the MLE for both direct imaging and SPADE, as before, for $N = 10, 100$ and 1000. The superiority of

SPADE over direct imaging might have been overestimated by the theory, which ignores the bias and finite sample effects, but it is still there for every scenario.

4.1.4 Strong thermal light

Here we step back from the weak incoherent light approximation $\epsilon \ll 1$ and consider the general case of thermal light with arbitrary strength. This section deviates from the rest of the chapter as it relies on a slightly different theory, which requires the full quantum treatment of Appendix D. The main results we refer to are outlined in Section D.4.

For the same case of two centred sources, the quantum state of the system at the image plane is given by Eq. (D.14) with a P representation (Eq. (D.61))

$$P(\boldsymbol{\alpha}) = \frac{1}{\pi|\Gamma|} e^{-\boldsymbol{\alpha}^\dagger \Gamma^{-1} \boldsymbol{\alpha}}, \tag{4.35}$$

where $\Gamma = \eta \mathrm{diag}(\Gamma_S, \Gamma_S)$, η is the imaging system's attenuation factor and Γ_S is the strength of each source at the object plane. We consider, without loss of generality, a single measurement over the quantum state, since we are interested in the per-photon Fisher information and the latter is simply linear over the number of measurements. This means that $M = 1$ and $N = \epsilon = \mathrm{Tr}\,\Gamma$. We also assume the system's PSF is inversion symmetric.

The (per-photon) quantum Fisher information for the state in Eq. (4.35) is [19]

$$\mathcal{K}(\theta) = \tau(0) - \gamma(\theta)^2 \left[\frac{(2+N)N}{(2+N)^2 - \Delta(\theta)^2 N^2} \right], \tag{4.36}$$

where $\gamma(\theta)$ is the derivative of the overlap term $\Delta(\theta)$ (See appendix F). The evaluation of Eq. (4.36) is a rather tricky and cumbersome process, which involves the geometrical method outlined in Section 2.3.2. The QFI for a Gaussian PSF is plotted in Fig. 4.10 for $N = 0.001, 1, 10$ and 100. Note that it consists of a constant term $\tau(0)$ and a N and θ-dependent term. For $N \ll 1$, the latter is zero and we are left with the QFI of Eq. (4.8). As can be seen from Fig. 4.10, that varying term introduces a dip, around the Rayleigh distance, in the otherwise constant QFI. That dip becomes deeper as N increases but doesn't affect $\mathcal{K}(\theta)$ for very small separations, which remains finite.

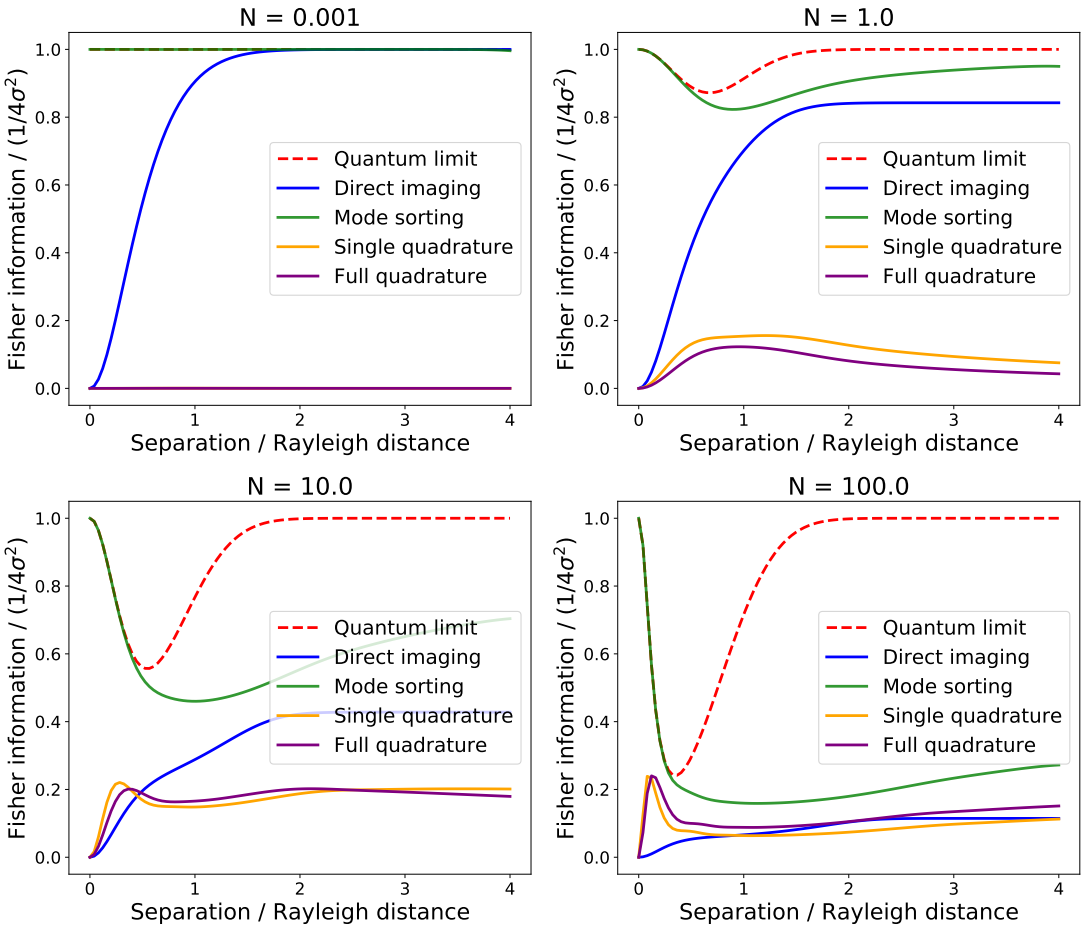


Figure 4.10: Normalised (per photon) quantum Fisher information (red dashed line) and Fisher information for direct imaging (solid blue line), mode sorting (green solid line), single (solid orange line) and full (solid purple line) quadrature mode resolved measurements, as a function of separation (normalised by Rayleigh distance), for a Gaussian PSF ($\sigma \approx 0.344$) and average photon number at the image plane $N = 0.001$ (upper left), $N = 1$ (upper right), $N = 10$ (lower left) and $N = 100$ (lower right).

For direct imaging measurements, the probability of detecting n photons at position x is (Eq. (D.62))

$$\Pr^{(x)}(n) = \frac{1}{\Gamma_x(\theta) + 1} \left(\frac{\Gamma_x(\theta)}{\Gamma_x(\theta) + 1} \right)^n , \tag{4.37}$$

where

$$\Gamma_x(\theta) = \frac{N}{2} \left(|\psi(x + \theta/2)|^2 + |\psi(x - \theta/2)|^2 \right) . \tag{4.38}$$

The Fisher information can be evaluated via reparametrization,

$$\mathcal{F}_{\mathrm{DI}}(\theta) = \int_{-\infty}^{+\infty} \mathrm{d}x \, \left(\frac{\partial \Gamma_x(\theta)}{\partial \theta} \right)^2 \mathcal{F}(\Gamma_x(\theta)), \tag{4.39}$$

where

$$\mathcal{F}(\Gamma_x) = \sum_{n=0}^{\infty} \left(\frac{\partial \mathrm{Pr}^{(x)}(n)}{\partial \Gamma_x} \right)^2 \left(\mathrm{Pr}^{(x)}(n) \right)^{-1} \tag{4.40}$$

is the single-mode Fisher information associated with the estimation of Γ_x. For a intensity measurement such as Eq. (4.37), that term is (see Appendix H)

$$\mathcal{F}^{(n)}(\Gamma_x) = \frac{1}{\Gamma_x \left(\Gamma_x + 1 \right)}. \tag{4.41}$$

The $\mathcal{F}_{\mathrm{DI}}(\theta)$, which can be evaluated numerically, is plotted in Fig. 4.10. As in the weak incoherent case, it drops drastically in the sub-Rayleigh regime, eventually going to zero. However, for strong light, it is no longer quantum-optimal even for large separations. This happens because the multiphoton events it detects contribute with less information than the single-photon ones, resulting in a decrease of overall per-photon efficiency.

For a mode sorting intensity measurement (such as SPADE), the probability of detecting n photons at the m-th mode component is (Eq. (D.62))

$$\mathrm{Pr}^{(m)}(n) = \frac{1}{\Gamma_m(\theta) + 1} \left(\frac{\Gamma_m(\theta)}{\Gamma_m(\theta) + 1} \right)^n, \tag{4.42}$$

where

$$\Gamma_m(\theta) = N R_m^2(\theta/2) = N \frac{Q^m}{m!} e^{-Q}, \tag{4.43}$$

with $Q = \theta^2/16\sigma^2$. The Fisher information is

$$\mathcal{F}_{\mathrm{MS}}(\theta) = \sum_{m=0}^{\infty} \left(\frac{\partial \Gamma_m(\theta)}{\partial \theta} \right)^2 \mathcal{F}(\Gamma_m(\theta)), \tag{4.44}$$

and the $\mathcal{F}(\Gamma_m(\theta))$ term is the same as in Eq. (4.41). Its numerical evaluation is graphed in Fig. 4.10 for a Gaussian PSF. For weak source strengths, we have $\mathcal{F}(\Gamma_m) \approx \Gamma_m^{-1}$, which means that

$$\mathcal{F}_{\mathrm{MS}}(\theta) \approx \frac{N}{4\sigma^2} \frac{e^{-Q}}{Q} \sum_{m=0}^{\infty} (m - Q)^2 \frac{Q^m}{m!} = \frac{N}{4\sigma^2}, \tag{4.45}$$

as before. For stronger light, it remains quantum-optimal for very small separations but loses its optimality for normal distances.

It is also possible to perform mode sorting quadrature measurements, as outlined in Section D.3. The probabilities for single and full quadrature measurements at the m-th order mode are

$$\Pr^{(m)}(q) = \frac{1}{\sqrt{\pi(2\Gamma_m + 1)}} e^{-q^2/(2\Gamma_m+1)} \tag{4.46}$$

and

$$\Pr^{(m)}(q,p) = \frac{1}{2\pi(\Gamma_m + 1)} e^{-(q^2+p^2)/2(\Gamma_m+1)}, \tag{4.47}$$

respectively. Then, the Fisher information is given by Eq. (4.44), but with the term $\mathcal{F}(\Gamma_m)$ as (see Appendix H)

$$\mathcal{F}^{(q)}(\Gamma_m) = \frac{2}{(2\Gamma_m + 1)^2}, \tag{4.48}$$

for single quadrature, and

$$\mathcal{F}^{(q,p)}(\Gamma_m) = \frac{1}{(\Gamma_m + 1)^2} \tag{4.49}$$

for a full quadrature measurement.

The quadrature Fisher informations are plotted in Fig. 4.10. For weak sources and Gaussian PSF, using $\mathcal{F}^{(q)}(\Gamma_m) \approx 2$ leads to

$$\begin{aligned} \mathcal{F}_{\text{single}}(\theta) &\approx \frac{N^2}{2\sigma^2} \frac{e^{-2Q}}{Q} \sum_{m=0}^{\infty} (m - Q)^2 \left(\frac{Q^m}{m!}\right)^2 \\ &= \frac{N^2}{\sigma^2} Q \left[I_0(2Q) - I_1(2Q)\right] e^{-2Q}, \end{aligned} \tag{4.50}$$

where $I_n(x)$ is the modified Bessel function of the first kind and $Q = \theta^2/16\sigma^2$. Similarly, $\mathcal{F}_{\text{full}}(\theta) = \mathcal{F}_{\text{single}}(\theta)/2$. Their poor performance for weak sources, compared with photon counting, is a well-known fact in the literature [6, 25]. This can be understood as follows. For intensity detection, mode sorting corresponds to a null measurement: for very small separations, the signal for the zero-th order mode is null and for higher-order modes is nonzero. This contrast makes the high-order modes stand out and allows us to extract information from them more easily. For quadrature measurements, however, all m-th order mode measurements are distributed as Gaussians with slightly different waists. It is much harder to

distinguish one from another, specially the zero-th order from the high-order ones as there is no clear-cut "discontinuity" between them.

Experimentally, quadrature measurements can be performed with homodyne (single quadrature) and heterodyne (full quadrature) detections. In these measurements, the signal field is mixed with a very strong local oscillator field, which brings excess vacuum noise, on a 50:50 beamsplitter. That added noise ends up degrading the resolution and even bringing back Rayleigh's curse, as can be noted in Fig. 4.10 when the separations approach zero. For strong thermal light, however, despite being weaker than mode sorting and still suffering from Rayleigh's curse, these measurements can still be, for very small separations, advantageous over direct imaging.

It is also possible to define, akin to binary SPADE, binary mode sorting schemes with quadrature measurements. In this case, instead of projecting on "all modes except the 0-th" we project on the first order one. The corresponding FI, for Gaussian PSF, is

$$\mathcal{F}^{(1)}_{\text{single}}(N,\theta) = \frac{2N^2\theta^2\left(1-\theta^2/16\sigma^2\right)^2}{\left(N\theta^2+8\sigma^2 e^{\theta^2/16\sigma^2}\right)}. \tag{4.51}$$

For small separations ($\theta \ll \sigma$) [25],

$$\mathcal{F}^{(1)}_{\text{single}}(N,\theta) \approx \frac{2N^2\theta^2}{\left(N\theta^2+8\sigma^2\right)}, \tag{4.52}$$

and $\mathcal{F}^{(1)}_{\text{full}}(N,\theta) = \mathcal{F}^{(1)}_{\text{single}}(N/2,\theta)/2$.

Finally, it is important to reiterate that, when quadratures measurements are experimentally implemented via sequential dyne detection, i.e., where the measurement itself takes care of the mode sorting stage, then we have to divide the total Fisher information by the number of modes measured. Although this is just a constant photon overhead, it can be significant in some cases.

4.1.5 Further studies

We have only scratched the surface of what has and can been done for the quantum metrology analysis of incoherent light two-point resolution. This seemingly

simple problem has sparked the whole field of quantum-inspired super-resolution (QISR) and continues to yield new discoveries and insights for scientists. In this closing section we summarise some of the research that hasn't been covered here.

In this section we considered equal intensity sources centred on a one dimensional transverse axis. Many studies have focused on generalising this scenario, performing simultaneous estimation with other parameters such as centroid [6, 21, 36, 49, 60, 72], relative intensity [37, 38] and axial displacement [74, 95, 97], eventually leading to the full 3D analysis of the problem [39, 40, 46, 50].

The literature has also devised other optimal measurement schemes. Akin to binary SPADE, SPLICE (*super-resolved position localisation by inversion of coherence along an edge*) manages to extract the first order mode via a π phase shift done by a phase plane in the image plane [17, 87], and SLIVER (*super-localization via image-inversion interferometry*) does the same with an image-inversion interferometer [90]. The theory of image-plane interferometry for two-point estimation was generalised and further developed in Ref. [51]. More ingenious approaches used heterodyne detection [89], collective photon measurements [92], array homodyning [59] and manipulations of the PSF [93, 96].

The statistical analysis has been improved via incorporation of spurious effects such as noise and misalignment [54, 55, 73]. Also, other statistical tools have been used, such as the Bayesian CRB [20] and hypothesis testing [23, 56].

Finally, the ideas of QISR have been experimentally tested and verified in a variety of settings for the case of two-point estimation [87–90, 92–100].

4.2 Intensity moments

While the theory of quantum-inspired super-resolution is well developed for the case of two point sources, it is at a much earlier stage for the general case of extended sources, which is much more difficult to study as it may potentially depend on infinitely many parameters.

A way to tackle that complexity is parametrise the object's intensity distribution by its geometric moments. Although an infinite number of them is necessary

to exactly recover the object, a pretty good reconstruction can be done with enough moments and post-processing techniques [148–151]. That way, the problem becomes of a finite dimensional parameter estimation task, for which the tools developed here are suitable.

In this section we state, without proof, some results from Refs.[22, 24, 26] regarding geometric moment estimation. Our main purpose is to develop the intuition and lay the groundwork for Section 4.3, whose ideas were based on these works, and show proper credit. Therefore, the reader can safely skip this section if uninterested.

4.2.1 Quantum limit

Consider an object plane intensity distribution $I(\vec{r})$, where $\vec{r}$ is a (dimensionless) vector in the 2D transversal plane. We assume the distribution is normalised,

$$\int \mathrm{d}\vec{r}\, I(\vec{r}) = 1, \tag{4.53}$$

and nonnegative. The *object width* is

$$W = \max_{\vec{r}_i,\vec{r}_j\in\mathcal{S}} |\vec{r}_i - \vec{r}_j|, \tag{4.54}$$

where $\mathcal{S}$ is the support of $I(\vec{r})$, i.e., the points for which it is nonzero. The sub-Rayleigh regime is characterised by $W \ll 1$ [24]. The object's $\vec{m} = (m_x, m_y)$-th order geometric moment is defined as

$$\Theta_{\vec{m}} = \int \mathrm{d}\vec{r}\, I(\vec{r})\vec{r}^{|\vec{m}|}. \tag{4.55}$$

Here we have used multi-index notation, where $|\vec{m}| = \sum_i m_i$, $\vec{m}! = \prod_i m_i!$, $\vec{r}^{|\vec{m}|} = \prod_i r_i^{m_i}$ and

$$\partial^{\vec{m}} = \prod_i \frac{\partial^{m_i}}{\partial r_i^{m_i}}. \tag{4.56}$$

Note that the 0-th moment is equal to the total intensity, $\Theta_0 = 1$. In the subdiffraction limit, the moments scale as [22, 24, 26]

$$|\Theta_{\vec{m}}| \leq \int \mathrm{d}\vec{r}\, I(\vec{r})|\vec{r}^{|\vec{m}|}| \leq \left(\frac{W}{2}\right)^{|\vec{m}|}. \tag{4.57}$$

As an example, consider the two-point resolution case, for which

$$I(x) = \frac{1}{2}(\delta(x+\theta/2) + \delta(x-\theta/2)). \tag{4.58}$$

The moments are then

$$\Theta_m = \left(\frac{1+(-1)^m}{2}\right)\left(\frac{\theta}{2}\right)^m. \tag{4.59}$$

In particular, $\Theta_0 = 1$ (total intensity), $\Theta_1 = 0$ (centroid) and $\Theta_2 = \theta^2/4$, which is related to the separation. In general, odd moments are null and even moments are even powers of the separation, $\Theta_{2m} = (\theta/2)^{2m}$.

For weak incoherent sources, the quantum state of the system is

$$\rho = \int \mathrm{d}\vec{r}\; I(\vec{r})\,|\psi_{\vec{r}}\rangle\,\langle\psi_{\vec{r}}|\,, \tag{4.60}$$

where

$$|\psi_{\vec{r}}\rangle = \int \mathrm{d}\vec{r'}\psi(\vec{r}-\vec{r'})\left|\vec{r'}\right\rangle. \tag{4.61}$$

We also assume that the system's PSF is inversion symmetric, $\psi_{-\vec{r}} = \psi_{\vec{r}}$. The direct imaging probabilities are simply

$$\Pr(\vec{r}) = I(\vec{r}) * |\psi(\vec{r})|^2 = \int \mathrm{d}\vec{r'} I(\vec{r'})|\psi(\vec{r}-\vec{r'})|^2. \tag{4.62}$$

We can use the multivariate Taylor expansion,

$$f(\vec{r}+\vec{r'}) = \sum_{\vec{m}} \frac{\partial^{\vec{m}} f(\vec{r})}{\vec{m}!}\vec{r'}^{|\vec{m}|}, \tag{4.63}$$

where the sum is taken over every $\vec{m} \in \mathbb{R}^D$ (D the dimension of the system), to expand $|\psi(\vec{r}-\vec{r'})|^2$ and rewrite the probabilities in terms of the object's moments:

$$\begin{aligned}\Pr(\vec{r}) &= \int \mathrm{d}\vec{r'} I(\vec{r'}) \sum_{\vec{m}} \frac{\partial^{\vec{m}}|\psi(\vec{r})|^2}{\vec{m}!}(-\vec{r'})^{|\vec{m}|} \\ &= \sum_{\vec{m}} \Theta_{\vec{m}} \frac{1}{\vec{m}!}(-\partial)^{\vec{m}}|\psi(\vec{r})|^2.\end{aligned} \tag{4.64}$$

In this way, the Fisher information matrix elements for moment estimation are

$$\mathcal{F}_{\vec{m},\vec{n}} = N\int \mathrm{d}\vec{r} \frac{\left[(-\partial)^{\vec{m}}|\psi(\vec{r})|^2\right]\left[(-\partial)^{\vec{n}}|\psi(\vec{r})|^2\right]}{\vec{m}!\vec{n}!\Pr(\vec{r})}. \tag{4.65}$$

In the sub-Rayleigh regime $W \ll 1$, it can be shown that the FI follows the general scaling [24]

$$\mathcal{F}_{\vec{m},\vec{n}} = N\mathcal{O}(1). \tag{4.66}$$

Finding the exact quantum Fisher information for moment estimation of the state in Eq. (4.60) is still an open problem. However, several bounds derived in the literature allow us to confidently conjecture that the diagonal elements of the QFI scale as [22, 26]

$$\mathcal{K}_{\vec{m},\vec{m}} = N\mathcal{O}(W^{-2\lfloor|\vec{m}|/2\rfloor}), \tag{4.67}$$

where $\lfloor x \rfloor$ is the *floor function*, which returns the greatest integer less than or equal to x. For even total order $|\vec{m}|$, $\mathcal{K}_{\vec{m},\vec{m}} = N\mathcal{O}(W^{-|\vec{m}|})$, while for odd $|\vec{m}|$ we have $\mathcal{K}_{\vec{m},\vec{m}} = N\mathcal{O}(W^{-(|\vec{m}|-1)})$.

For zero (total power) and first (centroids) order moments, both direct imaging FI and the quantum limit scale as $N\mathcal{O}(1)$. For second order moments ($|\vec{m}| = 2$) the QFI scales as $N\mathcal{O}(W^{-2})$ and we see the first advantage over direct imaging. Furthermore, it is at this specific order that we see the lifting of Rayleigh's curse, as exemplified by the two sources' case. For higher order moments, Rayleigh's curse makes a comeback, but we still have a growing advantage over direct imaging.

4.2.2 Mode sorting

Which measurements are quantum-optimal for geometric moment estimation? A good candidate would be PAD basis measurements of previous section. In multi-index notation, the PAD basis is given by

$$\phi_{\vec{m}}(\vec{r}) = i^{|\vec{m}|}\mathcal{F}\left[g_{\vec{m}}(\vec{u})\mathcal{A}(\vec{u})\right](\vec{r}), \tag{4.68}$$

where the $g_{\vec{m}}(\vec{u})$ are polynomials orthogonal with respect to $|\mathcal{A}(\vec{u})|^2$, the aperture function squared. Furthermore, since the PSF is inversion symmetric, the polynomials have a well-defined parity

$$g_{\vec{m}}(-\vec{u}) = (-1)^{|\vec{m}|}g_{\vec{m}}(\vec{u}). \tag{4.69}$$

The mode components are given by

$$R_{\vec{m}}(\vec{r}) = \phi^*_{\vec{m}}(\vec{r}) * \psi(\vec{r}) = (-i)^{\vec{m}} \int \mathrm{d}\vec{u}\; g_{\vec{m}}(\vec{u}) |\mathcal{A}(\vec{u})|^2 e^{i2\pi\vec{u}\cdot\vec{r}}. \tag{4.70}$$

Taylor expanding the exponential term leads to

$$R_{\vec{m}}(\vec{r}) = \sum_{\vec{k}} H_{\vec{m},\vec{k}} \vec{r}^{|\vec{k}|}, \tag{4.71}$$

where

$$H_{\vec{m},\vec{k}} = (-i)^{|\vec{m}|} \frac{(-2\pi i)^{|\vec{k}|}}{\vec{k}!} \int \mathrm{d}\vec{u}\; g_{\vec{m}}(\vec{u}) |\mathcal{A}(\vec{u})|^2 \vec{u}^{|\vec{k}|}. \tag{4.72}$$

Then, the mode sorting probabilities are

$$\Pr(\vec{m}) = \int \mathrm{d}\vec{r}\; I(\vec{r}) |R_{\vec{m}}(\vec{r})|^2 = \sum_{\vec{k},\vec{l}} H_{\vec{m},\vec{k}} H^*_{\vec{m},\vec{l}} \Theta_{\vec{k}+\vec{l}}. \tag{4.73}$$

From Eq. (4.73) it may appear that PAD basis mode sorting can fully reconstruct all geometric moments of a distribution. This measurement, however, has a fundamental limitation. In order to understand that, consider first the one dimensional case: the probabilities are $\Pr(m) = \sum_{kl} H_{mk} H^*_{ml} \Theta_{k+l}$ and

$$H_{mk} \propto \int_{-\infty}^{+\infty} \mathrm{d}u\; g_m(u) |\mathcal{A}(u)| u^k. \tag{4.74}$$

Recall that $\mathcal{A}(-u) = \mathcal{A}(u)$, $g_m(-u) = (-1)^m g_m(u)$, and note that $(-u)^k = (-1)^k u^k$. Therefore, the integrand has a well-defined parity dependent on the sum $m + k$. However, since integration of an odd function over a symmetric interval is zero, we end up having $H_{mk} = 0$ if m and k have opposite parities. Applying that to the summation terms $H_{mk} H^*_{ml} \Theta_{k+l}$ on the mode sorting probability, one can realise that the only nonzero terms are the ones for which k and l have the same parity, which consequently implies that $k + l$ is even. Therefore, mode sorting is only sensitive to the even moments of the object's intensity distribution [22, 89].

That restriction carries over to the two dimensional case: mode sorting is only sensitive to moments with all components even. Why does that happen? Remember that the PAD basis is tailored for the two-point resolution problem, which is highly symmetrical. In that scenario, as we have seen, the separation information is strictly contained in the even order moments, and the odd order moments are all zero.

Therefore, the PAD basis is made to extract maximum information from the former and ignore the latter. In a general imaging scenario, however, that strength becomes a weakness, and severely limits the usefulness of the basis.

In order to overcome that limitation, we have to somehow break the well-defined parity of the PAD modes. One possible solution is to displace the whole sample or, equivalently, the PAD modes' centroid, such that the object is strictly contained inside one quadrant of the transversal plane [89]. In that case, the distribution can be recovered from the even order moments only and PAD mode sorting works again.

Another solution is to use superpositions of PAD modes, such as the *interferometric-PAD* or *iPAD* modes [24]

$$\phi^{\pm}_{\vec{m},\vec{m'}}(\vec{r}) \equiv \frac{1}{\sqrt{2}}\left[\phi_{\vec{m}}(\vec{r}) \pm \phi_{\vec{m'}}(\vec{r})\right]. \tag{4.75}$$

Their mode projections follow linearly from Eq. (4.71),

$$R^{\pm}_{\vec{m},\vec{m'}}(\vec{r}) \equiv \left[\phi^{\pm}_{\vec{m},\vec{m'}}(\vec{r})\right]^* * \psi(\vec{r}) = \frac{1}{\sqrt{2}}\sum_{\vec{k}}\left(H_{\vec{m},\vec{k}} \pm H_{\vec{m'},\vec{k}}\right)\vec{r}^{|\vec{k}|}, \tag{4.76}$$

and the mode sorting probabilities are simply

$$\begin{aligned}\Pr^{\pm}(\vec{m},\vec{m'}) &\equiv \int \mathrm{d}\vec{r}\; I(\vec{r})\left|R^{\pm}_{\vec{m},\vec{m'}}(\vec{r})\right|^2 \\ &= \frac{\Pr(\vec{m}) + \Pr(\vec{m'})}{2} \pm \mathrm{Re}\left\{\sum_{\vec{k},\vec{l}} H_{\vec{m},\vec{k}} H^*_{\vec{m'},\vec{l}}\,\Theta_{\vec{k}+\vec{l}}\right\}.\end{aligned} \tag{4.77}$$

The last term on the RHS of Eq. (4.77) is slightly different from Eq. (4.73) as it depends on $\vec{m'}$ as well, and this extra freedom can be used to recover the odd order moments. Indeed, if $\vec{m}$ and $\vec{m'}$ have different parities, then $\vec{k}$ and $\vec{l}$ have different parities as well, and $\vec{k}+\vec{l}$ is no longer necessarily even.

How to incorporate these modes in a measurement basis? Refs. [22, 24] found that the iPAD modes that maximise the Fisher information are the ones for which $\vec{m}$ and $\vec{m'}$ are as close as possible. They also proposed the *iPAD basis*, which consists of seven sub-bases (since many of these modes are not orthogonal to

each other) $\{\phi^{\alpha}_{\vec{m}}(\vec{r})\}$ as

$$\phi^{0}_{\vec{m}}(\vec{r}) \equiv \phi_{\vec{m}}(\vec{r}), \tag{4.78}$$

$$\phi^{1}_{\vec{m}}(\vec{r}) \equiv \phi^{+}_{\vec{m},\vec{m}+(1,0)}(\vec{r}), \tag{4.79}$$

$$\phi^{2}_{\vec{m}}(\vec{r}) \equiv \phi^{-}_{\vec{m},\vec{m}+(1,0)}(\vec{r}), \tag{4.80}$$

$$\phi^{3}_{\vec{m}}(\vec{r}) \equiv \phi^{+}_{\vec{m},\vec{m}+(0,1)}(\vec{r}), \tag{4.81}$$

$$\phi^{4}_{\vec{m}}(\vec{r}) \equiv \phi^{-}_{\vec{m},\vec{m}+(0,1)}(\vec{r}), \tag{4.82}$$

$$\phi^{5}_{\vec{m}}(\vec{r}) \equiv \phi^{+}_{\vec{m},\vec{m}+(1,-1)}(\vec{r}), \tag{4.83}$$

$$\phi^{6}_{\vec{m}}(\vec{r}) \equiv \phi^{-}_{\vec{m},\vec{m}+(1,-1)}(\vec{r}). \tag{4.84}$$

The Fisher information for mode sorting in the iPAD basis is shown to scale as the QFI [22, 24], Eq. (4.67), so that measurement is, at least, near-quantum optimal.

In this section we have summarised the main literature results with respect to the estimation of geometric moments for weak incoherent light imaging. Many of the concepts and ideas introduced here, such as the issue with parity and its workarounds, quadrant displacement and iPAD basis, will be of fundamental importance to the rest of the text.

4.3 Imaging of point sources

An alternative to geometric moments estimation is to consider the case of imaging a finite, fixed number of point sources. Although not as general as an image can be, it is broad enough to be useful and simple enough to be tractable via the methods developed here. This section contains the results of our work in Ref. [108].

4.3.1 Quantum limit

Consider S weak incoherent sources at the positions $\{\vec{r}_i = (x_i, y_i)\}$ on the object plane, each with relative intensity p_i. The quantum state of the system at the image plane is (Appendix E, Eq. (E.12))

$$\rho = \sum_{i=1}^{S} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|, \tag{4.85}$$

where

$$|\psi_i\rangle = \int \mathrm{d}\vec{r}\psi(\vec{r}-\vec{r}_i)\,|\vec{r}\rangle\,, \tag{4.86}$$

and $|\vec{r}\rangle \equiv \hat{a}^\dagger_{\vec{r}}|\mathrm{vac}\rangle$ is the one-photon state at position $\vec{r}$. The direct imaging probabilities are (Eq. (E.28))

$$\Pr(\vec{r}) = \sum_{i=1}^{S} p_i |\psi(\vec{r}-\vec{r}_i)|^2, \tag{4.87}$$

and mode sorting probabilities are (Eq. (E.29))

$$\Pr(m) = \sum_{i=1}^{S} p_i |R_m(\vec{r}_i)|^2, \tag{4.88}$$

where

$$R_m(\vec{r}_i) = \int \mathrm{d}\vec{r}\;\phi^*_m(\vec{r})\psi(\vec{r}-\vec{r}_i) = \phi^*_m(\vec{r}_i) * \psi(\vec{r}_i) \tag{4.89}$$

is the projection of the i-th source field at the m-th mode. Our goal is to study the simultaneous estimation of all sources' parameters, namely, all relative intensities $\{p_i\}$ and positions $\{x_i\}$ and $\{y_i\}$. With these parameters in hand, plus the prior knowledge that there are exactly S sources, we can completely reconstruct the original object distribution.

While the constraint of fixed and known number of sources seems extremely restrictive, it still allows us to draw fairly general conclusions about imaging bounds. Also, the simultaneous estimation of position and intensity parameters seems like "comparing apples and oranges", but it is still useful since the results we look after are more qualitative than quantitative.

The Fisher information matrices for direct imaging and mode sorting can be easily evaluated via Eq. (2.11). The QFI can be calculated via the method of non-orthogonal bases (Section 2.3.2), which is done in Appendix I.

To showcase our results, we consider a fixed number of sources randomly distributed inside a disk of diameter W, with random relative intensities as well, as exemplified in Fig. (4.11). Then, we vary the diameter W and see how the Fisher information behaves as we go in and out of the subdiffraction limit. In order to have a scalar figure of merit, we choose

$$\mathcal{V}(\boldsymbol{\theta}) = 1/\,\mathrm{Tr}\left[N\mathcal{F}(\boldsymbol{\theta})^{-1}\right], \tag{4.90}$$

which is the inverse of the *total variance* [123] and obeys the CRB-like equation Eq. (2.14). Here we name it generically as *total precision*. Since each distribution of sources will return different values, we average over a large number of them in order to extract general trends.

We can also use the *generalized precision*

$$\mathcal{W}(\boldsymbol{\theta}) = 1/\det\left[N\mathcal{F}(\boldsymbol{\theta})^{-1}\right], \tag{4.91}$$

which is the inverse of the *generalized variance* and obeys the CRB-like equation Eq. (2.15). The generalized precision yields, qualitatively, very similar results compared to the total precision, but its plots are of slightly harder visualization. Therefore, in this chapter we focus on the total precision.

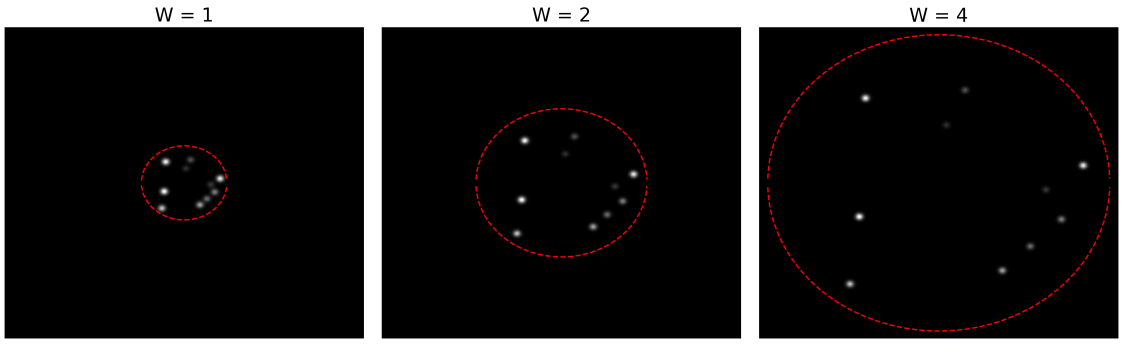


Figure 4.11: Example of object with $S = 10$ point sources, which have random intensities and are randomly located inside a circle (red dashed curve) of diameter $W = 1$ (left), $W = 2$ (centre) and $W = 4$ (right).

The total precision, for estimation of intensities and positions, for the quantum limit and direct imaging is plotted on the left of Fig. 4.12. We use a Gaussian PSF, so the Rayleigh distance becomes $d_R \approx 2.9/\sigma$ and the subdiffraction limit is characterized by $W \ll \sigma$. For simplicity, we use $\sigma = 1$. Note that, for regular distances, direct imaging's precision is fairly constant, while in the sub-Rayleigh limit it quickly deteriorates, as a consequence of Rayleigh's curse. The quantum limit follows the same behaviour, which shows that 1) direct imaging is quantum-optimal for regular distances and 2) Rayleigh's curse is still present for general

imaging. The second conclusion was already observed in the context of geometric moments' estimation.

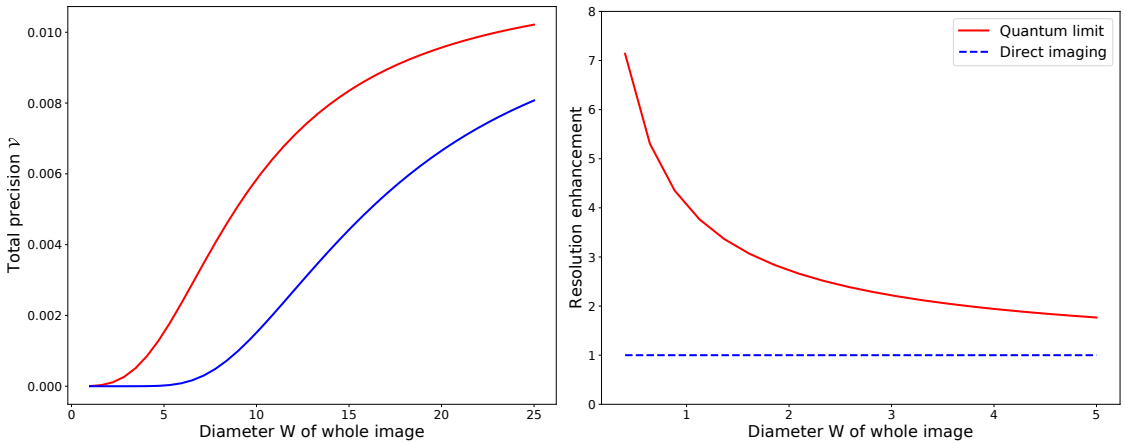


Figure 4.12: Total precision (left) and resolution enhancement (right), averaged over 100 objects for the simultaneous estimation of all intensities and positions, for the quantum limit (red solid curve) and direct imaging (blue solid curve). Gaussian PSF with $\sigma = 1$, objects with 5 sources each, intensities and positions are random.

On the left of Fig. 4.13 we plot the total precision, as before, but for the simultaneous estimation of all intensities, considering the positions known, and equally, on the right of Fig. 4.13, for the estimation of positions when the intensities are known. As we can see, all plots are qualitatively similar. Therefore, for now on we will focus on the full estimation scenario.

We also use another figure of merit, with more practical value, derived from Eq. (4.90), dubbed *resolution enhancement* (RE). If a given measurement, for a given object, for our system with numerical aperture NA, has a certain total precision $\mathcal{V}(\boldsymbol{\theta})$, we try to find which NA$'$ for direct imaging, for the same object, would lead to the same $\mathcal{V}(\boldsymbol{\theta})$. Then, we say that this given measurement has a resolution enhancement of

$$\mathcal{R} = \frac{\mathrm{NA}'}{\mathrm{NA}}. \tag{4.92}$$

Trivially, the RE of direct imaging is always unity. A $R < 1$ means that we performed worse than direct imaging, and $R > 1$ means that we performed better and, therefore, achieved super-resolution. The RE of any measurement is upper bounded by the quantum RE, which is defined via the $\mathcal{V}(\boldsymbol{\theta})$ of the QFI.

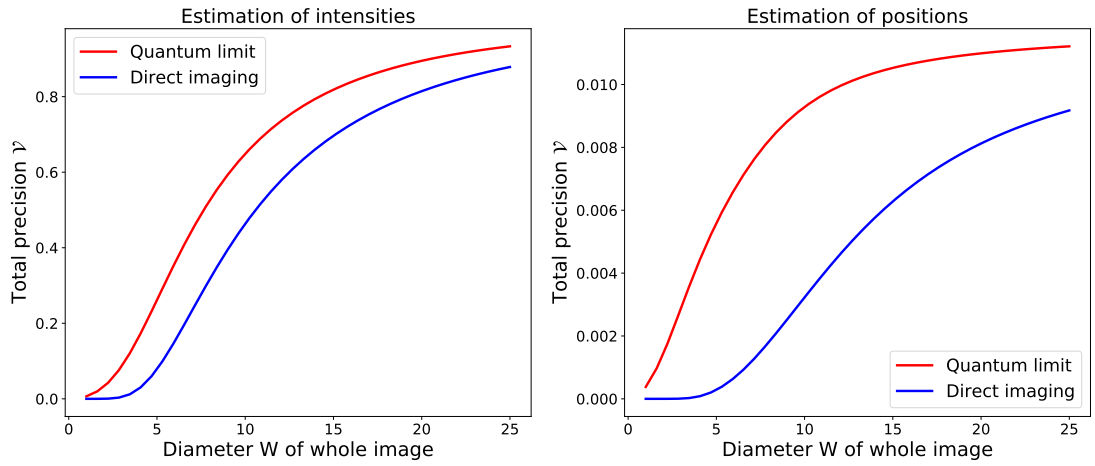


Figure 4.13: Total precision, averaged over 100 objects for the simultaneous estimation of intensities only (left) and positions only (right), for the quantum limit (red solid curve) and direct imaging (blue curves). Gaussian PSF with $\sigma = 1$, objects with 5 sources each, intensities and positions are random.

The resolution enhancements for the quantum limit and direct imaging are plotted on the right of Fig. 4.12. We have focused on the sub-Rayleigh regime (more specifically, $W < 5$), which can not be seen properly in the total precision plot. The RE plot shows that, although we cannot achieve the "infinite resolution" pot of gold as in the two-point case, in the subdiffraction regime there's still a huge room for improvement over conventional direct imaging.

The big question remaining is which measurements can achieve, or at least approach, the quantum limit. The compatibility matrix for this imaging scenario is identically zero (See Appendix I), which means that the quantum limit is attainable. However, we might need collective measurements to achieve it. The SLD matrices for point sources estimation do not commute, so they provide no common eigenbasis for a quantum-optimal measurement. Therefore, we need to exercise our creativity and try to find an optimal scheme on our own.

4.3.2 Mode sorting

A possible candidate for quantum-optimal measurement is the PAD basis intensity mode sorting scheme. For a Gaussian PSF $\psi(x, y)$, the optimal basis

consists of the 2D Hermite-Gaussian modes with the same waist σ as the PSF,

$$\phi_{mn}(x,y) = \frac{1}{\sqrt{2^{m+n}m!n!}} H_m\left(\frac{x}{\sqrt{2}\sigma}\right) H_n\left(\frac{y}{\sqrt{2}\sigma}\right) \psi(x,y), \tag{4.93}$$

a few of which are displayed in Fig. 4.14. However, this basis, as already seen in the geometric moments case, has several limitations. Firstly, its well-defined parity,

$$\phi_{mn}(-x,-y) = (-1)^{m+n}\phi_{mn}(x,y), \tag{4.94}$$

which is a feature in the highly symmetric two-point resolution problem, is a big issue for general imaging as it prevents us from properly recovering odd moments of the object distribution. In practice, this means that images we reconstruct are superpositions of the original one with three "ghost" images, which are its reflections over each quadrant, as depicted in Fig. 4.15.

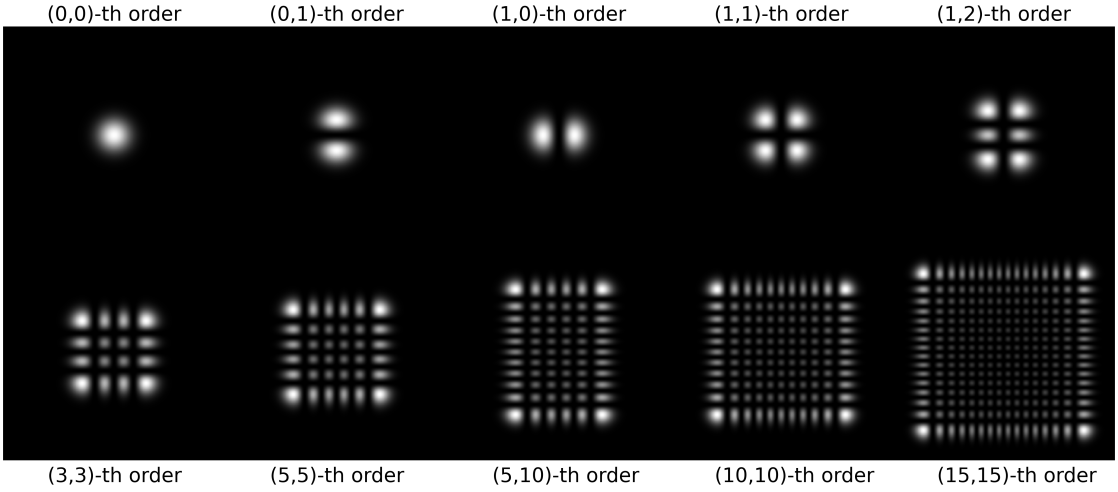


Figure 4.14: Examples of two dimensional Hermite Gaussian modes.

Secondly, the modes only cover a finite area, which is an issue for larger images. In Fig. 4.16 we plot the main area covered[1] by the mode ϕ_{mm} as a function of its order m. As we can see, it grows almost perfectly linearly with mode order. Although an infinite number of modes can cover any area, in any practical realisation of mode sorting we are restricted to a finite number, which poses a restriction to the size of objects that can be imaged.

[1] Although, mathematically, the support of HG modes covers the whole $\mathbb{R}^2$ field, in practice they are basically zero outside their main peaks. Here, we define "covered area" as the area enclosed by their outermost peaks.

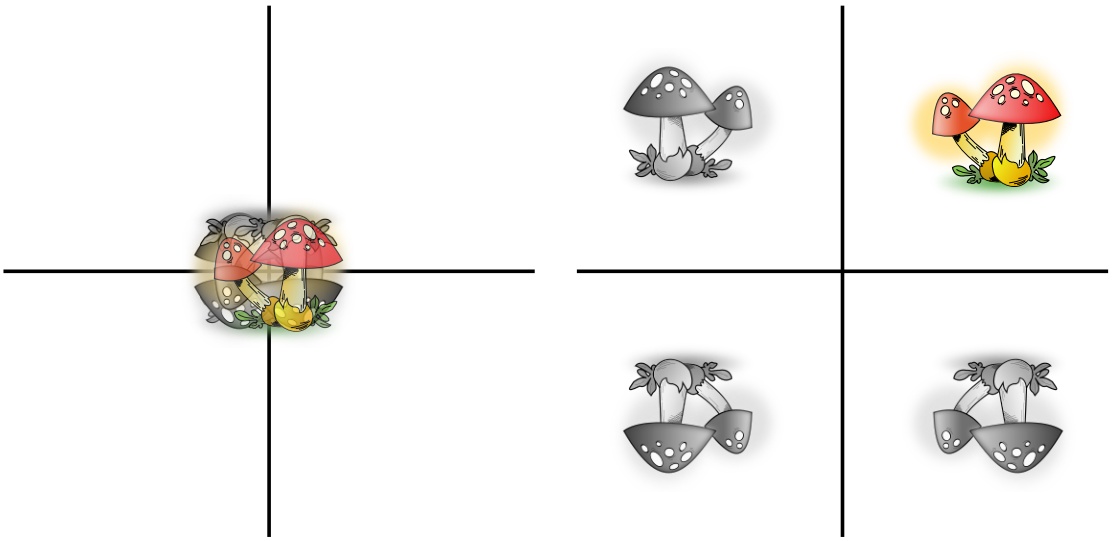


Figure 4.15: When the object's centroid coincides with the modes' (left), its reconstruction consists of the correct (coloured) image superposed with three ghost (greyscale) images. When the centroids are displaced from each other (right), these images drift apart.

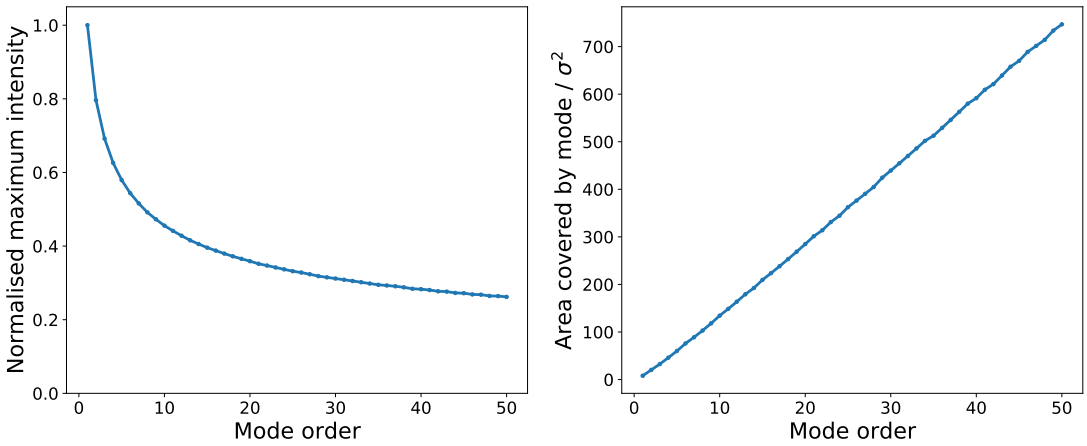


Figure 4.16: Normalised maximum intensity (left) and area covered (right) of a HG mode ϕ_{mm} as a function its order m.

Finally, the peak intensity of the modes decreases with mode order. As seen in Fig. 4.16, the maximum intensity of the mode ϕ_{mm} decreases as $1/m$. This means that, in practice, higher-order modes have increasingly smaller signal to noise ratios (SNR), which might make it harder to detect or generate them.

The performance of mode sorting versus direct imaging and the quantum limit is shown in Fig. 4.17. Its weak performance at large distances is mostly due

to the mode size limitation, which makes the outermost portions of the objects "invisible" to the microscope. Its efficiency is also affected by the parity limitation, where the ghost images decrease the precision. Even with this limitation, it still outperforms direct imaging in the deep subdiffraction limit. This feature hints at us that, once these limitations are somehow sorted, we might have a great measurement scheme in our hands.

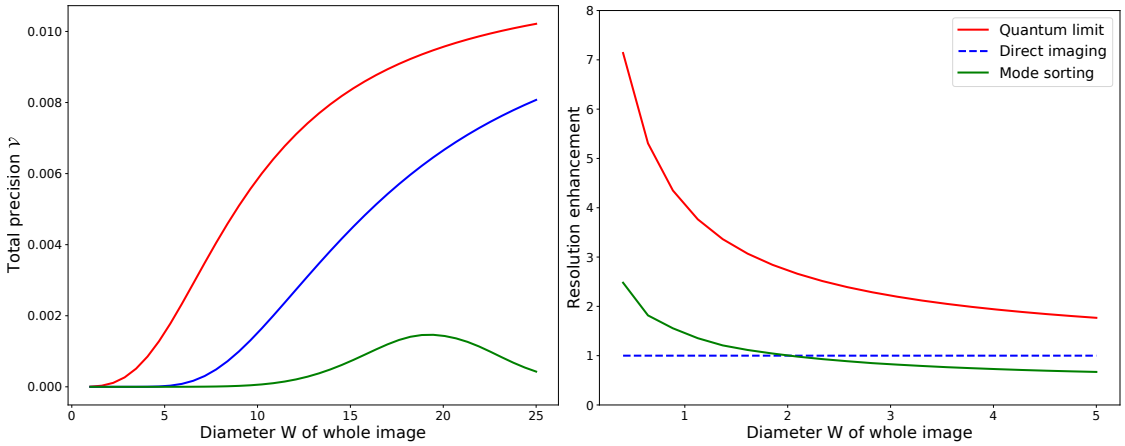


Figure 4.17: Total precision (left) and resolution enhancement (right), averaged over 100 objects for the simultaneous estimation of all intensities and positions. Gaussian PSF with $\sigma = 1$, objects with 5 sources each, intensities and positions are random, mode sorting with $M = 21$ modes along each direction.

Quadrant-displaced modes

A possible solution to the parity issue, suggested in Ref. [89], is to displace the modes or, equivalently, the object, such that the latter lies entirely inside one of the quadrants of the coordinate system centred at the modes' centroid position, as schematised in Fig. 4.15. In this case, the ghost images do not overlap with the desired object and we can perfectly reconstruct it. Therefore, we propose the *quadrant displaced mode sorting* scheme, where the measurement basis is now $\{\phi_{mn}(\vec{r} - \vec{d})\}$. The only free parameters are the direction and magnitude of the displacement vector $\vec{d}$.

Since we are interested in general images, a specific direction of the displacement vector is not too important. Therefore, we simply use a diagonal displacement along any of the four quadrants. For the displacement magnitude, there is a tradeoff:

too small d is not sufficient to break the symmetry, i.e., to split the ghost images, while too large d makes the object too far from the modes' support and, therefore, "invisible" to them. In Fig. 4.18 we graph the precision (normalised to one, as we're only interested in how it varies) as a function of d for different object sizes W, averaged over 1000 samples, for a HG basis up to the 20-th order in either direction. As we can see, the optimal displacement is independent of W. In Fig. 4.19 we fix $W = 5$ and plot the optimal displacement versus the the order of highest-order mode used m (for $\phi_{mm}(\vec{r})$). The optimal displacement now varies linearly with m, which shows us that the meaningful displacement is between object and modes' respective centroids, instead of between the latter and the object's "corners", as the ghost images example would lead us to believe.

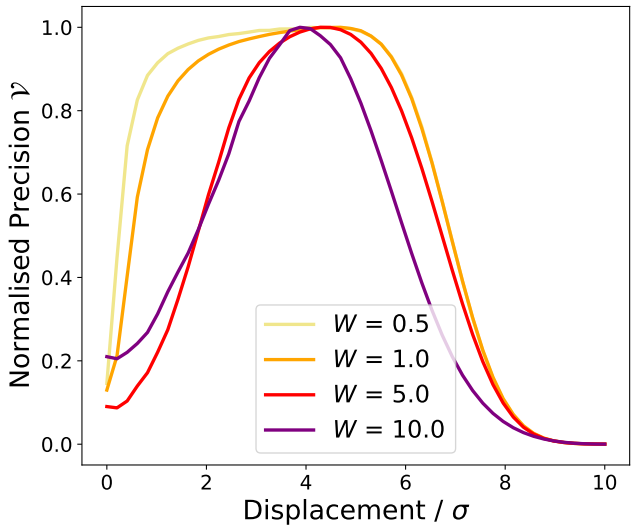


Figure 4.18: Precision (normalized by each curve's maximum) $\mathcal{V}$ versus mode displacement for object size $W = 0.5$ (yellow solid curve), $W = 1$ (orange solid curve), $W = 5$ (red solid curve) and $W = 10$ (purple solid curve). The curves are averaged over 1000 objects, each with $S = 5$ point sources.

In Fig. 4.20 we show the precision and resolution enhancement for quadrant displaced mode sorting, for a fixed displacement of magnitude 4.2. As expected, it behaves worse than even its regular variant for large distances, since the distance between the images' and modes' centroids is now larger: the modes are further

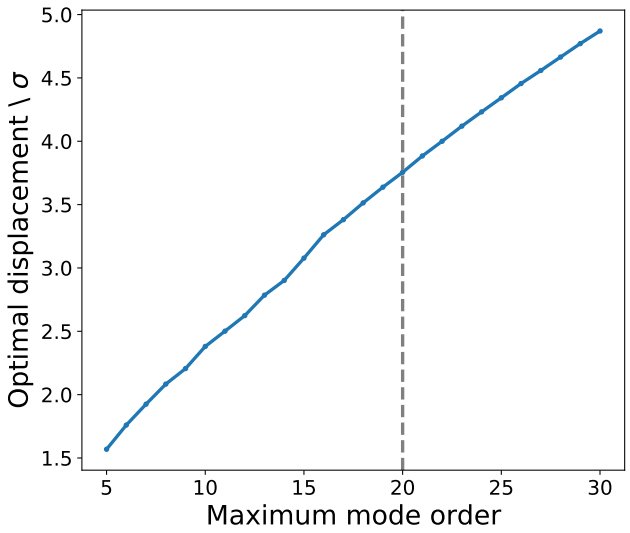


Figure 4.19: Optimal displacement as a function of the basis' maximum mode order. The dashed grey line highlights the mode order $m = 20$. The curve is averaged over 100 objects of size $W = 5$, each with $S = 5$ point sources.

away from the images. For small images, however, it surpasses previous methods and gets really close to the quantum limit.

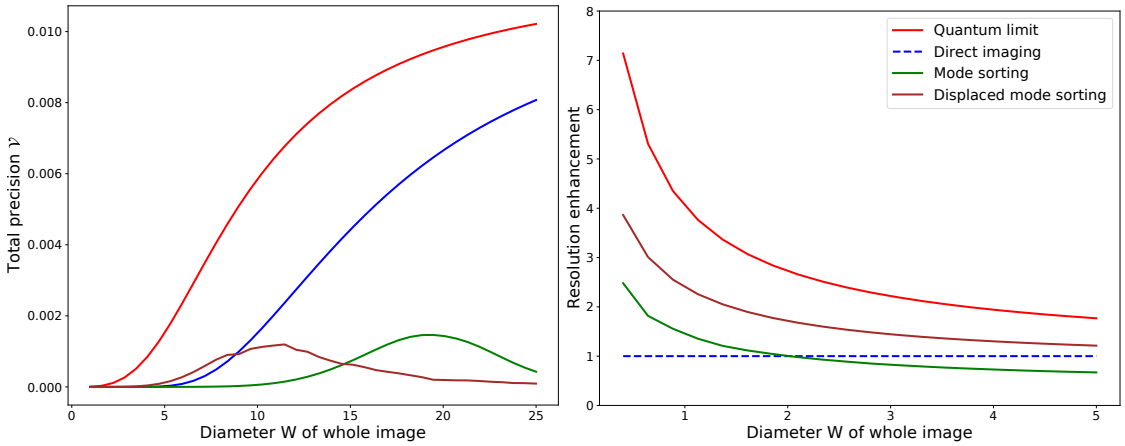


Figure 4.20: Total precision (left) and resolution enhancement (right), averaged over 100 objects for the simultaneous estimation of all intensities and positions. Gaussian PSF with $\sigma = 1$, objects with 5 sources each, intensities and positions are random, mode sorting with $M = 21$ modes along each direction, and mode displacement $d = 4.2$.

Mode scanning

A generalisation of the previous scheme is to perform *mode scanning*, where several mode-displaced measurements are done and the displacement vectors scan

the image plane. In this case, fixing the scanning window as the object $W \times W$ area, the only free parameter is the scanning stride. In Fig. 4.21 we plot the (normalised) precision versus the number $N_{\text{positions}}$ of scanned positions per dimension (x or y), which defines a scanning stride of $W/(N_{\text{positions}} - 1)$. Note that $N^2_{\text{positions}}$ measurements are performed, and the FI has to be normalised accordingly.

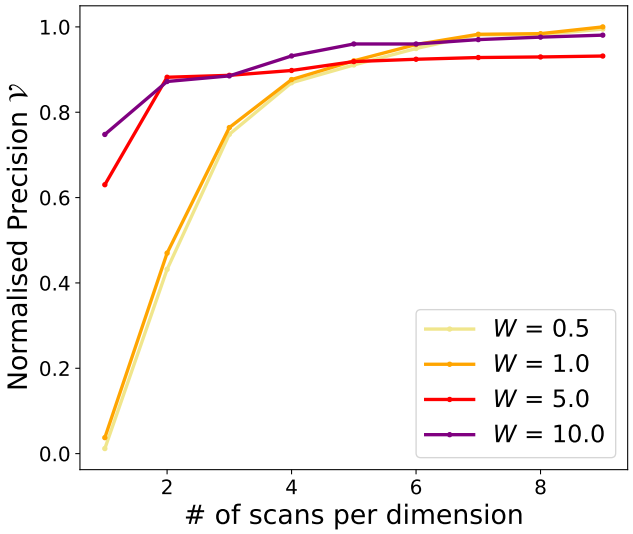


Figure 4.21: Precision (normalized by each curve's maximum) $\mathcal{V}$ versus number of scanned positions along each dimension for object size $W = 0.5$ (yellow curve), $W = 1$ (orange curve), $W = 5$ (red curve) and $W = 10$ (purple curve). The curves are averaged over 100 objects, each with $S = 5$ point sources.

We can see that the precision saturates for all object sizes W at around $N_{\text{positions}} = 5$, which corresponds to the number of sources we used. Indeed, in general the saturation number is around $N_{\text{positions}} = S$. The saturation occurs because the positive effect of having more modes' "points of view", due to the various displacements, balances the negative effect of performing more measurements, whose potential redundancy implies in a "waste" of photons.

In Fig. 4.22 we show the precision and resolution enhancement for scanning mode sorting, for $N_{\text{positions}} = S = 5$. This measurement is the best of two worlds between regular and displaced mode sorting: it is better for large distances, since there are always modes close to the object, as well as in the subdiffraction limit, where the various scans take care of the symmetry limitation.

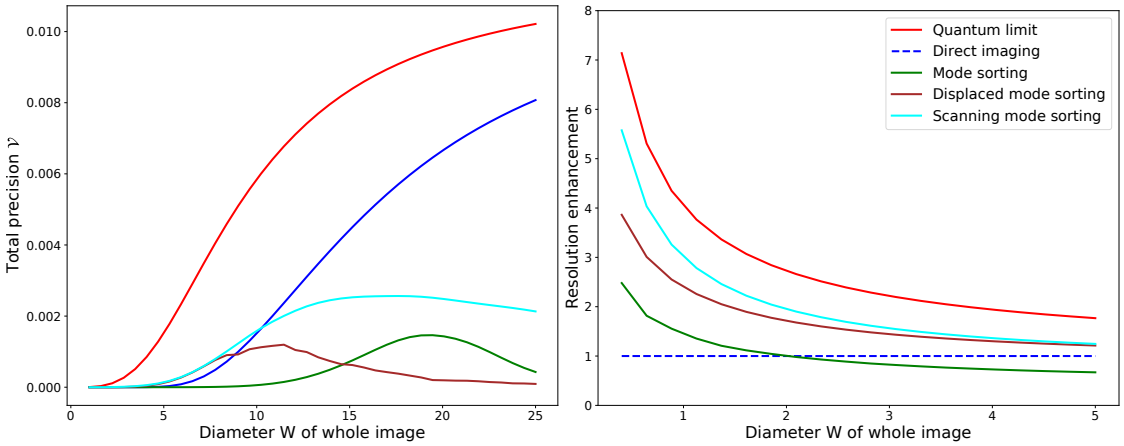


Figure 4.22: Total precision (left) and resolution enhancement (right), averaged over 100 objects for the simultaneous estimation of all intensities and positions. Gaussian PSF with $\sigma = 1$, objects with 5 sources each, intensities and positions are random, mode sorting with $M = 21$ modes along each direction, mode displacement $d = 4.2$ and $N_{\text{positions}} = 5$.

iPAD modes

Another way to break the symmetry of HG modes is to use superpositions of them, such as in the *interferometric PAD (iPAD) basis* introduced in Section 4.2.2. It consists of the seven sub-bases from Eqs. (4.78) to (4.84), which use the mode superpositions from Eq. (4.75): the interference between even and odd order HG modes breaks their well-defined parity and kills the ghost images. In Fig. 4.23 we display some elements of this basis.

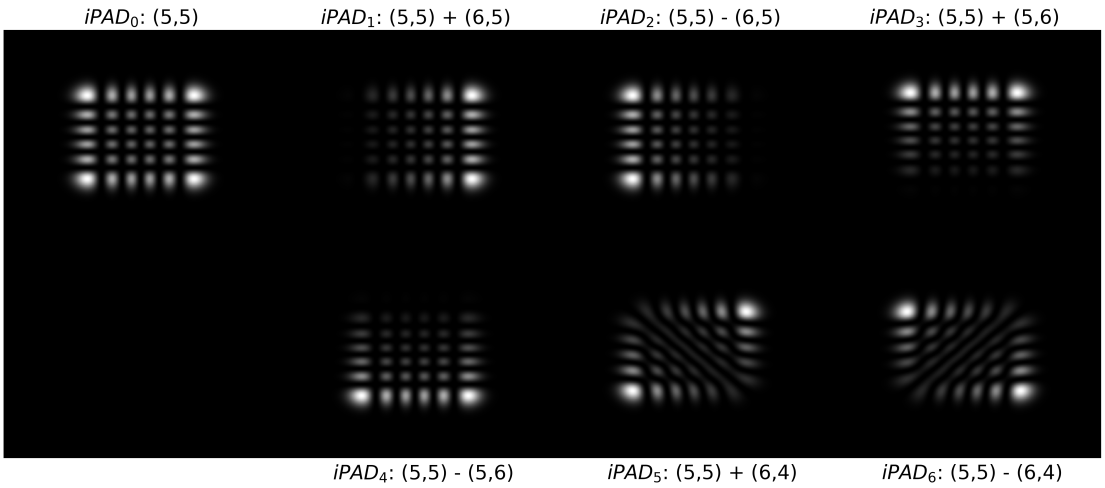


Figure 4.23: $(5, 5)$-th order element in each of the seven iPAD sub-bases.

In Fig. 4.24 we show the total precision and resolution enhancement for interferometric mode sorting. It is as good as mode scanning, and it's the closest we get to the quantum limit for the subdiffraction regime.

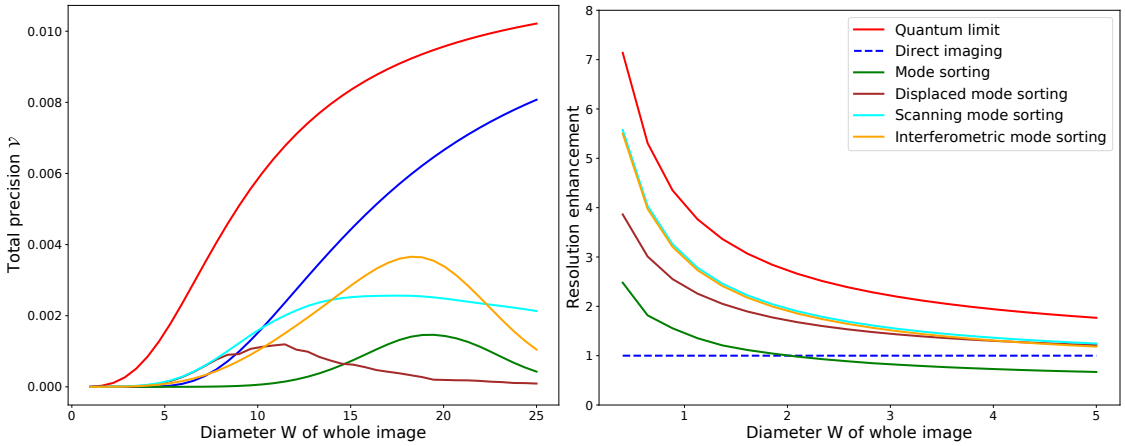


Figure 4.24: Total precision (left) and resolution enhancement (right), averaged over 100 objects for the simultaneous estimation of all intensities and positions. Gaussian PSF with $\sigma = 1$, objects with 5 sources each, intensities and positions are random, mode sorting with $M = 21$ modes along each direction, mode displacement $d = 4.2$ and $N_{\text{positions}} = 5$.

4.3.3 Maximum likelihood estimation

Now that we have seen the achievable precision, what the actual object reconstructions look like? Here we used the maximum likelihood estimator (MLE) to predict the parameters of a 3-source distribution for different object sizes. For each data acquisition session, its MLE was evaluated 100 times, and all those instances are shown so we can have an idea of their statistics (variance and bias). Furthermore, each MLE was evaluated over a dataset containing 10^6 detected photons. The procedure for the estimation was to find the minimum of the likelihood functions via the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, provided by the function "optimize.minimize" from the *Scipy* Python library.

The results of the estimation procedure are shown in Fig. 4.25. For small object diameters W we can see that direct imaging is very biased and more spread than displaced, scanning and interferometric mode sorting. The worst measurement is regular mode sorting, where we can furthermore see the aforementioned ghost images, as each source generates three more mirror images in the reconstruction.

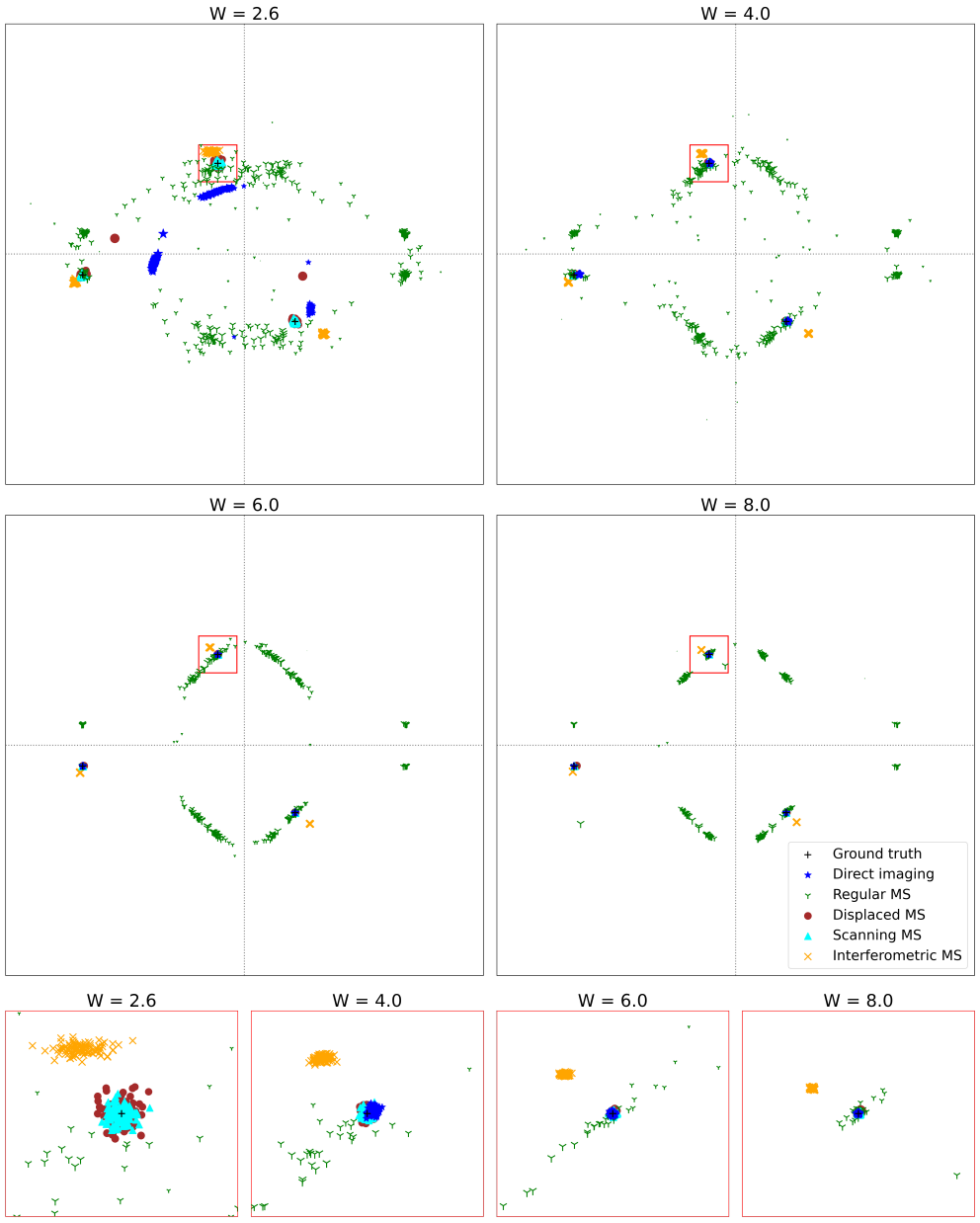


Figure 4.25: Maximum likelihood estimations of several independent measurement (100 instances for each), for a 3-source object with diameter $W = 2.6$ (upper left), $W = 4$ (upper right), $W = 6.0$ (lower left) and $W = 8.0$ (lower right). Below, zoomed sections of the bigger plots.

As we leave the subdiffraction regime, the spreads decrease and the reconstructions become more consistent. For regular separations, i.e., outside the Rayleigh regime, direct imaging is king and its estimates, which can be barely seen in our plots, are heavily centred above the ground truth. A surprise was a persistent bias of interferometric mode sorting even for large separations, which is something that

will need further investigations in the future.

4.3.4 Quadrature measurements

The mode sorting schemes designed here can also be implemented with quadrature instead of intensity measurements. Although the former are clearly disadvantageous over the latter for weak incoherent light, as discussed in Section 4.1.4, they are of much simpler experimental implementation, as we will see in later chapters, and there's hope that they can still outperform direct imaging in the subdiffraction regime.

In Fig. 4.26 we show the precision and resolution enhancement for interferometric mode sorting with full quadrature measurements. Despite its overall weak performance, it can be superior to direct imaging for very small image sizes.

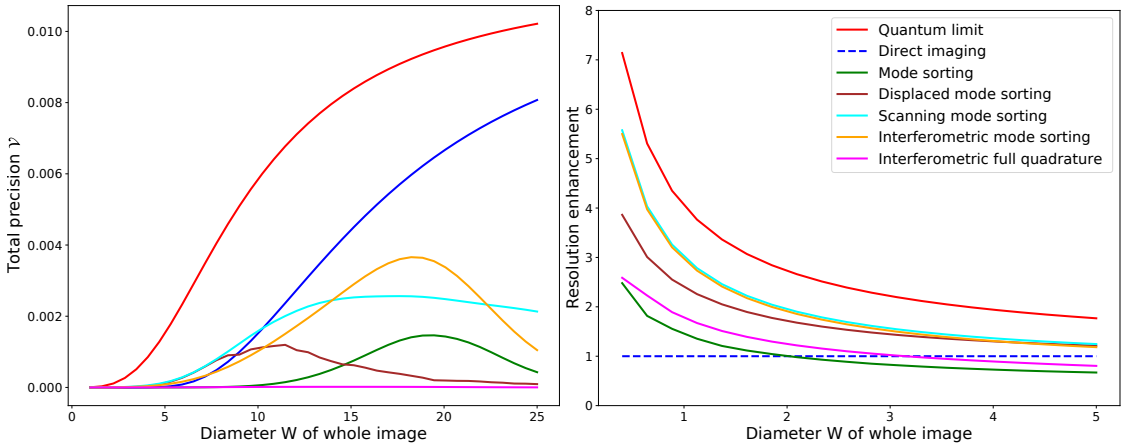


Figure 4.26: Total precision (left) and resolution enhancement (right), averaged over 100 objects for the simultaneous estimation of all intensities and positions. Gaussian PSF with $\sigma = 1$, objects with 5 sources each, intensities and positions are random, mode sorting with $M = 21$ modes along each direction, mode displacement $d = 4.2$ and $N_{\text{positions}} = 5$.

With all these measurement schemes in hand, one might wonder what would be the effect of changing the number of modes used, since so far we've restricted ourselves to $M = 21$ modes along each direction. In Fig. 4.27 we do the same as Fig. 4.26 but for $M = 11$ and $M = 31$. Using higher order modes obviously improves resolution for large distances, as the area covered by the measurement increases. However, that doesn't necessarily translate to higher resolution for very small objects. We conjecture that this is due to the fact that, similarly to the

two-point resolution case seen in Fig. 4.6, for very small separations only a few modes are necessary to get really close to the achievable (infinite-mode) precision.

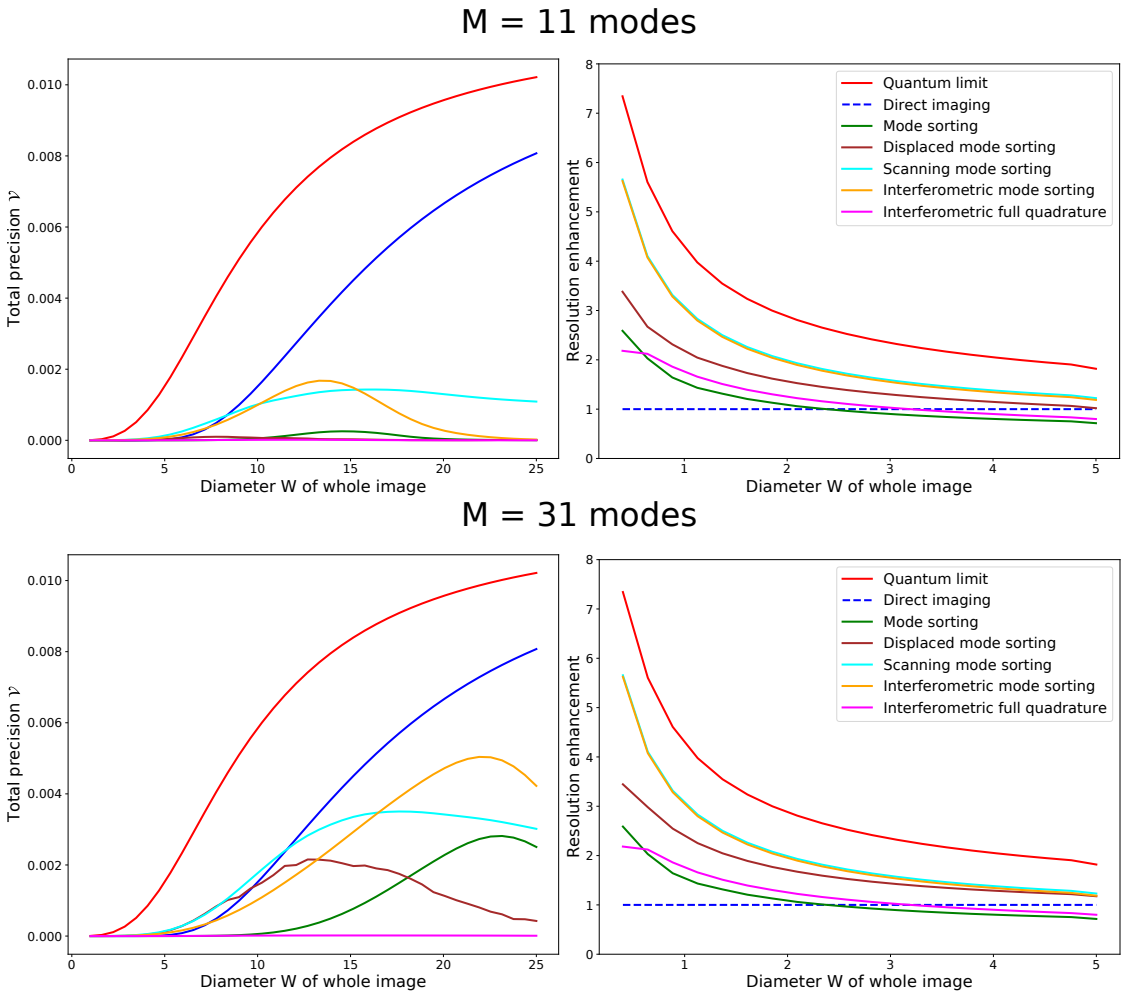


Figure 4.27: Total precision (left) and resolution enhancement (right), averaged over 100 objects for the simultaneous estimation of all intensities and positions, as a function of the diameter W of the whole image. W ranges from 0 to 25 on the left plots and from 0 to 5 on the right. Mode sorting was performed with $M = 11$ (upper panels) and $M = 31$ (lower panels) along each direction. Gaussian PSF with $\sigma = 1$, objects with 5 sources each, intensities and positions are random, mode displacement $d = 4.2$ and $N_{\text{positions}} = 5$.

4.3.5 Circular aperture

So far we have only considered imaging with a Gaussian PSF. In this section we study the case of a circular aperture, which corresponds to an Airy function PSF

$$\psi(\vec{r}) = \frac{1}{\sqrt{\pi}} \frac{J_1(\pi D|\vec{r}|)}{|\vec{r}|}, \tag{4.95}$$

where D is the aperture's diameter and $J_n(x)$ the n-th order Bessel function of the first kind. Its PAD basis is formed by the Fourier transforms of Zernike modes (See Appendix G), dubbed *Fourier-Zernike (FZ) basis*,

$$\phi_{mn}(\vec{r}) = \sqrt{\frac{2(n+1)}{1+\delta_{m,0}}} \frac{J_{n+1}(\pi D r)}{r} \sin\left(m\phi + \frac{\pi}{2}\mathcal{H}(m)\right), \tag{4.96}$$

where $\mathcal{H}(m) = 1$ for $m \geq 0$ and $\mathcal{H}(m) = 0$ for $m < 0$. A few of these modes are shown in Fig. 4.28. Zernike polynomials, and therefore their Fourier transforms, have some rules regarding their indices: the index n can be an arbitrary non-negative integer; the index m can be either positive or negative, but it must have the same parity as n (i.e., both must be either even or odd) and obey $|m| \leq n$. As seen in Fig. 4.28, increasing n increases the radial extent of the FZ mode, while increasing $|m|$ increases the number of "blobs" along the angular direction. Finally, their parity is defined as

$$\phi_{mn}(-\vec{r}) = (-1)^m \phi_{mn}(\vec{r}) = (-1)^n \phi_{mn}(\vec{r}), \tag{4.97}$$

since m and n are always both even or odd.

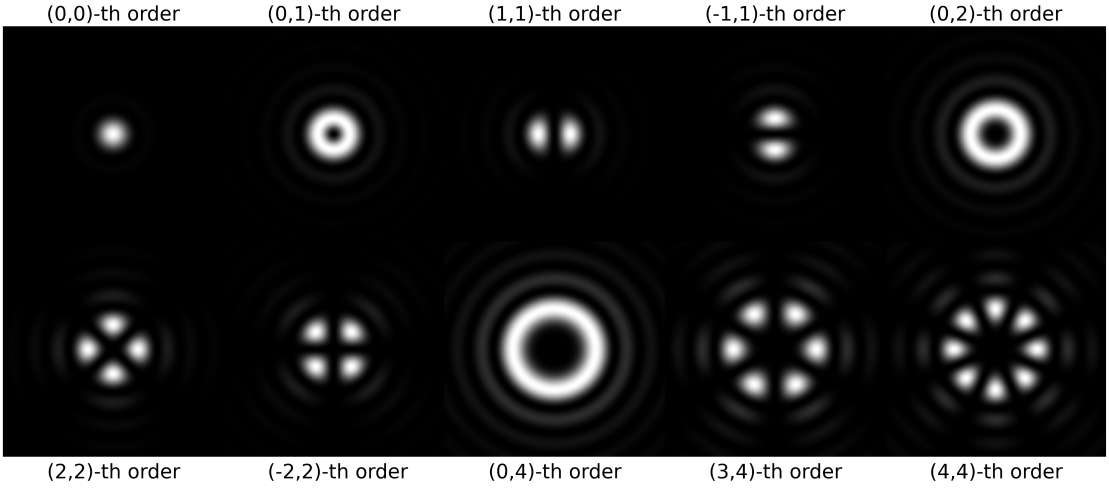


Figure 4.28: Examples of two dimensional Fourier-Zernike (FZ) modes.

In Fig. 4.29 we show the precision and resolution enhancement for several measurements with a circular aperture imaging system. Qualitatively, the results

are mostly similar to the case of a Gaussian aperture. Scanning mode sorting is superior to the interferometric scheme, and displaced mode sorting performs poorly compared to the Gaussian case. We believe these differences are due to the radial structure of the FZ modes.

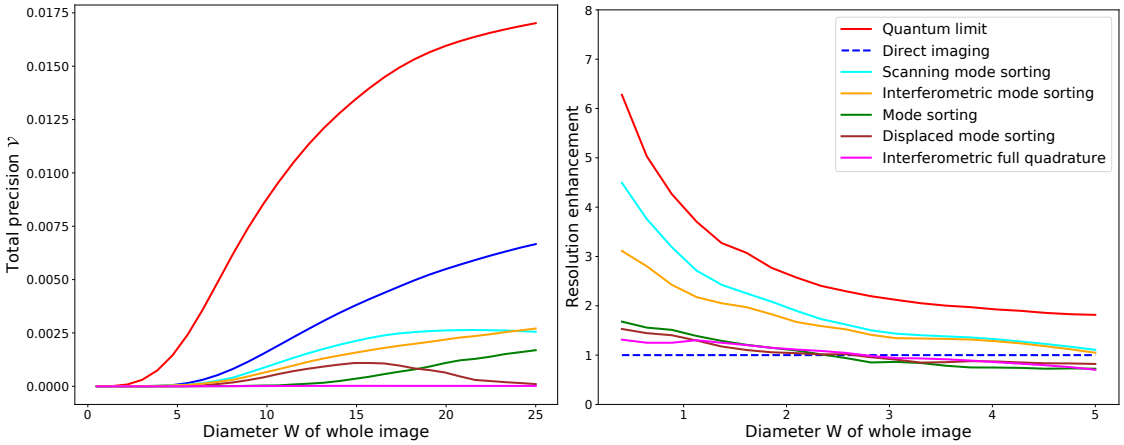


Figure 4.29: Total precision (left) and resolution enhancement (right), averaged over 100 objects for the simultaneous estimation of all intensities and positions. Airy PSF with $D = 0.42$, objects with 5 sources each, intensities and positions are random, mode sorting with $M = 21$ modes along each direction, mode displacement $d = 4.2$ and $N_{\text{positions}} = 5$.

4.3.6 Conclusion

In this section we have performed a quantum metrological analysis of the far-field imaging of a discrete 2D distribution of point sources. We have found that, unlike the case of two point sources, Rayleigh's curse is still present and unavoidable. However, a careful assessment of the quantum Fisher information shows that in the subdiffraction regime a resolution much higher than the one of conventional direct imaging methods can be achieved.

We have not found yet a measurement scheme that exactly achieves the quantum precision bound for all image sizes or, at least, for the sub-Rayleigh regime. Furthermore, such a measurement is very likely to be dependent on the actual object being imaged, which might require feedback control and adaptive schemes, and it might be reliant on collective measurements of photons, which might be impractical for current technology.

Regardless of that open problem, we have found classes of measurements that are superior to direct imaging and get really close to the quantum limit for very small separation. They are based on the PAD mode basis, which yielded quantum efficiency for the two source problem. However, when extended to general imaging, these modes behave poorly. The reason is that their parity symmetry, which is a feature for the highly symmetric two-point resolution problem, is a serious limitation in general since this makes them insensitive to the antisymmetric parts of images. In order to solve this, we proposed three adaptations to the PAD basis. The first, displaced mode sorting, would simply translate the modes in such a way that all the object's information would be contained in the symmetric part of their distribution, which the modes would be sensitive to. This method, although of poor performance for normal distances (i.e., outside the Rayleigh regime), achieved reasonable super-resolution for small separations. The second, scanning mode sorting, is a generalization of the first and displaces the modes' centroid to a grid of positions in the 2D image plane. This achieved remarkable performance, being superior to displaced modes for large distances and approaching the quantum limit for small ones. Finally, we proposed interferometric mode sorting, based on the iPAD basis [22, 24] of superpositions of PAD modes. This specific choice of modes breaks the parity symmetry and in doing so achieves resolutions comparable to scanning mode sorting.

Finally, we studied the resolution of mode sorting with quadrature measurements. While it is known that they perform poorly for weak incoherent light [25], we have shown that they can still beat direct imaging for very small objects. These measurements, albeit inferior to intensity mode sorting, can be implemented much more easily with current technology. This happens because they can be carried out via heterodyne detection, which has a "mode sorting" mechanism embedded in it and allows us to avoid the use of actual mode sorting implementations, which are currently very rudimentary and limited.

There are plenty of open questions and future research directions for this specific field. As already stated, we are unaware of a measurement that is quantum-optimal,

and therefore achieves the highest resolution permitted by nature. It would be desirable if such measurements were implementable in the laboratory, and even better if that could be accurately done with current technology.

The schemes presented here can also be further improved, specially with the aid of adaptive strategies. Instead of performing a fixed displacement or fixed grid scanning, we could use optimization techniques in order to find optimal mode displacements that achieve highest resolution. Or, we could try to find new superpositions of PAD modes, beyond the iPAD basis, that might be optimal for each imaged object. Hopefully, of one these adaptations might even lead to the elusive quantum-optimality.

This whole analysis can also be extended for sources of different nature, such as fully coherent, partially coherent, strong incoherent or even quantum light. It can also be extended to more general distributions of objects, such as point sources on a 3D volume, point sources of an arbitrary (not fixed beforehand) number, or even extended sources.

On the experimental front, a fully-fledged implementation of 2D imaging of any of these super-resolution schemes would be very interesting. We have preliminary results on such a task, but for macroscopic samples only, i.e., in a proof-of-principle experiment.

5 | Statistical analysis of coherent light imaging

In this chapter we perform a similar statistical analysis as in Chapter 4, but this time for fully coherent light sources. Since the quantum state of coherent light fields is pure, the statistical analysis is highly simplified and we are able to reach analytical formulas that are applicable to various imaging scenarios. However, now we have the added difficulty that the number of photons that reaches the image plane is not simply proportional to the number that left the object plane, due to interference effects. This simple fact complicates the statistical analysis and has created a fair bit of confusion in the literature. We hope this chapter also helps in clarifying this issue.

The results contained here are from our work in Ref. [109]. In order to keep up with the mathematics, we suggest the reading of Appendix D or, at least, acquaintance with the results derived in Section D.5.

In Section 5.1 we derive the formula for the quantum Fisher information for coherent light sources, and in Section 5.2 we analyse the corresponding Fisher information, with respect to intensity and quadrature mode-sorted measurements. This formalism then is applied to some imaging scenarios in Section 5.3.

5.1 Quantum Fisher information

The image plane quantum state of S uncorrelated coherent sources with vector of amplitudes $\boldsymbol{A_0}$ is, via Eqs. (D.69) and (D.15),

$$|\Psi\rangle = \prod_{j=1}^{S} e^{A_j \hat{\psi}_j^\dagger - A_j^* \hat{\psi}_j} |\mathrm{vac}\rangle \,, \tag{5.1}$$

where $\boldsymbol{A} = \sqrt{\eta}\boldsymbol{A_0}$, η is the imaging system's attenuation factor and $\hat{\psi}_j^\dagger$ are the PSF operators (Eq. (D.3))

$$\hat{\psi}_j^\dagger = \int \mathrm{d}\vec{r}\; \psi(\vec{r} - \vec{r}_j)\hat{a}_{\vec{r}}^\dagger, \tag{5.2}$$

with $\{\vec{r}_j\}$ the positions of the sources in the object plane. Since the PSF operators are hard to deal with, due to their non-orthogonality, we can rewrite Eq. (5.1) in a more suitable, orthonormal basis $\{\phi_m(\vec{r})\}$. Defining the mode operators

$$\hat{\phi}_m^\dagger = \int \mathrm{d}\vec{r}\; \phi_m(\vec{r})\hat{a}_{\vec{r}}^\dagger, \tag{5.3}$$

our quantum state can be decomposed as (Appendix D, Eq. (D.21))

$$|\Psi\rangle = \bigotimes_{m=0}^{\infty} |E_m\rangle_m \,, \tag{5.4}$$

where $|\alpha\rangle_m$ is a coherent state of $\hat{\phi}_m$ and

$$E_m = \int \mathrm{d}\vec{r}\phi_m^*(\vec{r})E(\vec{r}) \tag{5.5}$$

is the m-th mode component of the electric field (Eq. (D.22))

$$E(\vec{r}) = \sum_{j=1}^{S} A_j \psi(\vec{r} - \vec{r}_j). \tag{5.6}$$

The electric field can also be used to define the average photon number at the image plane,

$$\epsilon = \int \mathrm{d}\vec{r}\, |E(\vec{r})|^2 \,. \tag{5.7}$$

The quantum Fisher information can be easily evaluated for coherent states because they are pure states. Assume that our state depends on a vector of parameters $\boldsymbol{\theta}$. Then (Eq. (2.30)),

$$\mathcal{H}_{ij} = 4\left[\langle \partial_i\Psi|\partial_j\Psi\rangle - \langle \partial_i\Psi|\Psi\rangle \langle \Psi|\partial_j\Psi\rangle\right], \tag{5.8}$$

from which the QFI and compatibility matrices follow from $\mathcal{K} = \mathrm{Re}\{\mathcal{H}\}$ and $\Gamma = \mathrm{Im}\{\mathcal{H}\}$. The derivatives can be computed from decomposition (5.4) and the product rule,

$$|\partial_i\Psi\rangle = \sum_m |\partial_i E_m\rangle_m \otimes \left(\bigotimes_{m'\neq m} |E_{m'}\rangle_{m'}\right). \tag{5.9}$$

Since the mode basis is orthonormal, $[\hat{\phi}_m^\dagger, \hat{\phi}_n] = \delta_{mn}$, we have that $\langle E_m|_m$ and $|E_n\rangle_n$ live in orthogonal Hilbert spaces for $m \neq n$. This orthonormality allows the $\mathcal{H}$ matrix to be rewritten as

$$[\mathcal{H}]_{ij} = 4 \sum_m \langle \partial_i E_m | \partial_j E_m \rangle_m - \langle \partial_i E_m | E_m \rangle_m \langle E_m | \partial_j E_m \rangle_m . \tag{5.10}$$

Now, recall that the coherent states $|\alpha\rangle_m$ can be written via the displacement operator as

$$|E_m\rangle_m = e^{E_m \hat{\phi}_m^\dagger - E_m^* \hat{\phi}_m} |\text{vac}\rangle , \tag{5.11}$$

from which the derivatives can be evaluated,

$$|\partial_i E_m\rangle_m = \left[(\partial_i E_m) \hat{\phi}_m^\dagger - (\partial_i E_m^*) \hat{\phi}_m \right] |E_m\rangle_m . \tag{5.12}$$

Using the coherent states' definition, $\hat{\phi}_m |\alpha\rangle_m = \alpha |\alpha\rangle_m$ (equivalently, $\langle\alpha|_m \hat{\phi}_m^\dagger = \alpha^* \langle\alpha|_m$) and $\langle\alpha|_m \hat{\phi}_m \hat{\phi}_m^\dagger |\alpha\rangle_m = 1 + |\alpha|^2$, we can evaluate the terms of Eq. (5.10),

$$\begin{aligned} \langle \partial_i E_m | \partial_j E_m \rangle_m = {} & (\partial_i E_m^*)(\partial_j E_m) \\ & + \left[(\partial_i E_m^*)(\partial_j E_m) + (\partial_i E_m)(\partial_j E_m^*) \right] |E_m|^2 \\ & - (\partial_i E_m)(\partial_j E_m)(E_m^*)^2 - (\partial_i E_m^*)(\partial_j E_m^*) E_m^2 \end{aligned} \tag{5.13}$$

and

$$\begin{aligned} \langle \partial_i E_m | E_m \rangle_m \langle E_m | \partial_j E_m \rangle_m = {} & \left[(\partial_i E_m^*)(\partial_j E_m) + (\partial_i E_m)(\partial_j E_m^*) \right] |E_m|^2 \\ & - (\partial_i E_m)(\partial_j E_m)(E_m^*)^2 - (\partial_i E_m^*)(\partial_j E_m^*) E_m^2 . \end{aligned} \tag{5.14}$$

As can be easily seen, most of terms cancel each other, and we are left with

$$\begin{aligned} [\mathcal{H}]_{ij} &= 4 \sum_m (\partial_i E_m^*)(\partial_j E_m) \\ &= 4 \int_{\mathbb{R}^2} \mathrm{d}\vec{r} \mathrm{d}\vec{r'} \left(\frac{\partial E^*(\vec{r})}{\partial \theta_i} \right) \left(\frac{\partial E(\vec{r'})}{\partial \theta_j} \right) \sum_m \phi_m(\vec{r}) \phi_m^*(\vec{r'}) \\ &= 4 \int_{\mathbb{R}^2} \mathrm{d}\boldsymbol{r} \left(\frac{\partial E^*(\vec{r})}{\partial \theta_i} \right) \left(\frac{\partial E(\vec{r})}{\partial \theta_j} \right) . \end{aligned} \tag{5.15}$$

From there, the QFI and compatibility matrices $\mathcal{K} = \mathrm{Re}\,\mathcal{H}$ and $\Gamma = \mathrm{Im}\,\mathcal{H}$ follow simply. Eq. (5.15) is remarkably simple and elegant compared with its incoherent

counterparts, and boils down to a simple classical field evaluation. It is important to note that Eq. (5.15) is independent of the mode basis we used to derive it: the orthonormal mode decomposition is used simply for mathematical convenience.

Finally, note that if the electric field is real, which happens, for example, when the PSF is real and all sources are in phase, then the compatibility matrix Eq. (2.24) is identically zero, as the matrix $\mathcal{H}$ becomes purely real. In that case, all object parameters can be simultaneously estimated with optimality, i.e., the full QFI matrix can be achieved by some measurement. When the field is not real, however, we might encounter restrictions and quantum-optimal measurements might not exist.

5.2 Fisher information

In this section we evaluate the Fisher information for intensity and quadrature measurements over mode components of the state in Eq. (5.1). We put the spatial continuous basis (for direct imaging) in the same framework as discrete bases via a slight abuse of notation, using $\phi_{\vec{r}_0}(\vec{r}) = \delta^2(\vec{r} - \vec{r}_0)$.

For intensity measurements, we have the Poisson probabilities

$$\Pr^{(m)}(n) = \frac{e^{-|E_m|^2}|E_m|^{2n}}{n!}. \tag{5.16}$$

Since the probabilities only depend on the squared field $|E_m|^2$, the FI follows from reparametrisation[1] as

$$\mathcal{F}_{ij} = \sum_{m=0}^{\infty} \left(\frac{\partial|E_m|^2}{\partial\theta_i}\right)\left(\frac{\partial|E_m|^2}{\partial\theta_j}\right)\mathcal{F}(|E_m|^2), \tag{5.17}$$

where

$$\mathcal{F}(|E_m|^2) = \frac{1}{|E_m|^2}, \tag{5.18}$$

which follows from a simple reparametrization of (See Appendix H, Proposition 7)

$$\mathcal{F}(|E_m|) = 4. \tag{5.19}$$

[1]The sum is replaced by an integral for a continuous mode basis.

The FI can then be expanded as

$$\begin{aligned}\mathcal{F}_{ij} &= 2\,\mathrm{Re}\left\{\sum_{m=0}^{\infty}\frac{\partial E_m^*}{\partial\theta_i}\frac{\partial E_m}{\partial\theta_j}\right\}|E_m|^2\mathcal{F}(|E_m|^2)\\ &\quad+2\,\mathrm{Re}\left\{\sum_{m=0}^{\infty}\frac{\partial E_m}{\partial\theta_i}\frac{\partial E_m}{\partial\theta_j}(E_m^*)^2\right\}\mathcal{F}(|E_m|^2)\\ &=\frac{1}{2}\mathcal{K}_{ij}+2\,\mathrm{Re}\left\{\sum_{m=0}^{\infty}\frac{\partial E_m}{\partial\theta_i}\frac{\partial E_m}{\partial\theta_j}\frac{E_m^*}{E_m}\right\}.\end{aligned}\tag{5.20}$$

Note that the first term on the RHS is simply half of the QFI. If the electric field is real, then the second term on the RHS is equal to the first, and the FI is identically equal to the QFI. In that case, note that the FI is also basis-independent, as long as all (infinite) basis elements are measured: intensity mode sorting and direct imaging are theoretically equivalent and quantum-optimal. In practice, pixel size, number of modes, among several other parameters, can affect this equivalence.

For full quadrature measurements,

$$\mathrm{Pr}^{(m)}(q,p)=\frac{1}{2\pi}e^{-\left(q-\sqrt{2}\,\mathrm{Re}\{E_m\}\right)^2/2}e^{-\left(p-\sqrt{2}\,\mathrm{Im}\{E_m\}\right)^2/2},\tag{5.21}$$

and the FI is

$$\mathcal{F}_{ij}=\sum_m\iint_{-\infty}^{\infty}\mathrm{d}q\mathrm{d}p\left(\frac{\partial\ln\mathrm{Pr}^{(m)}(q,p)}{\partial\theta_i}\right)\left(\frac{\partial\ln\mathrm{Pr}^{(m)}(q,p)}{\partial\theta_j}\right)\mathrm{Pr}^{(m)}(q,p).\tag{5.22}$$

Since the distribution parameters are encoded into both quadratures $\mathrm{Re}\{A_m\}$ and $\mathrm{Im}\{A_m\}$, we need to use the chain rule as

$$\frac{\partial\ln\mathrm{Pr}^{(m)}(q,p)}{\partial\theta_i}=\frac{\partial\ln\mathrm{Pr}^{(m)}(q,p)}{\partial\,\mathrm{Re}\{E_m\}}\frac{\partial\,\mathrm{Re}\{E_m\}}{\partial\theta_i}+\frac{\partial\ln\mathrm{Pr}^{(m)}(q,p)}{\partial\,\mathrm{Im}\{E_m\}}\frac{\partial\,\mathrm{Im}\{E_m\}}{\partial\theta_i}.\tag{5.23}$$

After some algebra and solving the Gaussian integrals we arrive at

$$\mathcal{F}_{ij}=2\left(\sum_m\frac{\partial\,\mathrm{Re}\{E_m\}}{\partial\theta_i}\frac{\partial\,\mathrm{Re}\{E_m\}}{\partial\theta_j}+\frac{\partial\,\mathrm{Im}\{E_m\}}{\partial\theta_i}\frac{\partial\,\mathrm{Im}\{E_m\}}{\partial\theta_j}\right),\tag{5.24}$$

which can be reorganized as

$$\mathcal{F}_{ij}=2\,\mathrm{Re}\left\{\sum_m\left(\frac{\partial E_m^*}{\partial\theta_i}\right)\left(\frac{\partial E_m}{\partial\theta_j}\right)\right\}=\frac{1}{2}\mathcal{K}_{ij}.\tag{5.25}$$

The quadrature FI is always half the QFI, regardless of mode basis. In the case of real electric field, this means that it is always smaller than the FI for intensity

measurements, which are quantum-optimal. This happens because quadratures are complementary variables, and this non-commutativity decreases the information available through measurements. Experimentally, these measurements are usually performed via heterodyne detection, whose local oscillator field, which is mixed with the signal field and is much stronger than the latter, brings a lot of vacuum noise to the detectors.

In the presence of relative phases and/or complex PSFs, however, that disadvantage becomes a feature and quadrature measurements really shine, since they are sensitive to field phases. Intensity measurements, on the other hand, discard all phase information and are therefore less precise.

5.3 Imaging scenarios

5.3.1 Two-point resolution

We firstly go back to our simplest and dearest imaging scenario: two equal intensity sources centred on an one-dimensional axis. We assume their intensities sum to one, so the average photon number at the object plane is unity as well, $\epsilon_0 = 1$. For coherent light, there's one additional parameter in order to fully characterise the system: the relative phase between the sources. The classical electric field at the image plane then becomes

$$E(x) = \frac{1}{\sqrt{2}}\left[e^{i\Phi}\psi(x-\theta/2) + \psi(x+\theta/2)\right], \tag{5.26}$$

where θ is the separation between the sources and Φ the relative phase. With the electric field $E(x)$ we can evaluate all quantities of interest to us.

The average photon number at the image plane is proportional to (from now on we ignore the quantum efficiency term $\sqrt{\eta}$)

$$\epsilon = \int_{-\infty}^{+\infty} \mathrm{d}x\, |E(x)|^2 = 1 + \mathrm{Re}\Big\{e^{-i\Phi}\Delta(\theta)\Big\}, \tag{5.27}$$

where

$$\Delta(\theta) = \int_{-\infty}^{+\infty} \mathrm{d}x\ \psi^*(x-\theta)\psi(x). \tag{5.28}$$

For a Gaussian PSF, $\Delta(\theta) = e^{-\theta^2/8\sigma^2}$. The value of ϵ as a function of separation and for various relative phases is plotted in Fig. 5.1. For large separations, it is constant, as in the incoherent case. As we approach the subdiffraction regime, however, interference between the sources starts to occur, and ϵ starts to change. When sources are in phase, $\Phi = 0$, they interfere constructively and lead to a higher ϵ. It is important to note here that we do not actually have $\epsilon > \epsilon_0$ (ϵ_0 is the average photon number at the object plane), which would be nonsense: this apparent contradiction happens because we are ignoring the quantum efficiency η, which is usually very small [42, 51]. When sources are in quadrature, $\Phi = \pi/2$, they behave as if they were incoherent [53, 141] and we have a θ-independent ϵ. However, as they become out of phase, they interfere destructively and ϵ starts to decrease, eventually reaching zero for $\Phi = \pi$ and zero separation. However, it is important to note that this limiting case is unphysical, as the sources "become one" at $\theta = 0$, but that is incompatible with a nonzero relative phase.

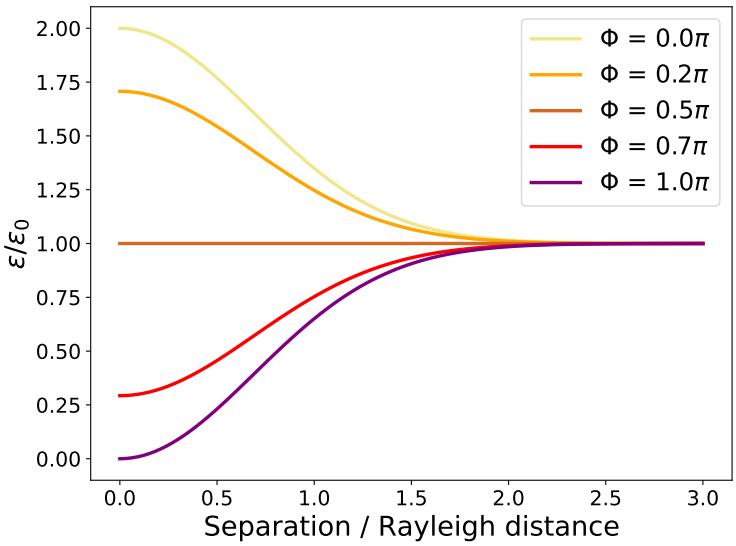


Figure 5.1: Average photon number at the image plane ϵ, normalised by its object plane value ϵ_0, for various values of relative phase of the sources, as a function of their separation.

The quantum Fisher information is, via Eq. (5.15),

$$\mathcal{K}(\theta) = \tau(0) - \mathrm{Re}\Big\{e^{-i\Phi}\tau(\theta)\Big\}, \tag{5.29}$$

where

$$\tau(\alpha-\beta) \equiv \int_{-\infty}^{+\infty} \mathrm{d}x \, \frac{\partial\psi^*(x-\alpha)}{\partial\alpha}\frac{\partial\psi(x-\beta)}{\partial\beta}. \tag{5.30}$$

For a Gaussian PSF (See Appendix F),

$$\tau(\theta) = \frac{1}{4\sigma^2}\left(1-\frac{\theta^2}{4\sigma^2}\right)e^{-\theta^2/8\sigma^2}, \tag{5.31}$$

and the QFI becomes

$$\mathcal{K}(\theta) = \frac{1}{4\sigma^2}\left[1-\cos(\Phi)\left(1-\frac{\theta^2}{4\sigma^2}\right)e^{-\theta^2/8\sigma^2}\right]. \tag{5.32}$$

In order to define the per-photon QFI, should we normalise $\mathcal{K}(\theta)$ per emitted photon or per detected photon? These two cases are shown on the left and right plots, respectively, of Fig. 5.2. Let us discuss them and try to find an answer.

Consider first the case of constructive interference, $\Phi = 0$. As we know, coherence is the enemy of resolution, since it leads to indistinguishability. Therefore, as the sources approach each other the QFI goes to zero, in a resurgence of Rayleigh's curse. When the sources are in quadrature ($\Phi = \pi/2$), they again behave as incoherent: the QFI is constant and the curse is lifted. Finally, for out-of-phase sources ($\Phi = \pi$), the normalisation changes the QFI behaviour. If we normalise by the emitted photons, then the per-photon QFI increases as we enter the sub-Rayleigh regime, eventually becoming twice as much as the incoherent case for zero separation. The reason for this is that out-of-phase sources are highly distinguishable, as could be seen in Fig. 3.5. If we normalise the QFI by the detected photons, however, then it blows up to infinity, which is a very odd thing that has intrigued researchers [27, 32, 41, 45, 53, 69]. This happens because, due to destructive interference, very few photons reach the image plane: we are trying to detect a signal in a dark fringe. If we consider that all the object information is contained only in these photons, then the FI blows up. If we consider the information is evenly spread among all emitted photons, the divergence disappears. Therefore, the controversy is caused by the dependence of ϵ, the average photon number at the image plane, on the separation.

Furthermore, this anomaly also happens because we consider the estimation of separation only and implicitly assume that all other parameters (centroid, relative

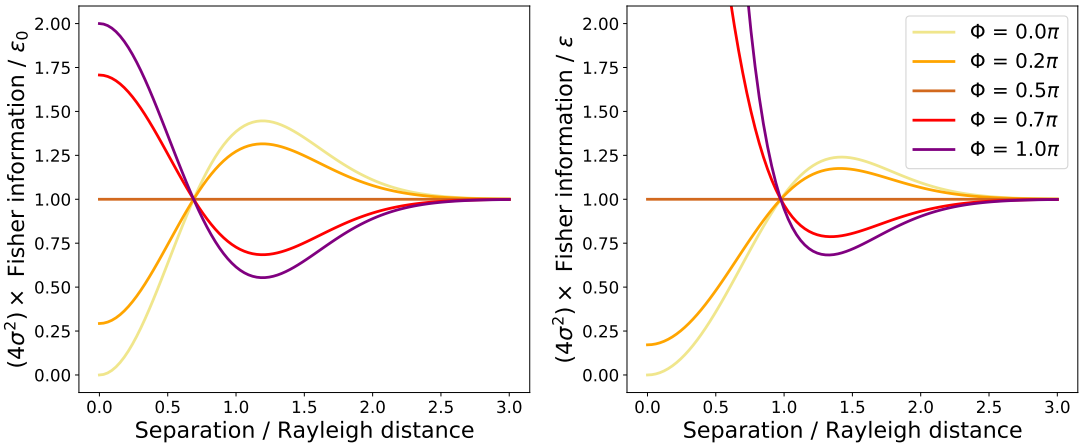


Figure 5.2: Quantum Fisher information for separation estimation, normalised by emitted (left) and detected (right) photons, as a function of separation, for several values of relative phase Φ.

intensities and phase) have been perfectly determined beforehand. However, in a real scenario, as we will see later, we would have to estimate all of them and, while the uncertainty in separation goes to "zero", for other parameters it "blows-up", in a "Heisenberg principle"-like behaviour, and when we consider their complete estimation all things cancel off.

Therefore, we have seen that the blowing-up of the per-photon QFI is an artifact of an unrealistic analysis. For this reason, I personally believe that normalising the QFI by either detected or emitted photons should be fine, and each researcher should choose what is most convenient for them. However, for replicability and overall consistency, everyone should always make sure to emphasise *which* definition they are using.

Finally, we evaluate the Fisher information for several types of measurements. For direct imaging, the FI is

$$\mathcal{F}_{DI}(\theta) = \int_{-\infty}^{+\infty} \mathrm{d}x \left(\frac{\partial |E(x)|^2}{\partial \theta} \right)^2 \frac{1}{|E(x)|^2}. \tag{5.33}$$

This integral can be evaluated numerically. For intensity mode sorting, it is

$$\mathcal{F}_{MS}(\theta) = 2\,\mathrm{Re}\left\{ \sum_{m=0}^{\infty} \frac{\partial E_m^*}{\partial \theta_i} \frac{\partial E_m}{\partial \theta_j} \right\} + 2\,\mathrm{Re}\left\{ \sum_{m=0}^{\infty} \frac{\partial E_m}{\partial \theta_i} \frac{\partial E_m}{\partial \theta_j} \frac{E_m^*}{E_m} \right\}, \tag{5.34}$$

where

$$E_m = \frac{1}{\sqrt{2}}\left[e^{i\Phi}R_m(\theta/2) + R_m(-\theta/2)\right], \tag{5.35}$$

and

$$R_m(\pm\theta/2) = \int \mathrm{d}x\ \phi_m^*(x)\psi(x \mp \theta/2). \tag{5.36}$$

If we work with a real and mirror inversion-symmetric aperture, such as the Gaussian and circular ones, we can use a mode basis with well-defined parity $\phi_m(-x) = (-1)^m\phi_m(x)$ in order to exploit that symmetry. In this case,

$$E_m = \frac{1}{\sqrt{2}}\left[e^{i\Phi} + (-1)^m\right]R_m(\theta/2), \tag{5.37}$$

and it is easy to see that

$$\frac{\partial E_m}{\partial\theta_i}\frac{\partial E_m}{\partial\theta_j} = \frac{\partial E_m^*}{\partial\theta_i}\frac{\partial E_m}{\partial\theta_j}\frac{E_m}{E_m^*}, \tag{5.38}$$

which implies that $\mathcal{F}_{MS}(\theta) = \mathcal{K}(\theta)$, i.e., mode sorting is quantum-optimal for any relative phase is this case. Finally, the FI for quadrature mode sorting is half of the QFI. In Fig. 5.3 we plot the QFI and the direct imaging FI (per emitted photon) for various phase values, ranging from $\Phi = 0$ (yellow) to $\Phi = \pi$ (purple). We didn't plot the FI for intensity and quadrature mode sorting separately due to redundancy, as they are equal to and half the QFI, respectively. Note that direct imaging is close to the quantum limit as the sources are in phase, but as they become out-of-phase its behaviour for small separations changes, namely Rayleigh's curse is still present except for when we exactly have $\Phi = \pi$.

5.3.2 Sources' parameters

Now let's look at the slightly more exciting imaging scenario of several point sources. Consider S sources on an one-dimensional object plane at positions $\{x_i\}$ and with complex amplitudes A_i. The $\mathcal{H}$ matrix for the simultaneous estimation of all intensities $I_i = |A_i|^2$ and positions is

$$\mathcal{H} = \begin{bmatrix} \mathcal{H}_I & \mathcal{H}_{I,x} \\ \mathcal{H}_{I,x}^\dagger & \mathcal{H}_x \end{bmatrix}, \tag{5.39}$$

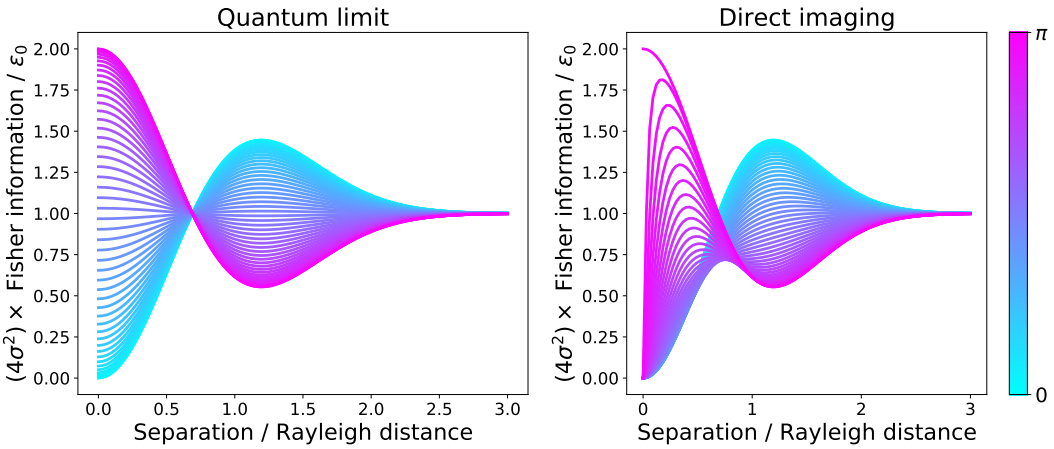


Figure 5.3: (Per emitted photon) Quantum Fisher information (left) and direct imaging Fisher information (right) as a function of separation estimation, for various relative phases. The phases range from $\Phi = 0$ (blue) to $\Phi = \pi$ (pink).

with submatrices given by

$$[\mathcal{H}_I]_{ij} = \frac{\Delta(x_i - x_j)}{A_i A_j^*}, \tag{5.40a}$$

$$[\mathcal{H}_x]_{ij} = 4A_i^* A_j \tau(x_i, x_j), \tag{5.40b}$$

$$[\mathcal{H}_{I,x}]_{ij} = -2\frac{A_j}{A_i}\gamma(x_i - x_j), \tag{5.40c}$$

where $\Delta(\alpha) = \int_{-\infty}^{\infty} \psi^*(x-\alpha)\psi(x)\mathrm{d}x$ is the overlap between sources at the image plane,

$$\gamma(\alpha) = \int_{-\infty}^{\infty} \psi^*(x-\alpha)\frac{\partial\psi(x)}{\partial x}\mathrm{d}x \tag{5.41}$$

and

$$\tau(\alpha,\beta) = \int_{-\infty}^{\infty} \frac{\partial\psi^*(x-\alpha)}{\partial\alpha}\frac{\partial\psi(x-\beta)}{\partial\beta}\mathrm{d}x. \tag{5.42}$$

We assume a Gaussian PSF $\psi(x) = \exp(-x^2/4\sigma^2)/(2\pi\sigma^2)^{1/4}$ with again $\sigma = 0.21\lambda/\mathrm{NA}$. In this case, the (coherent light) Rayleigh distance is $d_R = 0.82\lambda/\mathrm{NA} \simeq 3.9\sigma$. For the Gaussian PSF, we have that (See Appendix F) the overlap is

$$\Delta(x) = \exp(-x^2/8\sigma^2), \tag{5.43}$$

the $\gamma(x)$ term is

$$\gamma(x) = (-x/4\sigma^2)\Delta(x), \tag{5.44}$$

and

$$\tau(x) = \frac{1}{4\sigma^2}\left(1 - \frac{x^2}{4\sigma^2}\right)\Delta(x). \tag{5.45}$$

For visualization purposes, we consider five sources distributed inside an interval $[-W/2, W/2]$, where W is the total support of the sources' distribution. By varying W, we can evaluate how the parameter estimation task behaves as the object enters the subdiffraction regime. For simplicity, we also assume that all intensities sum up to unity. We use as figure of merit our beloved (per detected photon) total precision (Eq. (4.90))

$$\mathcal{V}/\epsilon = 1/\mathrm{trace}[(\mathcal{K}/\epsilon)^{-1}]. \tag{5.46}$$

Remember that ϵ is a quantity that depends not only on the intensities of the sources but also, due to interference effects, on their positions and relative phases.

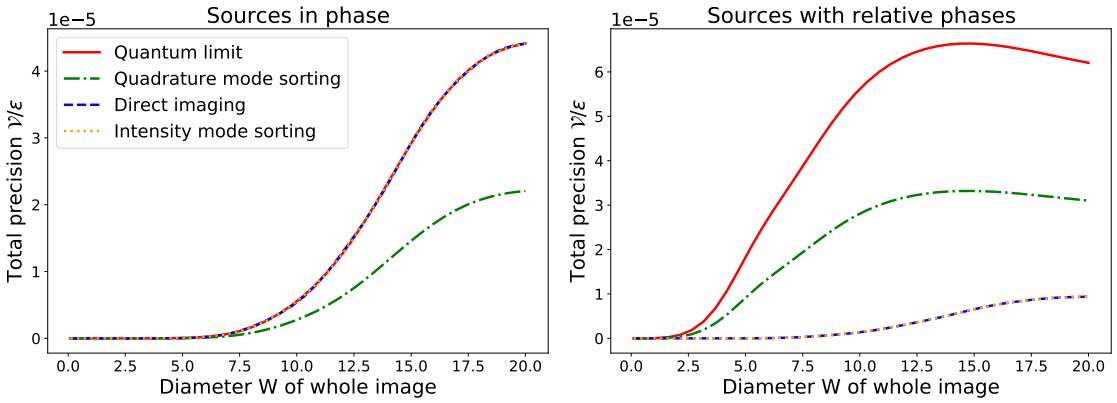


Figure 5.4: (per detected photon) Total precision for estimation of positions, for full quadrature measurement (dash-dotted green curve), direct intensity imaging (dashed blue curve) and mode sorting with Hermite-Gaussian modes followed by intensity detection (dotted orange curve), and the corresponding quantum limit (solid red curve), as a function of the object diameter W. We choose $\sigma = 1$, so that the Rayleigh limit is equal to 3.9. The amplitudes and positions of the sources were chosen randomly. On the left, all phases are zero, while on the right they are randomly picked between 0 and 2π.

When all sources are in phase, i.e., the relative phases $\Phi_i - \Phi_j$ are all zero, the compatibility matrix Γ is identically zero and the QFI $\mathcal{K}$ can be achieved. In fact, it is saturated by any mode resolved intensity measurement. Therefore, in this situation, intensity measurements are clearly desirable over full quadrature ones, which can achieve at most half of the quantum limit. This behaviour can be seen on

the left plot of Fig. 5.4, where the total precision of the estimation of the positions of the sources is plotted as a function of the object diameter W. Physically, this happens because the intensity measurements act directly on the light field, while the quadrature measurements involve interference with a local oscillator, which brings its own noise into play and degrades the achievable precision.

When there are nonzero phases, however, the Γ matrix is no longer zero and there is no measurement that can exactly achieve the quantum bound. As can be seen on the right of Fig. 5.4, where random phases were assigned to the sources, the previously optimal intensity measurements are now much weaker. The second term on the right hand side of Eq. (5.20), the only one that depends on the actual basis of modes chosen for the demultiplexing, is responsible for that. On the right of Fig. 5.4, intensity measurements for mode sorting and direct imaging coincide, but that doesn't need to be true. Meanwhile, the quadrature measurements are much more desirable now, even though they cannot achieve the quantum limit. Physically, this can be simply explained by the fact that they are sensitive to the image-plane wavefront, while intensity measurements are not. This advantage is more than enough to compensate the drawback of the added local oscillator noise.

Beyond their differences, all measurements and the quantum limit suffer from a degradation in precision as they venture further into the sub-wavelength domain, which is a manifestation of Rayleigh's curse. For the QFI, it stems from its nonzero off-diagonal terms, dominated by the overlap $\Delta(x_i - x_j)$. For large distances, it approximates a delta function and the sources can be independently measured, but for sub-Rayleigh distances it gets wider and "spreads" over the off-diagonal terms, decreasing the overall precision.

5.4 Conclusion

In summary, by deriving the QFI and compatibility matrix for the general imaging of a distribution of coherent light sources, we were able to identify the ultimate quantum limit in estimation precision. We showed that, in the absence of relative phases between the sources, this limit can be achieved by intensity

measurements. In their presence, however, the limit is no longer attainable, but full quadrature measurements, as they are sensitive to phases, can achieve half of it.

Future research on this topic can include a more complete statistical analysis of the problem: studying the effect of biases, using the relative phases as nuisance parameters, estimating the effects of measurement noise, among others. Our results could also be generalized to 3D imaging scenarios and for partially coherent light sources, which include coherent and incoherent sources as special cases.

6 | Spatial modulation of light

All the quantum inspired super-resolution techniques previously introduced involve in some way or another manipulating the transverse spatial profile of light. The devices that perform such transformations are called *spatial light modulators* (SLM). In this chapter we study phase modulation SLMs, which are the ones used in our microscopy experiments. In Section 6.1 we review how they work, and how they can be used for simultaneously encoding amplitude and phase. Then, in Section 6.2, which contains our work of Ref. [110], we report the modelling and performance optimization of a liquid crystal on silicon (LCoS) phase-only SLM. An accurate and well-calibrated SLM is crucial for the realisation of any QISR measurement.

6.1 Phase-only modulators

6.1.1 Diffraction gratings

A phase-only spatial light modulator is a device that can arbitrarily shape the wavefront of a light beam. Consider an electric field with transverse spatial profile $E_{\text{in}}(x)$ incident on a reflective SLM surface. The beam reflected off the SLM will have a spatial profile

$$E_{\text{out}}(x) = E_{\text{in}}(x)e^{i\psi(x)}, \tag{6.1}$$

where $\psi(x) \in [0, 2\pi]$ is the SLM's phase mask or *hologram*. The extension of this theory to two dimensional transversal planes is straightforward. For simplicity, from now on we also consider an uniform illumination beam, $E_{\text{in}}(x) \equiv 1$.

Our goal is to perform a completely arbitrary phase *and* amplitude modulation of light. There are schemes which can do that with several modulators [152], but

these are hard to align and obviously more costly. A simpler and easier approach can be done with a single phase-only SLM. In order to see how that is possible, consider a blazed phase grating hologram

$$\psi(x) = \mathrm{mod}\left(\frac{2\pi x}{\Lambda}\right), \tag{6.2}$$

where Λ is the period of the grating and $\mathrm{mod}(x)$ is the remainder of the division of x by 2π. Since 2π shifts in the phase are not observable, what the incoming beam on the SLM "sees" is a perfect mirror, as shown Fig. 6.1, with slope $2\pi/\Lambda$. The far-field outgoing beam, which is the Fourier transform of $E_{\mathrm{out}}(x)$, is simply

$$E_{\mathrm{far}}(x) = \delta(x - 1/\Lambda). \tag{6.3}$$

This means that the grating works, as expected, as a mirror, simply displacing the beam by a distance $1/\Lambda$ at the far-field plane.

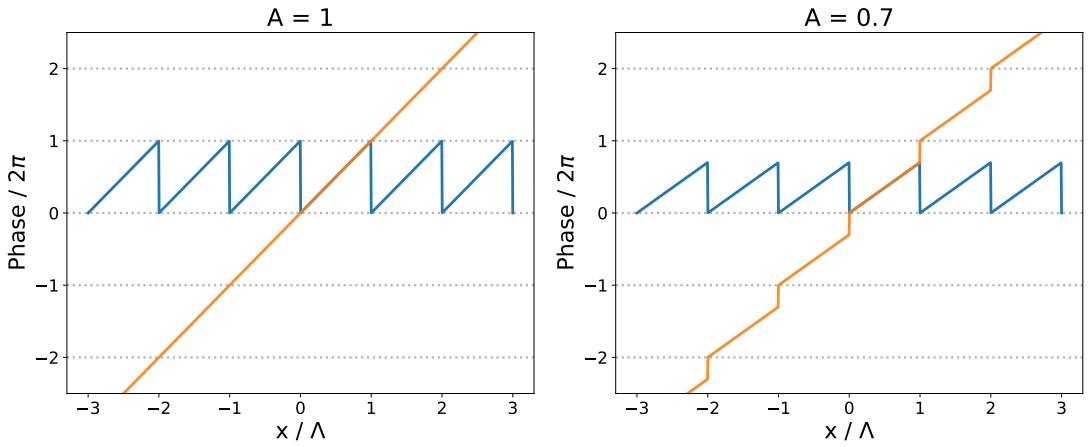


Figure 6.1: Phase grating (solid blue curve) and its "phase-unwrapped" version (solid orange curve) for grating amplitude $A = 1$ (left) and $A = 0.7$ (right).

Now consider a slightly different grating,

$$\psi(x) = A\ \mathrm{mod}\left(\frac{2\pi x}{\Lambda}\right), \tag{6.4}$$

where $A \in [0, 1]$ is the *grating amplitude*. Although it looks like the same grating, what the field "sees" is now a "slated" mirror, as shown again in Fig. 6.1. These

"defects" make the mirror not work perfectly, with light diffracted to other directions. In order to see that, let's calculate the far-field beam. Using the mod property

$$\mathrm{mod}\,(x) = x - 2n\pi,\ x \in [2\pi n, \pi(n+1)] \tag{6.5}$$

and the *Dirac comb* definition,

$$\Lambda \sum_{n=-\infty}^{+\infty} e^{-in2\pi x\Lambda} = \sum_{n=-\infty}^{+\infty} \delta(x - n/\Lambda), \tag{6.6}$$

the far-field outgoing beam is

$$E_{\mathrm{far}}(x) = e^{-i\pi\Lambda(x-A/\Lambda)}\,\mathrm{sinc}(\Lambda(x - A/\Lambda)) \sum_{n=-\infty}^{+\infty} \delta(x - n/\Lambda), \tag{6.7}$$

whose intensity is plotted in Fig. 6.2. Now we have beams at all positions n/Λ, which are known as the n-th orders of diffraction. The further A is from unity, the more power is distributed from the first order to all other orders. The zeroth, first and second order intensities as a function of A are shown in Fig. 6.3.

The slated grating example gives us a hint on how a phase-only SLM can also modulate the amplitude of the field in a spatially-resolved fashion. By adding a grating to a phase hologram, the amplitude of the grating itself can be spatially modulated in order to control the distribution of power between diffraction orders. Therefore, if we are interested in a specific order, say, the first one, we can easily block the other ones with an iris and control its amplitude profile by the grating modulation. The only drawback is the loss of power, which is a reasonable trade-off.

6.1.2 Phase and amplitude modulation

Suppose we have an uniform incoming beam hitting our SLM. We want to generate the target field

$$E_{\mathrm{target}}(x) = A(x)e^{i\phi(x)}, \tag{6.8}$$

where $A(x)$ and $\phi(x)$ are the desired amplitude and phase profiles. In order to do that, we put on the SLM the hologram

$$\psi(x) = M(x)\ \mathrm{mod}\ (F(x) + g(x))\,, \tag{6.9}$$

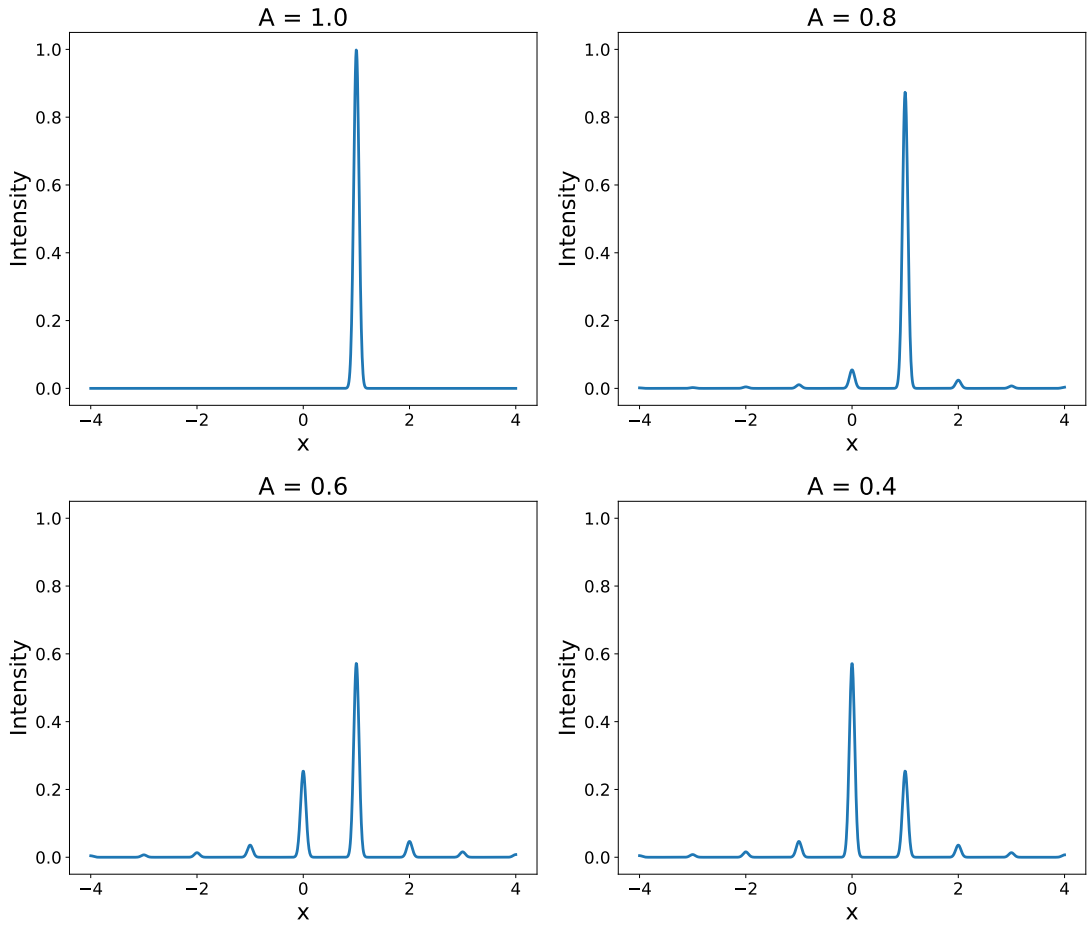


Figure 6.2: Far-field intensity distributions for different values of grating amplitude A. We used $\Lambda = 1$. For an easier visualisation, we considered a Gaussian incoming beam (with $\sigma = 0.05$), which is convoluted with the far-field result.

where $M(x) \in [0, 1]$ and $F(x) \in [0, 2\pi]$ are functions to be determined and $g(x)$ is a grating. We then let the outgoing beam propagate to the Fourier plane, where the diffraction orders will spatially separate, as shown in Fig. 6.4. We want to encode the target field in the N-th order of diffraction, so we use an iris to block all other orders. We then let the remaining beam propagate to the image plane, where it can be detected or manipulated as we want. By choosing suitable functions $M(x)$ and $F(x)$, we can have an arbitrarily shaped spatial profile on that beam in the image plane, with only some loss in power.

Recall the top-hat function $\Pi(x)$, which is unity for $x \in [0, 1]$ and zero otherwise. The function $\Pi(x/\Lambda - n)$ covers the interval $[n\Lambda, (n + 1)\Lambda]$ and can be used as

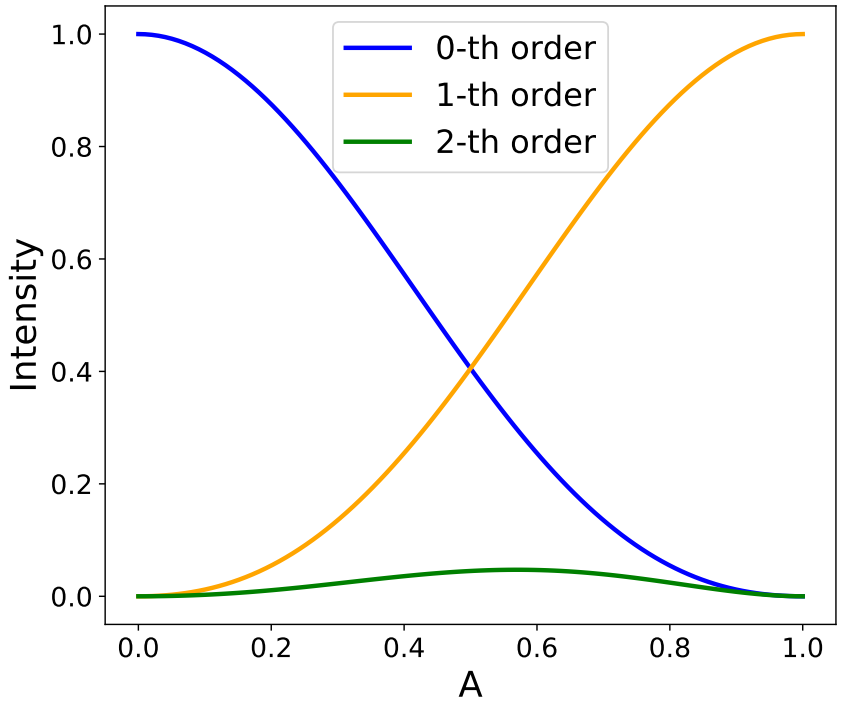


Figure 6.3: Intensity of the zeroth (solid blue curve), first (solid orange curve) and second (solid green curve) orders of diffraction beams as a function of the grating amplitude A.

a resolution of the identity,

$$1 = \sum_{n=-\infty}^{+\infty} \Pi(x/\Lambda - n). \tag{6.10}$$

Consider also a blazed grating $g(x) = \text{mod}\left(\frac{2\pi x}{\Lambda}\right)$. We assume now that the grating period Λ is very small and the $M(x)$ and $F(x)$ functions are slowly-varying over it, which means that they can be considered constant over each grating period: for $x \in [n\Lambda, (n+1)\Lambda]$, we have $M(x) = M(n\Lambda) \equiv M(x_n)$ and $F(x) = F(x_n)$. This means that inside each grating interval the hologram $\psi(x)$ is constant except for the grating, which implies that the outgoing field right after the SLM screen $E(x)$ is periodic over each grating interval.

Using the resolution of the identity, the $E(x)$ field is

$$E(x) = \sum_{n=-\infty}^{+\infty} E(x)\Pi(x/\Lambda - n). \tag{6.11}$$

Now, consider the family of fields $\{E_n(x)\}$, which are equal to $E(x)$ inside $[n\Lambda, (n+$

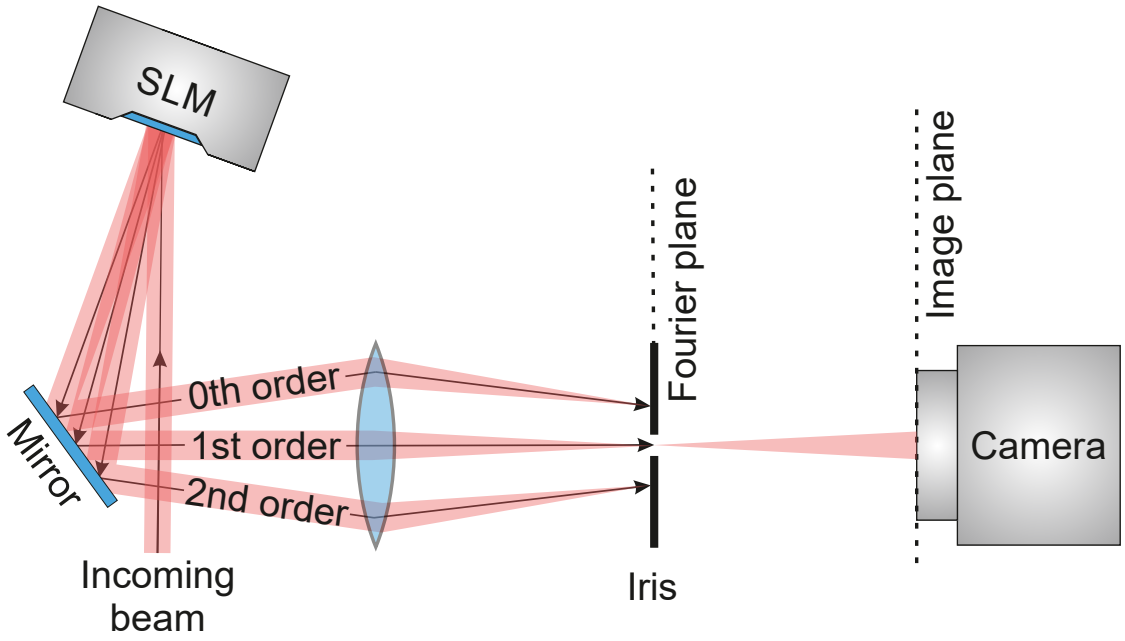


Figure 6.4: Depiction of the complex amplitude modulation scheme. A chosen order of diffraction coming from the SLM is spatially filtered in the Fourier plane, where all other orders are blocked. It is then propagated to the image plane, where it can be detected by a camera or used for something else.

$1)\Lambda]$ but repeat periodically outside that interval. Using this trick, we have

$$E(x) = \sum_{n=-\infty}^{+\infty} E_n(x)\Pi(x/\Lambda - n), \tag{6.12}$$

but now each

$$E_n(x) = e^{iM(x_n) \bmod (F(x_n)+g(x))} \tag{6.13}$$

is a periodic function and, therefore, can be Fourier expanded,

$$E_n(x) = \sum_{k=-\infty}^{+\infty} c_k^{(n)} e^{i2\pi kx/\Lambda}, \tag{6.14}$$

where

$$c_k^{(n)} = \frac{1}{\Lambda}\int_0^{\Lambda} \mathrm{d}x\ E_n(x)e^{-i2\pi kx/\Lambda}. \tag{6.15}$$

The total field can then be rewritten with a reordering of the sums,

$$\begin{aligned} E(x) &= \sum_{n=-\infty}^{+\infty} \left[\sum_{k=-\infty}^{+\infty} c_k^{(n)} e^{i2\pi kx/\Lambda} \right] \Pi(x/\Lambda - n) \\ &= \sum_{k=-\infty}^{+\infty} \left[\sum_{n=-\infty}^{+\infty} c_k^{(n)} \Pi(x/\Lambda - n) \right] e^{i2\pi kx/\Lambda} \\ &\equiv \sum_{k=-\infty}^{+\infty} E_{\text{appr}}^k(x) e^{i2\pi kx/\Lambda}. \end{aligned} \tag{6.16}$$

This final expression shows that, as $E(x)$ propagates to the far-field, the terms $e^{i2\pi kx/\Lambda}$ become $\delta(x - k/\Lambda)$, which means that the field at each k-th order of diffraction is $E_{\text{appr}}^k(x)$. We are interested in the N-th order, so we physically block all others and let it propagate again, so the second Fourier transform "cancels" the first and we are left with $E(x) = E_{\text{appr}}^N(x)$. Since we want that field to be our target field, our method amounts to finding functions $M(x)$ and $F(x)$ that satisfy the equation

$$A(x)e^{i\phi(x)} = E_{\text{appr}}^N(x). \tag{6.17}$$

Note, however, that $E_{\text{appr}}^N(x)$ is made of step-functions and can only approximate the target field, as shown in Fig. 6.5. We therefore only require Eq. (6.17) to be exactly satisfied for every x in the limit $\Lambda \to 0$. It suffices for us that it is satisfied at all points of the discrete grid $\{x_n = n\Lambda\}$, where $E_{\text{appr}}^N(x) = c_N^{(n)}$, which amounts to inverting the equation

$$\frac{1}{\Lambda} \int_0^\Lambda \mathrm{d}x \; e^{iM(x_n) \bmod (F(x_n)+g(x))} e^{-i2\pi Nx/\Lambda} = A(x_n)e^{i\phi(x_n)}. \tag{6.18}$$

In order to solve Eq. (6.18), first note that, inside the integration interval $[0, \Lambda]$, the grating is simply $g(x) = 2\pi x/\Lambda$, which means that we can use the change of variables $F(x_n) + g(x) = u$ to get

$$\frac{e^{iF(x_n)}}{2\pi} \int_{F(x_n)}^{F(x_n)-2\pi} \mathrm{d}u \; e^{iM(x_n) \bmod (u)} e^{-iNu} = A(x_n)e^{i\phi(x_n)}. \tag{6.19}$$

Furthermore, note that, since both $F(x_n)$ and $g(x)$ belong to the interval $[0, 2\pi]$, we have that $u \in [0, 4\pi]$. The mod function then implies that, for $u \in [0, 2\pi]$, $\mathrm{mod}(u) = u$ and, for $u \in [2\pi, 4\pi]$, $\mathrm{mod}(u) = u - 2\pi$. We can use that to split

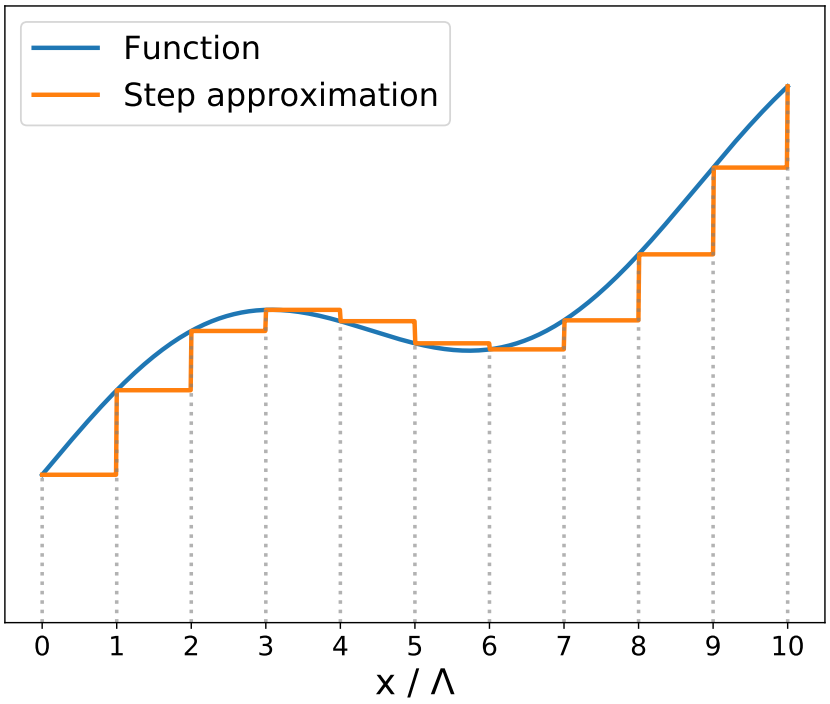


Figure 6.5: Arbitrary function (solid blue curve) and its step-wise approximation (solid orange curve). As the grating period Λ becomes smaller, the approximation gets closer to the real function.

the integration interval into $[F(x_n), 2\pi]$ and $[2\pi, F(x_n) + 2\pi]$, and after some algebra we reach

$$e^{i[\pi(M(x_n)-N)+NF(x_n)]}\,\mathrm{sinc}(M(x_n)-N) = A(x_n)e^{i\phi(x_n)}. \tag{6.20}$$

This equation can be exactly inverted and, extrapolating the discrete grid $\{x_n\}$ to the whole real axis, leads to the conditions [153]

$$M(x) = N + \mathrm{sinc}^{-1}(A(x)), \tag{6.21}$$

$$F(x) = [\phi(x) - \pi(M(x) - N)]/N. \tag{6.22}$$

If Eqs. 6.21 and 6.22 are satisfied, the resulting beam of the setup in Fig. 6.4 will approximately have the target field as its complex transverse spatial profile, with the approximation becoming exact in the limit $\Lambda \to 0$. Eqs. (6.21) and 6.22 for $N = 1$ are the celebrated result of Ref. [153]. In Fig. 6.6 we show an example of such holograms.

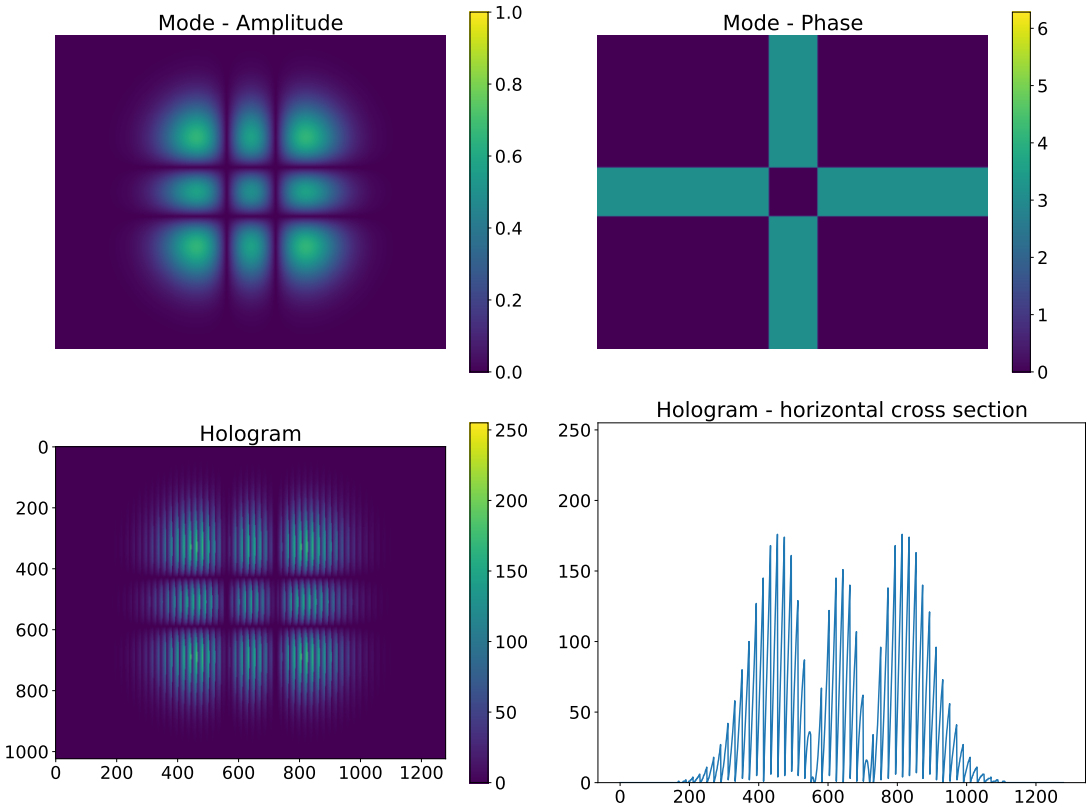


Figure 6.6: Amplitude (upper left) and phase (upper right) profiles of the $HG_{2,2}$ mode. Its corresponding hologram (lower left) on a 1280×1024 SLM screen with 256 gray levels and a 20 pixel period grating, and its central horizontal cross section (lower right).

Our solution derived so far requires a perfect SLM, i.e., one that exactly implements the transformation of Eq. (6.1). In practice, that is usually not the case, and we might have to consider a general

$$E_{\text{out}}(x) = E[\psi](x), \tag{6.23}$$

where the brackets $[\cdot]$ denote a functional dependence. Now, define the function

$$G[M(x_n), F(x_n)] \equiv \frac{1}{\Lambda}\int_0^{\Lambda} \mathrm{d}x \; E[M(x_n), F(x_n)](x) e^{-i2\pi N x/\Lambda}. \tag{6.24}$$

In order to have our desired field, we have to solve

$$G[M(x_n), F(x_n)] = A(x_n) e^{i\phi(x_n)}. \tag{6.25}$$

Eq. (6.25) can be inverted numerically as follows. First, set up a sufficiently fine grid of points $\{x_n\}$ covering the transverse plane. For each of them, evaluate the

LHS of Eq. (6.25) for a high enough number of possible combinations of M and F, sampled from the interval $\Sigma = [0, 1] \times [0, 2\pi]$. We then have a look-up table for each point x_n, and the solution will be

$$M(x_n), F(x_n) = \arg \min_{M', F' \in \Sigma} \left| G[M', F'] - A(x_n) e^{i\phi(x_n)} \right|. \tag{6.26}$$

6.2 Modelling and calibration of LCoS SLMs

6.2.1 Operating principle of LCoS SLMs

A reflective, phase-only, liquid crystal on silicon spatial light modulator (LCoS SLM) consists of a birefringent liquid crystal (LC) screen which is electrically addressed by a silicon back panel. The voltage applied by the panel changes the orientation of the LC molecules, which affects their refraction index. This, in turn, changes the optical path traversed by light. Since every pixel electrode in the back panel applies an independent voltage to the LC layer right above it, we are able to perform an arbitrary $\phi(x, y)$ transformation to the light's wavefront. It is important to note that only the light polarized along the SLM's *director axis* is modulated, while the orthogonal polarization remains unaffected.

The phase shift experienced by the light depends on the voltage at the back panel, the LC layer's thickness and its own wavelength. The voltage is controlled by a computer-generated matrix, whose values are known as *gray levels*. They usually range from 0 to 255 (8-bit encoding) or from 0 to 1023 (10-bit). The relationship between gray levels and phase shift is provided by manufacturers in the form of a look-up table, and is usually linear. Since it depends on the light wavelength, it is either provided for several λ values or the SLM is simply designed to be used with a single, fixed wavelength. The gray level that induces a 2π phase shift is known as the 2π *voltage* $A_{2\pi}$. It is expected that SLMs provide, at least, a full 2π phase shift. This also depends on the LC layer's thickness: the thicker, the wider the phase shift range. However, thicker layers also come with their own host of problems (as will be seen in Section 6.2.2), so a balance must be found by manufacturers.

6.2.2 Imperfections

Am ideal phase-only SLM should implement the transformation

$$E_{\text{out}}(x, y) = E_{\text{in}}(x, y)e^{i\psi(x,y)}. \tag{6.27}$$

In practice, there are unavoidable imperfections which change this relationship. In this section we study how these imperfections come to be, how they affect the SLM's behaviour, and how they can be compensated.

Back panel curvature

The layered structure [154] of a reflective phase-only SLM is schematized in Fig. 6.7. Due to an unavoidable manufacturing issue, the SLM back panel is almost always curved to some degree [155–157]. This curvature implies in a non-uniform thickness of the LC layer, which leads to a spatially dependent extra phase retardation term $\alpha(x, y)$, as in

$$E_{\text{out}}(x, y) = E_{\text{in}}(x, y)e^{i(\psi(x,y)+\alpha(x,y))}. \tag{6.28}$$

Due to its strong effect on the modulation, the back panel curvature is the most known SLM imperfection. There are several good methods to characterise and compensate it [155–158], and its (wavelength-dependent) correction is usually provided by manufacturers. For a phase-only modulation experiment, the curvature can be simply subtracted from the hologram. For our complex amplitude modulation scheme, Eq. 6.22 can be modified to

$$F(x, y) = [\phi(x, y) - \pi(M(x, y) - N) - \alpha(x, y)]/N. \tag{6.29}$$

Inner cavities

The multilayered structure of the SLM implies in Fabry-Pérot cavities inside its screen [158, 159], as shown in Fig. 6.7. This can lead to undesirable interference effects, as the beams coming from the cavities overlap with each other. In order to quantify them, consider the simplified three-layer model of Fig. 6.8. The terms r_1, r_2 and r_3 correspond to the air-cover glass, cover glass-LC and LC-back panel interfaces

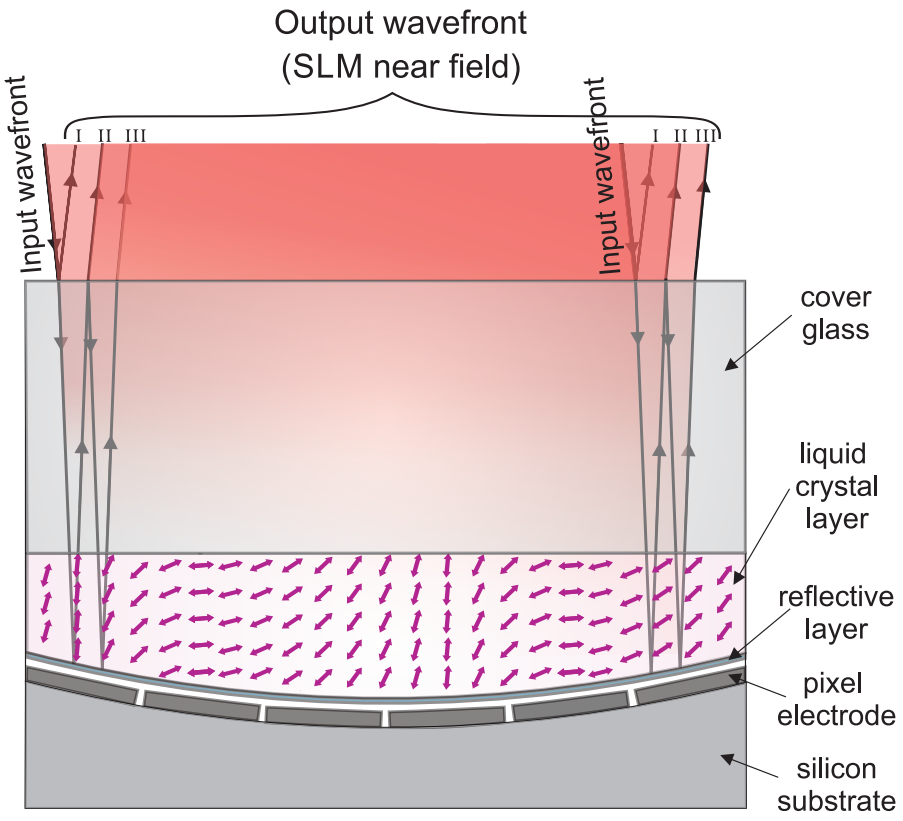


Figure 6.7: Schematic cross-section of a phase-only SLM illuminated by a wide laser beam. I, II, III outline the wavefronts produced by the Fabry-Perot cavity formed by the air-glass interface and the backpanel.

reflectivities, respectively, and t_i are the transmissivities such that $r_i^2 + t_i^2 = 1$. For simplicity, consider that everything happens above a single electrode pixel, which is reasonable since SLMs are usually operated with incoming beams E_{in} perpendicular to the screen. Finally, note that $n_3 > n_2 > n_1$, which means that extra π phase shifts only happen in reflections from above (in the orientation of Fig. 6.8).

The first beam,

$$E_0 = (e^{i\pi} r_1) E_{\text{in}}, \tag{6.30}$$

simply bounces off from the glass cover and doesn't even interact with the liquid crystal. The second outgoing beam is

$$E_1 = (t_1)(e^{i\pi} r_{23})(e^{i\phi_1})(t_1) E_{\text{in}}, \tag{6.31}$$

where r_{23} accounts for the interference that happens in the second cavity and ϕ_1 is the optical path difference accumulated during a "round-trip" through the glass

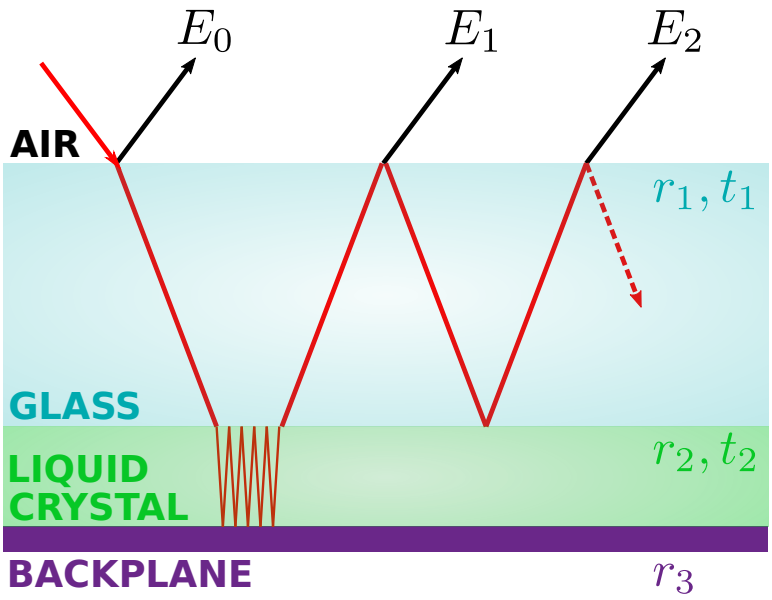


Figure 6.8: Simplified model of the SLM's layered structure.

layer. From that, we can see that the n-th beam will be

$$E_n = (t_1)^2 (e^{i\pi} r_{23} e^{i\phi_1})^n (r_1)^{n-1} E_{\text{in}}. \tag{6.32}$$

Using the geometric sum $\sum_{n=0}^{\infty} x^n = (1-x)^{-1}$ (when $|x| < 1$), the net field will be

$$E_{\text{out}} = \sum_{n=0}^{\infty} E_n = -\frac{r_1 + r_{23} e^{i\phi_1}}{1 + r_1 r_{23} e^{i\phi_1}} E_{\text{in}}. \tag{6.33}$$

Likewise, the effect of the second cavity will be expressed in r_{23},

$$r_{23} = -\frac{r_2 + r_3 e^{i\phi_2}}{1 + r_2 r_3 e^{i\phi_2}}. \tag{6.34}$$

For most SLMs, it is safe to assume that $r_3 = 1$ and ϕ_1 is constant along x and y. For an ideal SLM, $r_1 = r_2 = 0$, which means that $E_{\text{out}} = e^{i(\phi_1+\phi_2)} E_{\text{in}}$, where the phase ϕ_1 is simply a global phase and ϕ_2 corresponds to the desired SLM modulation ψ. In practice, r_1 is very small [159] but not negligible. The same happens with r_2 but, since the LC layer is significantly thinner (around 20 μm) than the cover glass, its net effect is mostly a small amplitude modulation of the final beam. That effect is usually already accounted for in manufacturer's look-up tables, which correct the small non-linearities of the grey level to phase

relation. Finally, the back panel curvature is reflected in the LC layer optical path, yielding $\phi_2(x,y) = \psi(x,y) + \alpha(x,y)$.

At a sufficiently high incidence angle and narrow incoming light, beams I, II, III (and higher-order ones), as per Fig. 6.7), will be physically displaced from each other. This is an easy and unambiguous way to see and confirm the cavity effects. Ref. [159] used this trick to measure the cavity parameters directly. However, this method is only able to find the reflection coefficients for high incidence angles. For small angles, which is the usual domain of operation with SLMs, it is not very useful.

Compensation of the cavity effects can be done by Eq. (6.26), which can be solved numerically.

Pixel crosstalk

In Fig. 6.7, a 6-pixel period blazed grating is printed on the SLM. The LC molecules, however, do not align perfectly to the voltage. This is a manifestation of *pixel crosstalk* [160–163], where the molecules are slightly aligned with their neighbours. This is due to the viscosity of the LC material and the gradual change of the electric field applied to the electrodes[1], which impede the refraction index of undergoing abrupt changes. The net effect of pixel crosstalk is a smoothing of the overall phase distribution, which is an issue for blazed gratings.

The pixel crosstalk can be modelled by a convolution of the ideal phase profile with a kernel called *crosstalk PSF*, usually considered as a generalised Gaussian function

$$g(x,y) \propto e^{-\left(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right)^{\gamma}}, \tag{6.35}$$

where γ is an integer. In our calibration work, we consider $\gamma = 1$ and $\sigma_x = \sigma_y = \sigma$. The latter ranges from a fraction of one pixel to several pixels, depending on the SLM model [163, 164], and quantifies the strength of the effect. The resulting phase on the SLM then becomes

$$\psi(x,y) = \theta(x,y) * g(x,y) + \alpha(x,y), \tag{6.36}$$

[1]For this reason, this effect is also known as the *fringing field* effect.

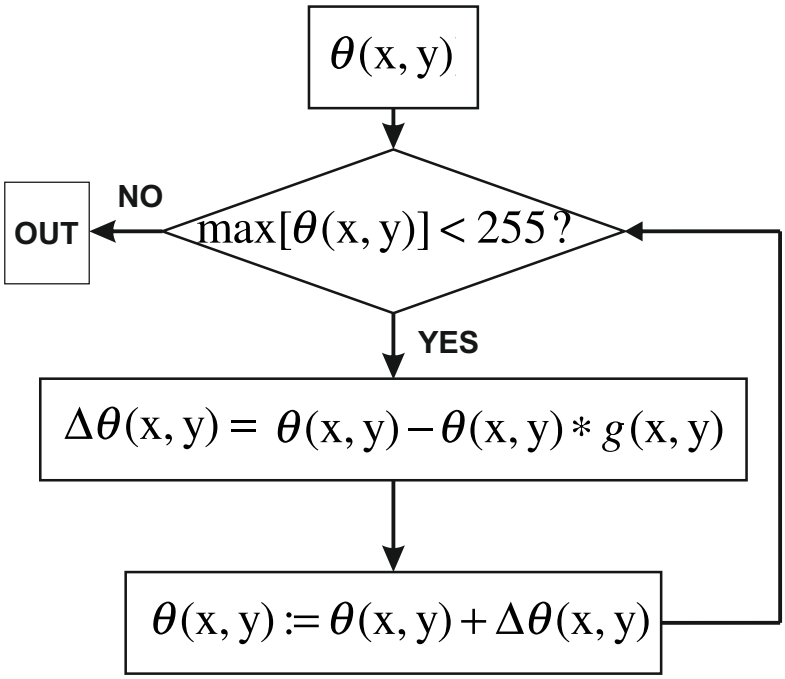


Figure 6.9: Algorithm used to calculate an SLM pattern compensating for crosstalk.

where $\theta(x, y)$ is the printed phase. The crosstalk parameters can be found by fitting the first order intensity distributions as a function of the grating amplitude [162].

To compensate for the crosstalk effect, we modify the hologram $\theta(x, y)$ applying the iterative algorithm shown on Fig. 6.9, where $g(x, y)$ is the fitted crosstalk Gaussian kernel. The iterations have to be stopped when the new hologram values are about to exceed the available range of phase modulation. Compared to other crosstalk compensation methods, this approach is not restricted to a specific type of holograms [162], does not reduce the spatial resolution [165] or involve complicated modelling [166].

6.2.3 Experiment

In previous works, pixel crosstalk [162, 163], the backpanel curvature [155–158] and the cavity [159] have been studied as individual phenomena, but their joint influence on beam modulation has not been investigated. Yet, since these effects act simultaneously, neglecting one of them leads to imprecise estimation of the others, hindering their correct compensation. In this section, we propose [110] a simple characterization procedure, which relies on exploiting the interference

occurring within the SLM's inner cavity and hence avoids the need for an external interferometer [157, 158], which would require careful alignment. Its further advantage is that the measurement results are specific to the SLM and not significantly affected by the external optics or its alignment.

Our method consists in generating blazed grating holograms of varying amplitude and measuring the position-dependent image-plane intensity of the diffracted light in the first and second orders as a function of the grating amplitude. The second order is crucial for the accurate prediction of the model parameters, since it is more sensitive to pixel crosstalk and inner cavity effects than the first order.

Fig. 6.4 illustrates the optical setup used to characterize the SLM. We illuminate the whole screen of a reflective phase-only LCoS-SLM (Hamamatsu X13138-02) using a continuous laser beam at 785 nm. The angle between the incoming and reflected beams is less than 5°. The polarization of the incoming beam is linear and aligned with the liquid crystal extraordinary axis. The SLM resolution is 1272×1024 pixels and the pixel pitch is 12.5 μm. At each pixel, the SLM is calibrated by the manufacturer to have a linear phase response to the gray levels. No other manufacturer-issued correction information is used aside from assuming that the phase-to-gray-level response of the liquid crystal layer is linear.

To characterize the spurious effects, we print on the SLM screen a blazed grating hologram of a 20-pixel period, varying its amplitude from 0 to 248 gray levels in steps of 2. The reflected beam is focused by a 2-inch aperture lens ($f = 250$ mm) and subsequently re-imaged by another lens onto a CCD camera (UI-2140SE). An iris diaphragm in the focal plane of the first lens selects either the first or the second diffraction order.

Fig. 6.10(a) shows the near-field intensity images for three different values of grating amplitude. We notice an annular structure in the intensity profiles, which is necessarily a sign of the previously mentioned Fabry-Perot cavity effect. Specifically, it is a result of interference of the fields reflected from the SLM after different number of passes through the liquid crystal layer of spatially varying thickness,

marked as I, II, and III in Fig. 6.7. For high grating amplitudes, this structure is much more pronounced in the second order than in the first one.

Intuitively this can be understood as follows. Field I is always emitted into the zeroth order. Because of the low reflectivity of the air-glass interface, beam II is generally much stronger than beam III. But when the grating voltage is 2π, the effect of the grating upon the single-pass beam (II) becomes equivalent to that of a plane mirror directing the entire optical field towards the first order. By the same token, the double-pass beam (III) experiences a grating of amplitude 4π, so all of its energy goes into the second order. As a result, in the neighborhood of that voltage the amplitudes of wavefronts II and III in the second order become comparable with their relative phase dependent on the liquid crystal layer thickness.

This effect is further visible in Fig. 6.10(b), where we plot the integrated intensity response for the three 20×20 sections of the SLM screen centered as marked in Fig. 6.10(a). The second order intensity is low, with the behavior strongly dependent on the cavity thickness, for the grating amplitudes around the 2π voltage (~ 200 gray levels), resulting in well-defined rings.

6.2.4 Results

The field distribution Eq. (6.36) depends upon the following set of parameters: $\{r, \sigma, A_{2\pi}(x,y), \alpha(x,y)\}$. We assume that r and σ are spatially independent, while $\alpha(x,y)$ and the 2π voltage $A_{2\pi}(x,y)$ are functions of transverse position. The voltage applied to the SLM corresponds to a blazed grating with its lines along the y axis:

$$\theta(x,y) = \frac{A}{A_{2\pi}} \cdot \mathrm{mod}\left(\frac{2\pi x}{\Lambda}, 2\pi\right), \tag{6.37}$$

where A and Λ are the grating amplitude and period, respectively.

We numerically calculate the Fourier transform of Eq. (6.36) and obtain the theoretical response curve of the first (second) diffraction order,

$$I^{\mathrm{th}}_{1^{\mathrm{st}}(2^{\mathrm{nd}})}(A, x, y, \{r, \sigma, A_{2\pi}, \alpha\}), \tag{6.38}$$

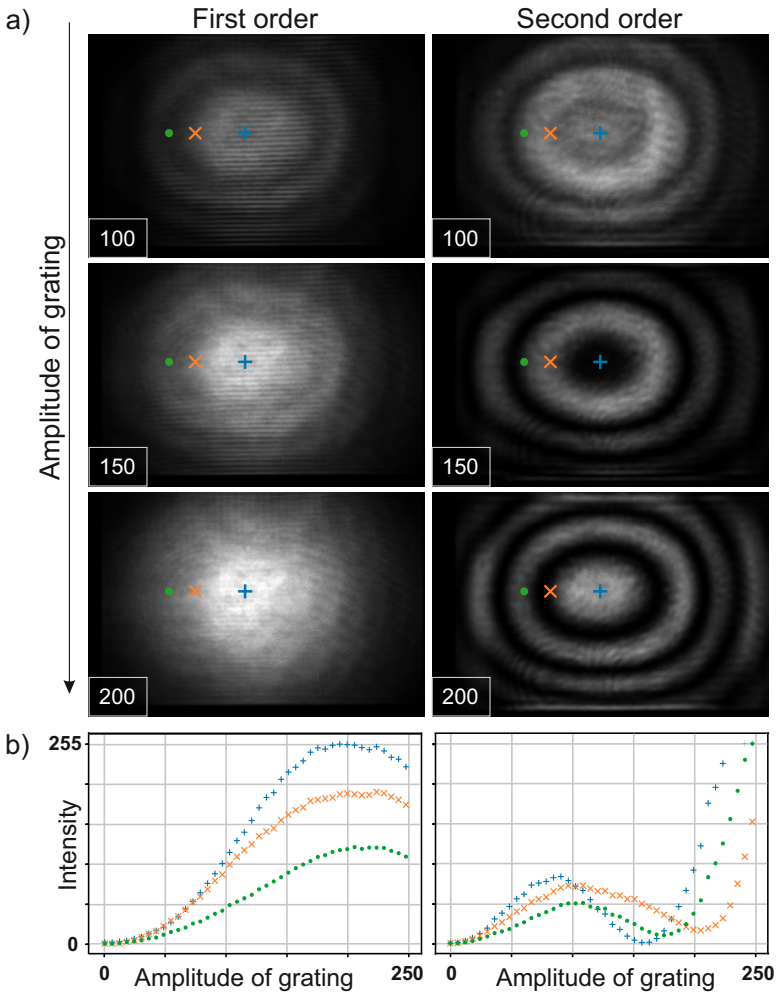


Figure 6.10: (a) Near-field images of the first and second diffraction orders for three different values $A = 100, 150, 200$ of grating amplitude, at three different SLM sections (green, orange, blue). (b) Diffracted intensity versus grating amplitude for the three SLM sections marked in Fig. 6.10(a).

which is a function of the grating amplitude and model parameters. We fit both diffraction orders simultaneously. To evaluate the quality of the fit, we use chi-squared

$$\chi^2(x,y) = \sum_{o=1}^{2}\sum_{i=0}^{n} \frac{\left(I_{o,i}^{\mathrm{exp}} - I_{o,i}^{\mathrm{th}}\right)^2}{I_{o,i}^{\mathrm{th}}}, \tag{6.39}$$

where $o = 1, 2$ are the two diffraction orders and $n = 125$ is the number of voltage values for which the data were acquired.

We estimate the parameters by fitting $I^{\mathrm{exp}}(A, x, y)$, the acquired experimental curves, for each 20×20 section of the SLM screen (examples are shown in

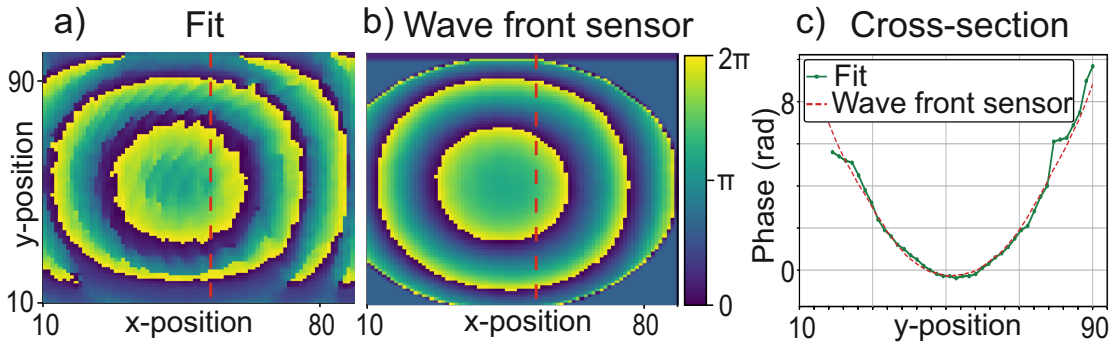


Figure 6.11: Phase map $\alpha(x, y)$ (a) calculated from our model fit and (b) measured with a wavefront camera. The profiles of the cross-sections marked by dashed red lines are shown in (c).

Fig. 6.10(b)). We search for the optimal set in the following intervals: $r \in [0, 0.15]$ in steps of 0.005, $\sigma \in [0, 2]$ pixels in steps of 0.05, $\alpha \in [0, 2\pi]$ in steps of 0.1, and $A_{2\pi} \in [190, 220]$ in steps of 1. The following two-step optimisation process is used. The goal of the first step is to find the optimal pair (r, σ). To this end, we first choose 20 spatial points (x_p, y_p) along the central horizontal cross-section of the raster. We then apply the grid search for each point to find $\alpha(x_p, y_p)$ and $A_{2\pi}(x_p, y_p)$ which minimize $\chi^2(x_p, y_p)$ for each (r, σ). We then find, via grid search, the pair of r and σ which gives the lowest average $\langle \chi^2(x_p, y_p) \rangle_{x_p, y_p}$. In the subsequent second step, we fit $\alpha(x, y)$ and $A_{2\pi}(x, y)$ over the entire SLM screen. The overall optimization procedure is computed in about 10 hours on a notebook with an Intel i5 quad-core processor.

Fitting both orders together significantly improves the accuracy of the parameter estimation. For our SLM, the results were as follows: coverglass reflection coefficient $r = 0.055$, crosstalk Gaussian point spread function width $\sigma = 0.75$ pixels, and the 2π voltage map dropping from 206 in the central part of the SLM to around 197 towards the borders.

Fig. 6.11(a) reports $\alpha(x, y)$ resulting from the fit. As seen, $\alpha(x, y)$ resembles an elliptical paraboloid, with a peak-to-valley value of about 12 rad (1.9λ). To verify the accuracy of the reconstructed $\alpha(x, y)$, we measure it using a commercial wavefront sensor camera (Phasics SID4). To eliminate the contributions of other optical elements, we also acquire a reference wavefront measurement for a setting in which the SLM is replaced by a flat mirror. Figure 6.11(b) shows the difference

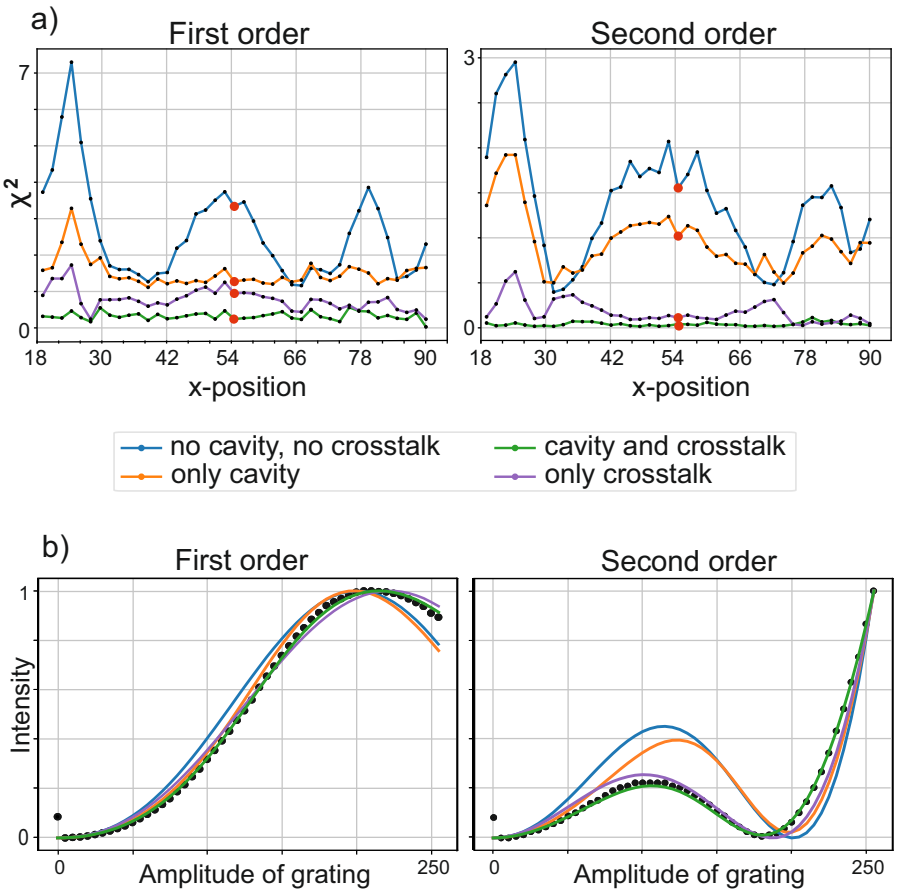


Figure 6.12: Comparison of the fit quality for the different estimation models. a) Fit quality parameter $\chi^2(I^{\text{th}}, I^{\text{exp}})$ calculated along an SLM cross-section. b) Measured (black dot) and fitted (coloured line) intensity response curves for an arbitrary SLM screen position (red dot in (a)).

between the phase profiles produced by the two measurements. This difference is consistent with the phase map reconstructed by our method, as evidenced further by comparing one of its cross-sections (Fig. 6.11(c)).

In Fig. 6.12 we compare the accuracy of our model with three simpler models, in which the crosstalk and/or cavity effects are neglected, i.e. σ and/or r are set to zero. The model neglecting both spurious effects (blue curve) is the least accurate (χ^2 is the highest). The quality of the fit improves if the model includes either cavity (orange curve) or crosstalk (purple curve), and is maximum if both effects are taken into account (green curve).

6.2.5 Compensation

We test the compensation encodings proposed in Section 6.2.2 with high-order Hermite-Gaussian (HG) modes[2]. As a figure of merit, we evaluate the fidelity of the produced mode with respect to the ideal. For this purpose, we acquire with a camera the near and far field intensity images and apply an iterative maximum-likelihood reconstruction method [168] to obtain the first-order normalized coherence profile

$$g^{(1)}(x, y, x', y') = \frac{\langle E(x, y)E^*(x', y')\rangle}{\int \langle E(x, y)E^*(x, y)\rangle \mathrm{d}x\mathrm{d}y}. \tag{6.40}$$

The fidelity of the experimental mode is then calculated as

$$F = \int E^*_{\mathrm{th}}(x, y)E_{\mathrm{th}}(x', y')g^{(1)}(x, y, x', y')\mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y',$$

where $E_{\mathrm{th}}(x, y)$ is the normalised theoretical profile of the ideal mode.

An example for $\mathrm{HG}_{12,12}$ is shown in Fig. 6.13, comparing the theoretical mode profile with those generated by non-, partially- and fully-corrected holograms. The non-compensated hologram (a) produces a mode whose intensity profile has a ring-shaped modulation in the near field (especially visible in the marginal plot below the 3D photograph), due to the Fabry-Perot cavity effect, and distorted in the far field, as a consequence of the curved back panel. The fidelity of this mode with the ideal one is 89.3%. The curvature-corrected hologram (b) generates a mode with a higher fidelity, 92.6%. The far field profile is no longer distorted, but the near field intensity distribution is still modulated by the cavity interference fringes. Hologram (c) compensates both the curvature and cavity effects, producing a mode which best approximates the ideal mode profile (e), with a fidelity of 94.5%.

To correct for the crosstalk, we apply the iterations according to Fig. 6.9 to the previously calculated curvature- and cavity-compensated hologram. While the fidelity of the produced mode is unchanged, the diffraction efficiency increases by 28%, as seen in Fig 6.13(d). The crosstalk compensation primarily modifies the areas of discontinuity in the phase profile $\theta(x, y)$ leading to a sharper blazed

[2]More details about this procedure can be found in our Python implementation [167].

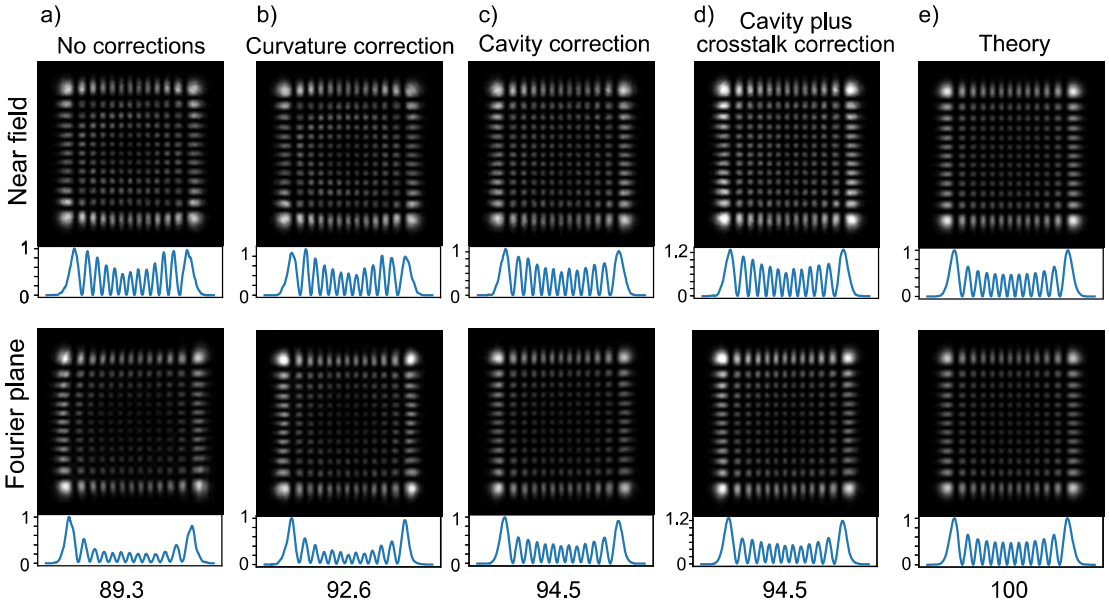


Figure 6.13: Near- (left) and far-field profiles of the $HG_{12,12}$ as produced by the SLM and measured by a camera (a-d) and expected theoretically (e). Each panel indicates the type of correction applied to the SLM hologram and the evaluated fidelity. For each image, the horizontal cross-sections through the mode centre is shown in blue line.

grating, whereas the smooth modulation functions $M(x, y)$ and $F(x, y)$ are not affected, which explains the improvement.

To show the sensitivity of the proposed method to the three spurious effects, we performed additional numerical simulations. To analyse the cavity plus curvature correction, we model an SLM that has a glass-air interface reflection coefficient r that varies from 0 to 0.15, and the backplane curvature that is similar to our SLM (Fig. 6.11), but multiplied by a factor within the interval $[0, 2.0]$.

In Fig. 6.14(a), we calculate the fidelity of $HG_{12,12}$ without any compensation, i.e. we use the algorithm of Ref. [153] to compute the SLM hologram directly. The fidelity decreases as the SLM curvature and reflection coefficient increase. In Fig. 6.14(b), we compensate only for its curvature, neglecting the cavity effect. For $r = 0$ (no cavity), we notice a significant improvement in the calculated fidelity, which however decreases with increasing r as the cavity effect becomes more significant. Finally, in Fig. 6.14(c), we use an SLM hologram that accounts for both the curvature and cavity effect. We see that our algorithm is capable of generating a mode with high fidelity, even when these two effects are significant.

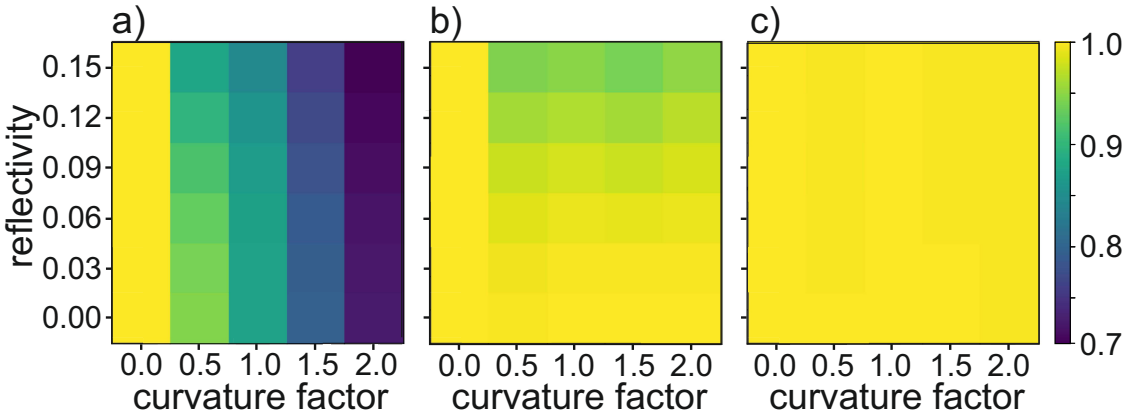


Figure 6.14: Simulated $HG_{12,12}$ fidelity as a function of the reflection coefficient r and curvature multiplying factor, for (a) uncorrected, (b) curvature corrected and (c) cavity corrected holograms.

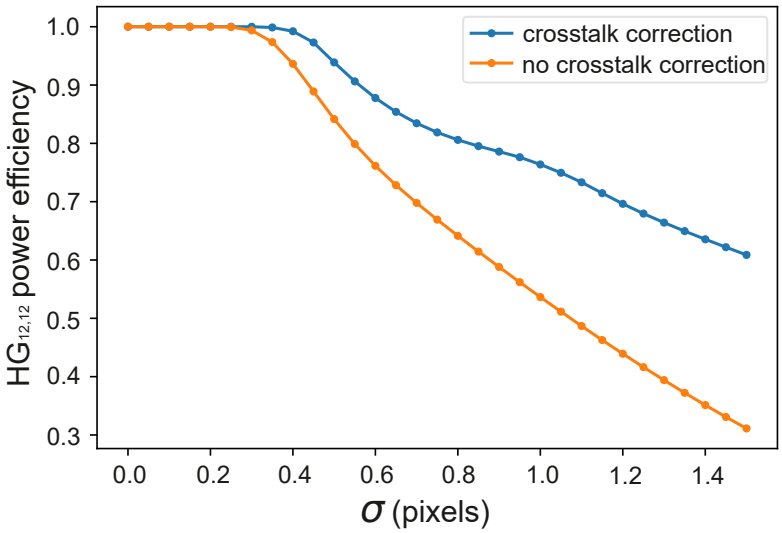


Figure 6.15: Simulated power efficiency as a function of crosstalk PSF width with and without the correction.

To characterize the crosstalk correction, we model the diffraction efficiency of the mode produced by the hologram corrected for the crosstalk effect and compare it with a hologram with no corrections. We plot the efficiency versus the width of the crosstalk point spread function in Fig. 6.15. As seen, the crosstalk correction significantly improves the efficiency for high widths. Both the reflectivity of the air-glass interface and curvature of the back panel are set to zero for this simulation. As mentioned previously, the influence of the crosstalk on the fidelity is insignificant.

6.3 Conclusion

In this chapter we have seen how phase-only LCoS spatial light modulators work, how they can be use for full complex field modulation, how they can be calibrated, and how we can use this calibration in order to improve their performance.

We did the calibration by accounting for all the major SLM spurious effects: the curvature of the back panel, the pixel crosstalk and the low-finesse internal cavity. Our technique of measuring the back panel curvature is reliant on the Fabry-Perot interference between the front cover glass and the back panel. As such, it is robust to possible curvature of the wave front incident onto the SLM or any other alignment imperfections. In this sense, it is complementary to Ref. [156] which permits one to study the cumulative spatially-dependent phase shift caused by all optical elements including the SLM.

The presented model, characterization and compensation method can be readily applied to any phase-only LCoS SLM, with potential benefits for a vast range of applications. In particular, we used this method to improve the performance of an optical prime numbers sieve, whose implementation is reported in Appendix J.

7 | Microscopy experiment

In this chapter, which contains our work in Ref. [111], we present Hermite-Gaussian microscopy (HGM) and report its first experimental implementation, which was also the first implementation of a quantum-inspired super-resolution technique for a full 2D imaging task.

The theory of QISR was laid out in chapters 4 and 5, where we have seen that measuring the projections of the light field along certain mode bases can lead to super-resolution. We have also seen, however, that performing direct mode sorting in a real experiment is still a technologically challenging task. One alternative was proposed: successive heterodyne detections with local oscillators as the basis' modes would also yield the necessary projections. This would shift the bottleneck from mode sorting to mode generation, which can be well accomplished with current technology. Indeed, in Chapter 6 we have seen how to experimentally generate spatial modes with high fidelity.

We now have all the tools to perform QISR in the laboratory. In Section 7.1 we introduce HGM [89], which provides estimators with which we can reconstruct images using the measured photocurrents. We show that a direct analytical implementation of HGM is not feasible in a real experimental setting, and it needs to be aided by a machine-learning algorithm in order to provide meaningful results. In Section 7.2 we present the implementation of HGM for coherent light illumination, with Section 7.3 showing the results: with a basis of 21 spatial modes in both transverse dimensions, we were able to perform two-dimensional imaging with twofold resolution enhancement beyond the diffraction limit.

7.1 Hermite-Gaussian microscopy

Hermite-Gaussian microscopy (HGM), first theoretically proposed in Ref. [89], is a method to reconstruct an object via its mode-sorted components. Although it was devised in the context of quadrature mode sorting with heterodyne detection, it can be applied to any scheme which decomposes the image field into its HG projections. HGM is outlined below.

7.1.1 Coherent light

In this section we show how a coherent object can be recovered from its measured HGM photocurrents, as first demonstrated in Ref. [89].

Consider a plane object illuminated by a uniform coherent light beam. The resulting electric field $E(x, y)$ contains all the information about that object's distribution. If it is imaged by an objective lens, then at its image plane we have the field distribution

$$E'(x, y) = \int \int_{-\infty}^{+\infty} E(x', y')\psi(x - x', y - y')dx'dy', \tag{7.1}$$

where (x, y) and (x', y') are the image and object plane coordinates, respectively, and $\psi(x, y)$ is the point-spread function (PSF) of the lens. Here we use the soft-aperture approximation, where the PSF is assumed Gaussian,

$$\psi(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x^2+y^2)/4\sigma^2}, \tag{7.2}$$

and $\sigma \approx 0.21\lambda/\mathrm{NA}$ [143].

Now consider the 2D Hermite-Gaussian basis

$$\phi_{mn}(x, y) = \frac{1}{\sqrt{2^{m+n}m!n!}} H_m\left(\frac{x}{\sqrt{2}\sigma}\right) H_n\left(\frac{y}{\sqrt{2}\sigma}\right) \psi(x, y), \tag{7.3}$$

with the same waist σ as the PSF. As we have seen in Chapter 4, the HG basis is the PAD (optimal) basis for the Gaussian PSF, in the case of incoherent imaging.

The projection of the image field $E'(x,y)$ along the (m,n)-th order mode, i.e., their overlap, is

$$\begin{aligned} J_{mn} &\propto \int \mathrm{d}\vec{r}\; E'(\vec{r})\phi^*_{mn}(\vec{r}) \\ &= \int \mathrm{d}\vec{r}\; E(\vec{r})R_{mn}(\vec{r}) \\ &= \frac{1}{\sqrt{m!n!}} \iint_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}y\; E(x,y) \left(\frac{x}{2\sigma}\right)^m \left(\frac{y}{2\sigma}\right)^n e^{-(x^2+y^2)/8\sigma^2}. \end{aligned} \tag{7.4}$$

In a typical mode sorting experiment, the image field is decomposed into spatially separated Gaussian beams, each corresponding to a ϕ_{mn} mode. The complex amplitude of each of these beams is proportional to J_{mn}. In a quadrature mode sorting experiment, we perform subsequent heterodyne detections using as signal beam our image field, and as local oscillator (LO) all the ϕ_{mn} HG modes. Each resulting complex photocurrent from the heterodyne detector is then be proportional to J_{mn}. For this reason, we call the latter, from now on, simply *photocurrents*.

The goal of HGM is to use the photocurrents J_{mn} to perform an approximate reconstruction of the original object. Firstly, note that the form of Eq. (7.4) closely resembles the object's geometric moment Θ_{mn}, and reduces to it in the subdiffraction limit, where $e^{-(x^2+y^2)/8\sigma^2} \approx 1$. Therefore, what we basically do is reconstruct an image from its moments.

The idea is to expand the object field in the HG basis,

$$E(x,y) = \frac{1}{2} \sum_{m,n=0}^{\infty} c_{mn}\phi_{mn}\left(\frac{x}{\sqrt{2}}, \frac{y}{\sqrt{2}}\right), \tag{7.5}$$

where

$$c_{mn} = \iint_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}y\; \phi^*_{mn}\left(\frac{x}{\sqrt{2}}, \frac{y}{\sqrt{2}}\right). \tag{7.6}$$

We then have to find the coefficients c_{mn} via the measured photocurrents. In accordance with Ref. [89], we look for, instead, the closely related coefficients β_{mn}, defined as

$$\beta_{mn} = \sqrt{2^{m+n}m!n!2\pi\sigma^2}c_{mn}. \tag{7.7}$$

Expanding the Hermite polynomials in their power series,

$$H_m(x) = \sum_{k=0}^{m} \alpha_{mk}x^k, \tag{7.8}$$

we have that (using Eq. (7.3))

$$\beta_{mn} = \sum_{k=0}^{m}\sum_{l=0}^{n}\sqrt{k!l!}\alpha_{mk}\alpha_{nl}J_{kl}. \tag{7.9}$$

Therefore, if we performed an experiment and acquired the J_{mn} photocurrents for $m, n \in 0, ..., M$, we can calculate the β_{mn} coefficients up to $(m, n) = (M, M)$ and have a reconstructed image

$$E_{\text{rec}}(x, y) = \sum_{m,n=0}^{M} \frac{\beta_{mn} H_m\left(\frac{x}{2\sigma}\right) H_n\left(\frac{y}{2\sigma}\right) e^{-(x^2+y^2)/8\sigma^2}}{2^{m+n+2} m!n!\pi\sigma^2}. \tag{7.10}$$

As we detect more modes, our reconstruction approaches the original object, with equality for $M \to \infty$. Eq. (7.10) is what ultimately defines Hermite-Gaussian microscopy.

What is the improvement of HGM? Instead of resorting to complicated statistical arguments, we go back to the good old Rayleigh criterion. Consider the resolvable distance for the two-point resolution problem, i.e., the minimum separation for which we can distinguish two centred point sources. In Fig. 7.1 we plot the resolvable distance, normalised by the direct imaging Rayleigh distance, as a function of M. Note that, as we increase M, there are diminishing returns with respect to the resolution improvement. Since higher-order modes are increasingly harder to detect or generate, we consider a sweet spot to be $M = 20$. For that value, the resolvable distance is $\approx 0.389 d_R$, which corresponds to an improvement factor of 2.57 in resolution over direct imaging.

7.1.2 Shot-noise analysis

The HGM equations are derived in a noiseless, ideal scenario. In a real experiment, even if all sources of systematic errors are eliminated, the unavoidable shot noise will still be present. As can be seen from Eq. (7.4), the photocurrents get exponentially smaller with the HG mode order, and therefore become heavily affected by shot noise. These small photocurrents, however, are amplified via the factorial terms in Eq. (7.9). The result is that even the slightest noise leads to a complete breakdown of the reconstruction in Eq. (7.10).

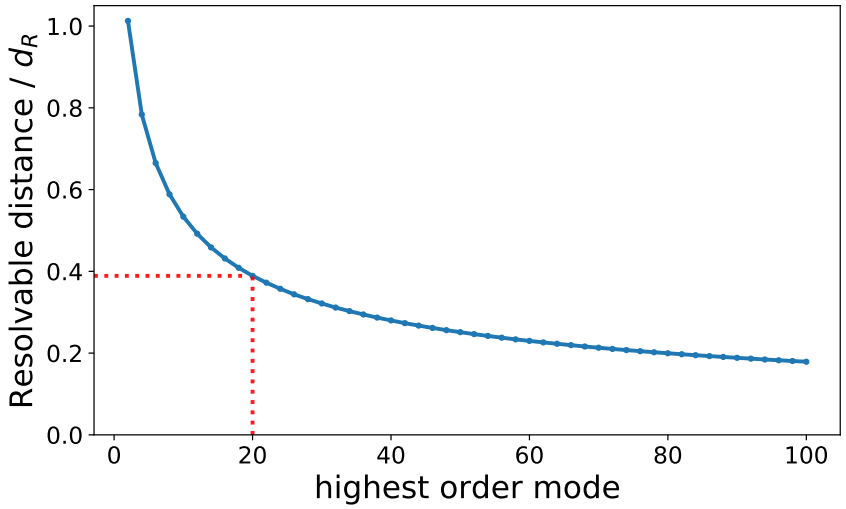


Figure 7.1: Resolvable distance, normalised by the Rayleigh distance d_R, for coherent light HGM as a function of the highest order M of HG modes used, with the dashed red lines highlighting $M = 20$.

In order to see that, we considered the sample object of Fig. (7.2), which consists of $3 \times 10 = 30$ sub-objects which are independently imaged by HGM. We simulated an idealized scenario where there are no experimental imperfections and the only source of error is shot noise (which is the inverse square root of the number of photons available). For heterodyne detection, as per Eq. (5.21), we simulate a Gaussian shot noise, while for mode sorting, Eq. (5.16), the noise is Poissonian. The results are shown in In Fig. 7.3. They confirm our suspicion that higher-order modes are more susceptible to shot noise, and HGM yields nonsensical reconstructions if the noise is not low enough. We see that 10^{17}–10^{18} photons per image section are necessary for HGM to work when HG modes up to (20,20) modes are used, which is an enormous photon overhead. Even when limiting the mode order to (5,5), still 10^4–10^5 photons are necessary to produce any meaningful result.

Therefore, practical HGM requires an algorithm that regularizes the noise from the higher-order photocurrents while at the same time being able to extract information from them. In this work, we addressed this issue by using a neural network, as will be shown in the experimental section.



Figure 7.2: Oxford logo, split into $3 \times 10 = 30$ sub-images, which are each independently imaged by HGM.

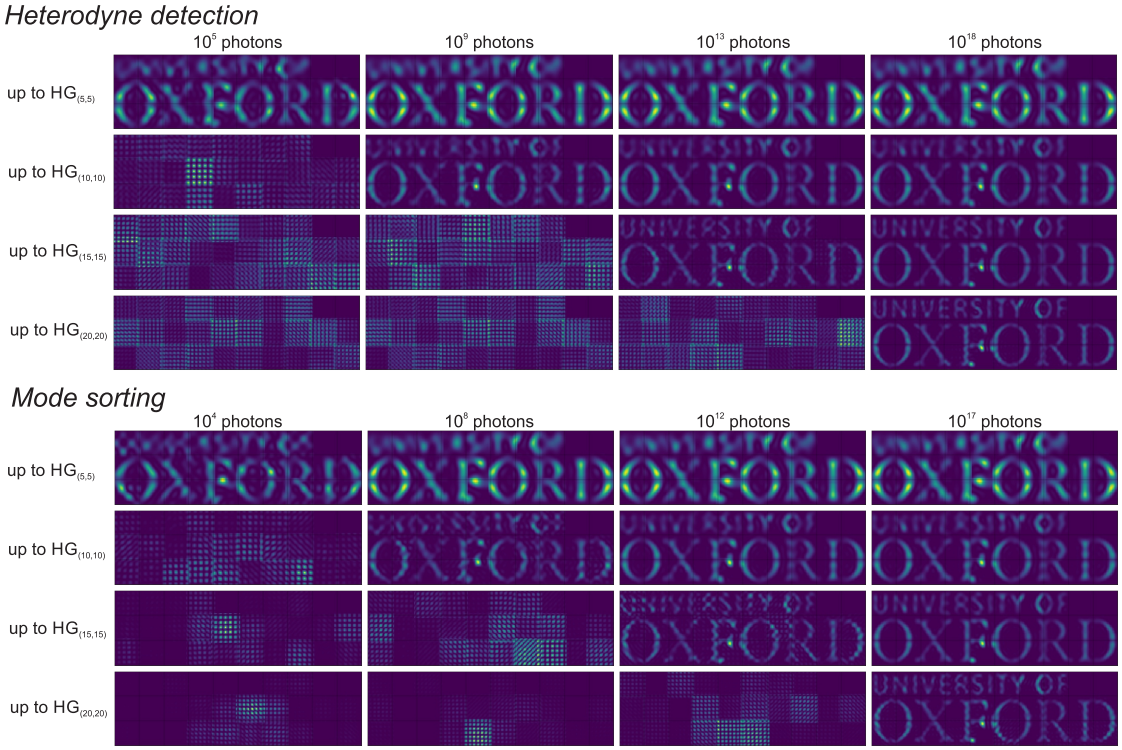


Figure 7.3: Oxford logo reconstruction using the closed-form expression (7.10) in the presence of shot noise. Simulations for two detection methods, different sizes of HG mode basis and different number of photons available per image section (DMD frame) are displayed. When fewer photons are available, the relative shot noise is more significant and reasonable reconstruction is possible only with smaller HG bases.

7.1.3 Incoherent light

If the object is illuminated by or emits incoherent light, the photocurrents J_{mn} at the heterodyne detector will have a random phase. In practice, we will only be able to measure the detector power

$$P_{mn} = \langle |J_{mn}|^2 \rangle = \iint \mathrm{d}\vec{r}\mathrm{d}\vec{r'} \ \left\langle E'(\vec{r})E'^*(\vec{r'}) \right\rangle \phi^*_{mn}(\vec{r})\phi_{mn}(\vec{r'}), \tag{7.11}$$

where

$$\left\langle E'(\vec{r})E'^*(\vec{r'}) \right\rangle = \iint \mathrm{d}\vec{\rho}\mathrm{d}\vec{\rho'} \ \left\langle E(\vec{\rho})E^*(\vec{\rho'}) \right\rangle \psi(\vec{\rho}-\vec{r})\psi^*(\vec{\rho'}-\vec{r'}). \tag{7.12}$$

The statistics of incoherent light imply that

$$\left\langle E(\vec{\rho})E^*(\vec{\rho'}) \right\rangle = \delta^2(\vec{\rho}-\vec{\rho'})I(\vec{\rho}), \tag{7.13}$$

where $I(\vec{\rho})$ is the object plane intensity distribution. This means that the heterodyne power is

$$\begin{aligned} P_{mn} &= \int \mathrm{d}\vec{r} \ I(\vec{r})|R_{mn}(\vec{r})|^2 \\ &= \frac{1}{m!n!} \iint_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}y \ I(x,y) \left(\frac{x}{2\sigma}\right)^{2m} \left(\frac{y}{2\sigma}\right)^{2n} e^{-(x^2+y^2)/4\sigma^2}. \end{aligned} \tag{7.14}$$

Analogously to the coherent case, let's expand the object intensity in the HG basis as

$$E(x,y) = \sum_{m,n=0}^{\infty} c_{mn}\phi_{mn}(x,y), \tag{7.15}$$

where

$$c_{mn} = \iint_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}y \ \phi^*_{mn}(x,y). \tag{7.16}$$

We similarly define the β_{mn} coefficients as $\beta_{mn} = \sqrt{2^{m+n}m!n!2\pi\sigma^2}c_{mn}$, with respect to which the reconstructed intensity is

$$I_{\mathrm{rec}}(x,y) = \sum_{m,n=0}^{M} \frac{\beta_{mn}H_m\left(\frac{x}{\sqrt{2}\sigma}\right)H_n\left(\frac{y}{\sqrt{2}\sigma}\right)e^{-(x^2+y^2)/4\sigma^2}}{2^{m+n+1}m!n!\pi\sigma^2}. \tag{7.17}$$

All seems well, but if we try to obtain the β_{mn} from the heterodyne powers we get

$$\beta_{mn} = \sum_{k,l=0}^{m,n} \sqrt{2^{k+l}}\alpha_{mk}\alpha_{nl}P_{\frac{k}{2},\frac{l}{2}}. \tag{7.18}$$

Since there is no P_{mn} for m, n non-integer, we simply cannot reconstruct the coefficients, and therefore the image, with HG projections alone.

What can we reconstruct with the data we have? If we try to get the coefficients with the P_{mn} only,

$$\beta'_{mn} = \sum_{k,l=0}^{m,n} 2^{k+l} \alpha_{m,2k} \alpha_{n,2l} P_{k,l}, \tag{7.19}$$

we end up reconstructing the alternative distribution

$$I'(x,y) = \frac{1}{4}\left[I(x,y) + I(-x,y) + I(x,-y) + I(-x,-y)\right], \tag{7.20}$$

which is the original intensity superposed with the three *mirror* or *ghost images* already encountered in Chapter 4, which are depicted in Fig. 4.15. One possible solution is to use quadrant-displaced modes: if the image is entirely contained inside one quadrant, the ghost images do not superpose with it. Another solution is to use interferometric HGM, i.e., measure also the superpositions of HG modes

$$\phi^{\pm}_{m+m',n+n'}(x,y) \equiv \frac{1}{\sqrt{2}}\left[\phi_{m+m',n+n'}(x,y) \pm \phi_{m,n}(x,y)\right], \tag{7.21}$$

We therefore consider the basis $\{\phi_{m,n}, \phi^{+}_{m+1,n}, \phi^{-}_{m+1,n}, \phi^{+}_{m,n+1}, \phi^{-}_{m,n+1}, \phi^{+}_{m+1,n+1}, \phi^{-}_{m+1,n+1}\}$, which is a slight variation[1] of the iPAD basis, Eqs. (4.78) to (4.84). We denote the obtained heterodyne powers as $\{P^{(i)}_{mn}\}$, for $i = 0, .., 6$. After some algebra, we can obtain the β_{mn} coefficients via

$$\beta_{mn} = \sum_{k,l=0}^{m,n} \sqrt{2^{k+l}} \alpha_{mk} \alpha_{nl} P'_{kl}, \tag{7.22}$$

where

$$P'_{mn} = \begin{cases} (\frac{m}{2})!(\frac{n}{2})! P^{(0)}_{\frac{m}{2},\frac{n}{2}}, \ m,n \text{ even}, \\ (\frac{m-1}{2})!(\frac{n}{2})!\sqrt{\frac{m+1}{8}}\left(P^{(1)}_{\frac{m-1}{2},\frac{n}{2}} - P^{(2)}_{\frac{m-1}{2},\frac{n}{2}}\right), \ m \text{ odd}, n \text{ even}, \\ (\frac{m}{2})!(\frac{n-1}{2})!\sqrt{\frac{n+1}{8}}\left(P^{(3)}_{\frac{m}{2},\frac{n-1}{2}} - P^{(4)}_{\frac{m}{2},\frac{n-1}{2}}\right), \ m \text{ even}, n \text{ odd}, \\ (\frac{m-1}{2})!(\frac{n-1}{2})!\sqrt{\frac{m+1}{8}}\sqrt{\frac{n+1}{8}}\left(P^{(5)}_{\frac{m-1}{2},\frac{n-1}{2}} - P^{(6)}_{\frac{m-1}{2},\frac{n-1}{2}}\right), \ m,n \text{ odd}. \end{cases} \tag{7.23}$$

The reconstructed intensity is then given by Eq. (7.17).

[1] The difference is in the $n-1$ term in bases 5 and 6.

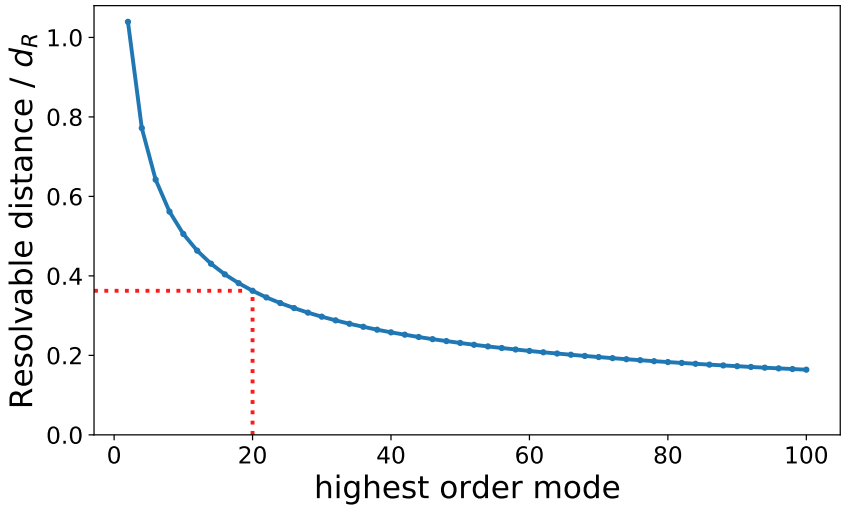


Figure 7.4: Resolvable distance, normalised by the Rayleigh distance d_R, for incoherent light HGM as a function of the highest order M of HG modes used, with the dashed red lines highlighting $M = 20$.

The resolvable distance for incoherent light HGM is shown in Fig. 7.4. Due to the parity symmetry of the two centred sources problem, the interferometry scheme is not necessary. At $M = 20$ modes, the resolvable distance is approximately 0.363 times the Rayleigh distance, which corresponds to a 2.75 resolution improvement factor.

7.2 Experiment

In our experiment, we did a proof-of-principle HGM implementation with $441 = 21 \times 21$ HG modes via heterodyne detection with coherently illuminated samples. We used a very low numerical aperture imaging system, since our objects were micrometer sized, and the resolution enhancement was considered with respect to that aperture. In order to make HGM work amidst experimental imperfections and shot noise, we used a neural network trained on thousands of sample objects. With this, we were able to achieve a twofold resolution improvement over the diffraction limit. Below, we report this experimental implementation in a detailed and comprehensive manner.

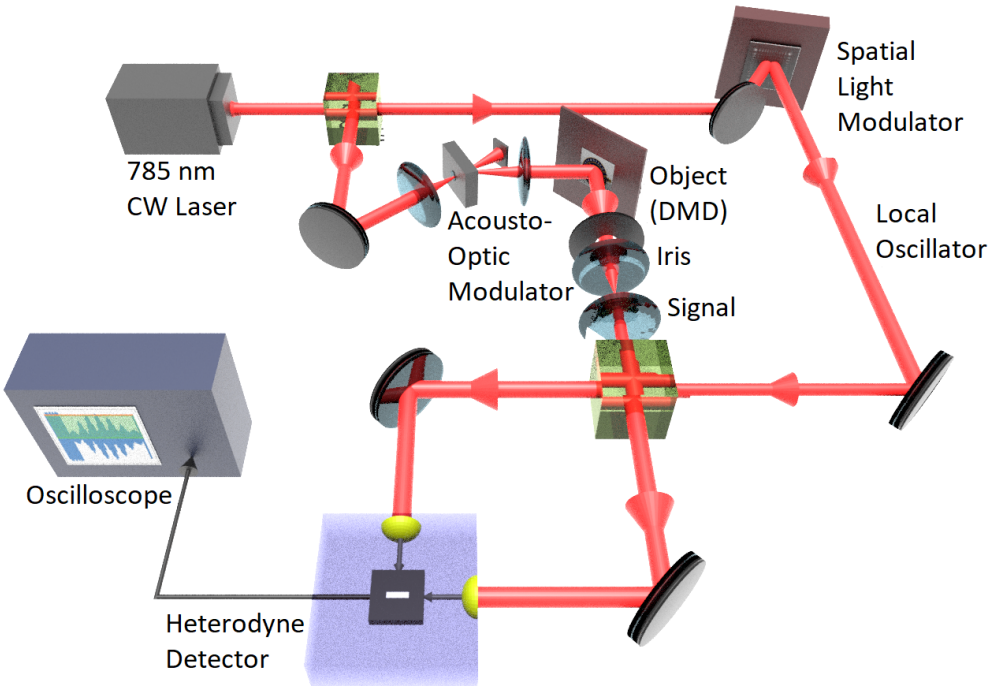


Figure 7.5: Schematic of the experiment

7.2.1 Setup

The experimental setup is displayed in Fig. 7.5. The beam of a continuous wave, distributed feedback diode laser (Eagleyard EYP-DFB-0785), operating at 785 nm, is sent through a single-mode fibre (Thorlabs P3-780PM-FC-1) in order to obtain a Gaussian spatial profile. The laser diode is able to produce about 84 mW of optical power and has a relatively narrow bandwidth of 2 MHz. A half-wave plate (HWP) and a polarizing beam splitter (PBS) split the laser beam into two paths, which will correspond to the local oscillator (LO) and signal of the heterodyne detector (HD).

Optical path

In the LO path, the laser beam is magnified by a telescope (f_1 = 50 mm, f_2 = 200 mm) in order to fully and (almost) uniformly illuminate the screen of a reflective phase-only LCoS SLM, as shown in Fig. 7.6. The incident beam hits the SLM screen almost perpendicularly, so that the angle between the incoming and reflected waves is smaller than 5 degrees, and polarized parallel to the SLM's LC director axis, so that all the incident light is modulated. The SLM holograms

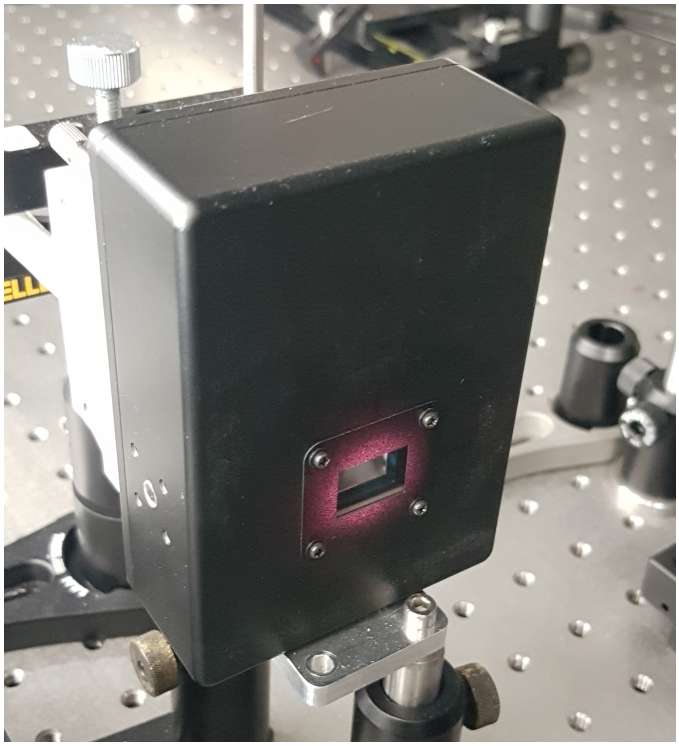


Figure 7.6: LCoS SLM.

output the desired phase and amplitude profiles in the first order of diffraction, which is selected by a telescope ($f_1 = 250$ mm, $f_2 = 100$ mm) and an iris at its focal plane, in the same manner as in Fig. 6.4.

In the signal path, the beam is sent through an acousto-optic modulator (AOM) in order to shift its frequency for the heterodyne detection. The purpose of this step is to reduce flicker noise. The produced first diffraction order mode from the AOM is incident upon a digital micromirror device (DMD), which modulates its amplitude with the binary bitmap of the object to be imaged. The imaging system aperture is defined by an iris whose diameter is set to 3.5 ± 0.03 mm and which is placed at a distance of 245.5 ± 1 cm from the DMD, corresponding to an optical system of numerical aperture (NA) of $(7.1 \pm 0.07) \times 10^{-4}$. The iris diameter was measured by placing a metal ruler next to it and imaging both these objects with a high-resolution camera, as shown in Fig. 7.7.

After the iris, a set of three telescopes magnify and collimate the signal beam in order to match the waist and wave-vector (propagation direction) of the LO

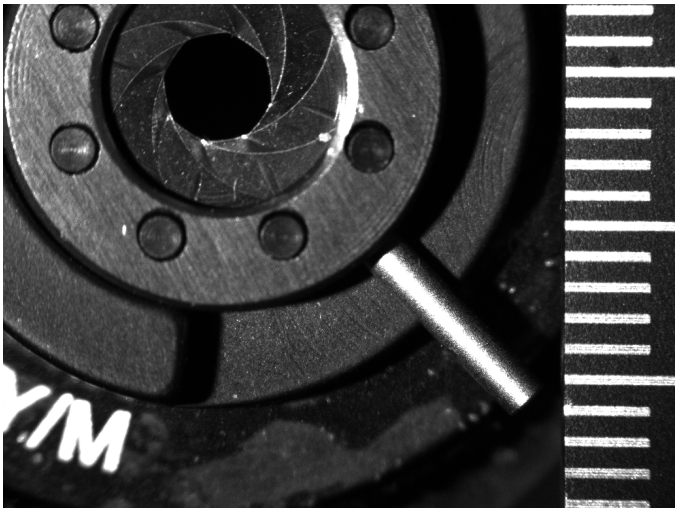


Figure 7.7: Iris and metal ruler.

beam in the 0-th order HG mode. The first telescope ($f_1 = 100$ mm, $f_2 = 50$ mm) collects the light transmitted through the iris. In the second ($f_1 = 50$ mm, $f_2 = 100$ mm) and third telescopes ($f_1 = 75$ mm, $f_2 = 75$ mm), the "eyepiece" lenses are mounted on translation stages to independently control the signal beam's diameter and divergence, hence enabling mode matching to the LO. The signal and LO paths are reunited at a PBS, whose output beams feed the photodetectors of a homemade balanced detector [169, 170], as shown in Fig. 7.8.

Spatial light modulator

We used a phase-only LCoS SLM (Hamamatsu X13138-02, 1272×1024 pixel resolution and 12.5 μm pixel pitch) to generate the Hermite-Gaussian modes for the LO path. In order to do that, we illuminated the SLM by a wide Gaussian beam, such that it was approximately uniform along the extent of its screen. Then, we used the scheme of Section 6.1.2, Eq. (6.22), to create holograms that modulate the amplitude and phase of the HG modes in the first order of diffraction of the SLM's reflected wave [153].

Fig. 7.9 illustrates a few examples of the holograms used. When computing the holograms, we compensate for the incoming beam's non-uniform profile and the SLM backplane curvature using our calibration procedure [110] outlined in Section 6.2. We didn't perform the cavity and crosstalk corrections to speed up the hologram

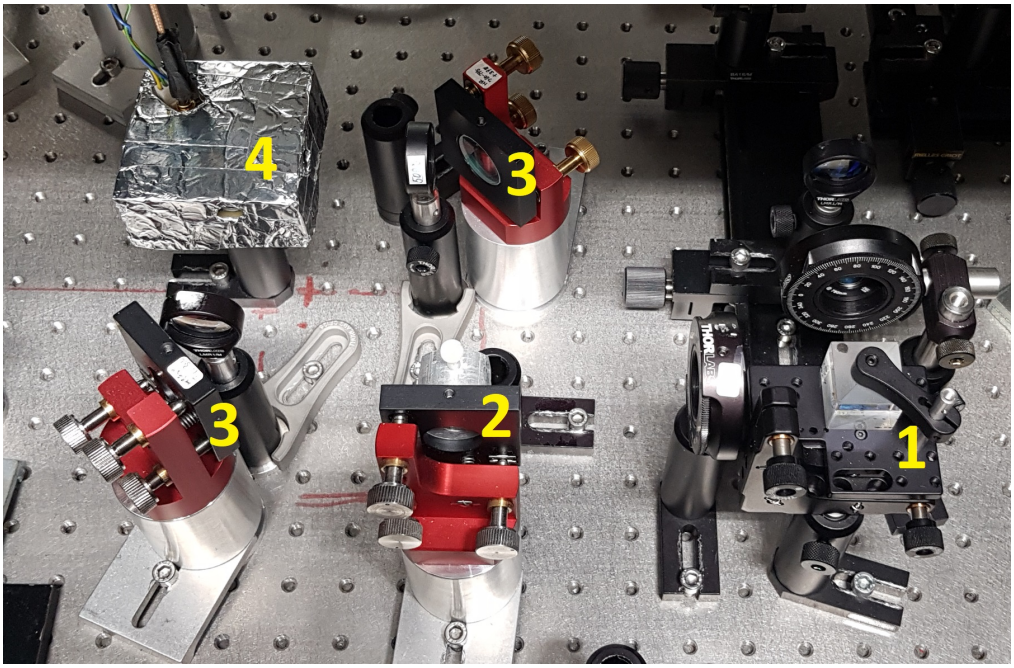


Figure 7.8: Heterodyne setup. Signal and LO beams merge at a PBS (1), and one of the outputs is further split at another PSB (2).Then, the beams are directed via mirrors (3) to the input ports of the heterodyne detector (4).

generation, but the decrease of fidelity due to this omission was compensated by the neural network in the data post-processing stage. The holograms could be modified further to alter the waist, displacement and orientation of the generated modes, which was used for the alignment procedures.

Digital micromirror device

The DMD (DLP LightCrafter 6500), shown in Fig. 7.10, consists of 1920×1080 micromirrors of pitch 7.56 μm. Each of them can be set to the ON or OFF state, corresponding, respectively, to a tilt by $+12°$ or $-12°$ along their diagonal axis and relative to their flat position. We used the DMD to generate the sample objects of the microscope. For that, the signal beam illuminated the central area of the DMD, where 210×210 micromirrors would display the objects to be imaged. The DMD pixels outside that area were set to the OFF state. The reason for using only this reduced area, of 1.588 mm, was to reduce optical aberrations due to the intrinsic DMD screen curvature. For a NA of $(7.1 \pm 0.07) \times 10^{-4}$, the theoretical coherent light Rayleigh limit $1.64\lambda/2\text{NA}$ is 906 μm, which corresponds to 120 DMD pixels,

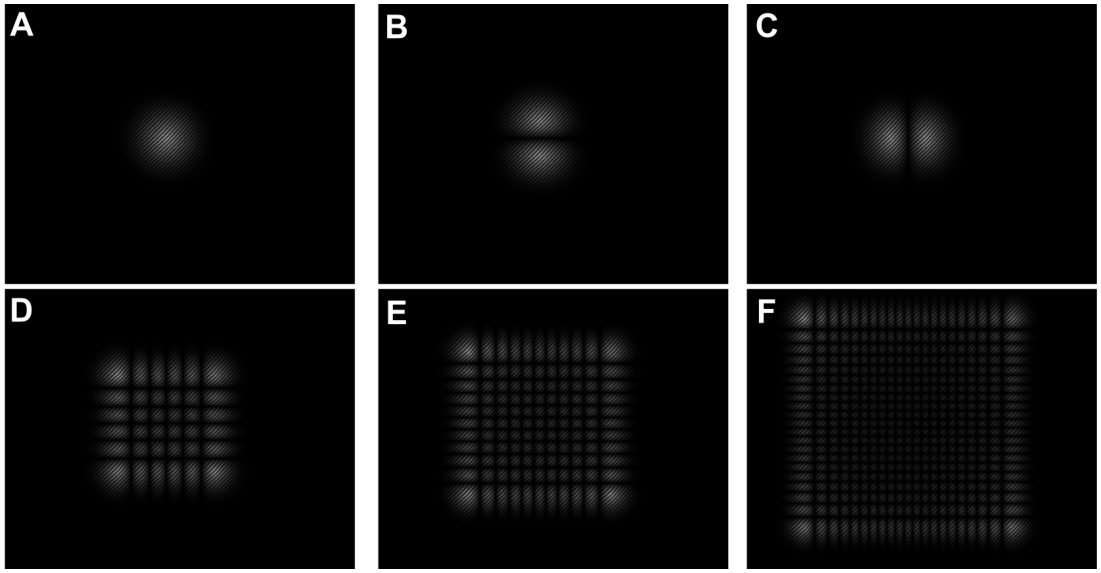


Figure 7.9: Examples of SLM holograms used to produce HG modes. (A) $HG_{0,0}$, (B) $HG_{0,1}$, (C) $HG_{1,0}$, (D) $HG_{5,5}$, (E) $HG_{10,10}$, (F) $HG_{20,20}$.

while the incoherent limit $1.22\lambda/2\text{NA}$ equals 674 μm or 89 DMD pixels. In either case, the Rayleigh distance was more or less half of the working area.

The DMD was driven in 'pattern on-the-fly' mode, in which a sequence of 400 binary bitmap images is loaded into the internal memory of the DMD controller board via USB. The DMD can update images every 105 μs ($\approx$ 10 kHz) at its fastest rate.

Acousto-optic modulator

The AOM (Isomet 1205C-1), driven at 92.05 MHz, was used to shift the frequency of the signal beam. The AOM works by producing a sound wave which modulates the refractive index of a crystal. A beam that passes through it experiences Bragg diffraction, which each m-th diffraction order undergoing a Doppler frequency shift equal to m times the AOM driving frequency. AOMs are usually calibrated to diffract most power in the first order, which is the one we chose as signal beam, while other orders were blocked. The 92.05 MHz RF driving signal was generated using a Direct Digital Synthesizer (DDS, AD9959 EVAL MODULE), and was amplified via a voltage-controlled amplifier in order to maximize the optical power.

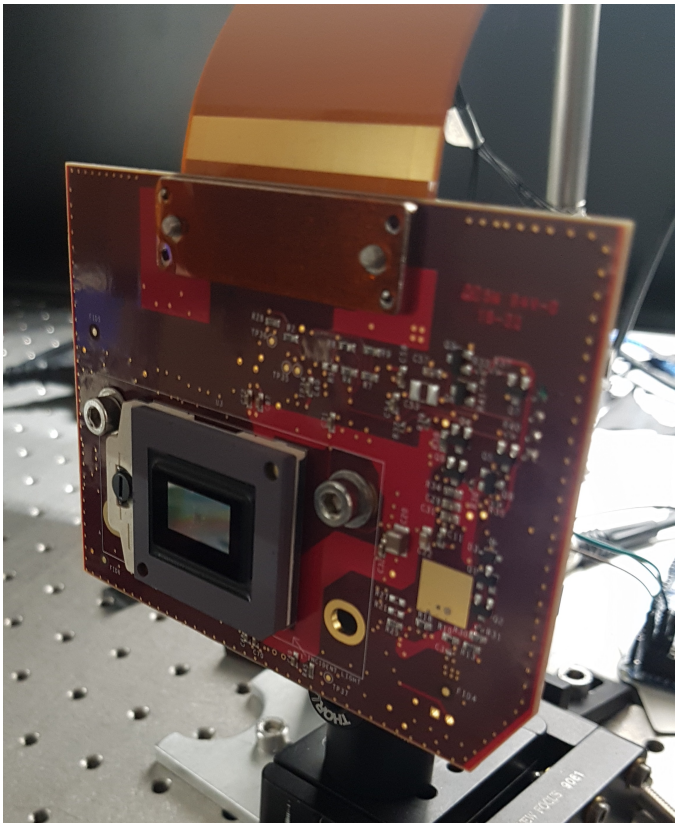


Figure 7.10: Digital micromirror device.

Electronics

The electronics scheme is shown in Fig. 7.11. The DDS board generates two sinusoidal signals at 92 MHz (DDS-CH0) and 92.05 MHz (DDS-CH1). The signal from DDS-CH0 goes from a 50:50 power splitter (Mini-circuits ZMSC-2-1+) to the frequency mixers 1 and 2 (Mini-circuits ZX-5-1L-S+), while the signal from DDS-CH1 goes from a 90:10 power splitter (Mini-circuits ZFDC-10-1-75) to the AOM (90%) and to mixer 2 (10%). Before reaching the AOM, the signal is amplified (amp 2, Mini-circuits ZX60-8008E and ZHL-32A in series).

The homemade [169] heterodyne detector (HD) outputs a 90.05 MHz signal which goes to a 90:10 splitter, with 10% for the spectrum analyzer (Rigol DSA800) and 90% for mixer 1.

The output of the frequency mixers is a superposition of the sum and difference frequencies. The output of mixer 1 is amplified (amp 1, Mini-circuits ZFL-500-BNC and ZFL-1000LN+ in series) before going through the low pass filter LP1

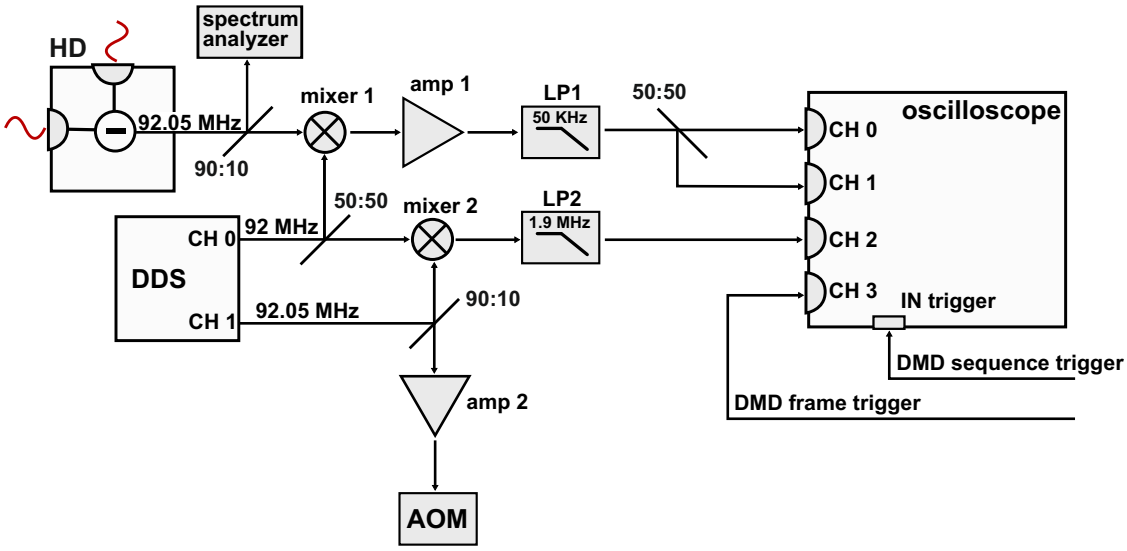


Figure 7.11: Electronics setup.

(Mini-circuits BLP 1.9+ and Thorlabs EF 124 in series), which transmits only the difference frequency at 50 kHz (92.05 MHz - 92 MHz). The downconverted signal is 50:50 split between two oscilloscope (Tektronix MSO 5104) channels CH0 and CH1 and consists of the measured HD data.

The reason we reduce the frequency of the HG signal to 50 kHz is the fact that the DMD refresh rate is at most 10 kHz. A 92.05 MHz signal would contain thousands of times the same DMD frame photocurrents while only a few are needed for the measurement, which would be a huge computational burden. Therefore, the downconversion was used to decrease the size of data and speed up data transfer and saving.

The output of mixer 2 is also downconverted to 50 kHz in the low pass filter LP2 (Mini-circuits BLP 1.9+) and enters channel CH2 of the oscilloscope. This is the *reference signal*, which is necessary for measuring the phase of the photocurrents. Finally, channel CH3 and Trigger IN of the oscilloscope are connected to the frame and sequence (recall that a sequence consists of 400 DMD frames) trigger outputs of the DMD, respectively.

7.2.2 Procedures

Mode matching

It is crucial for HGM to have a good mode-matching between the $HG_{0,0}$ mode and the system's PSF. This could be done by matching the mode in the LO with the image of a single DMD pixel at the centre of the screen, which would represent a point source, but this image would have a very low signal-to-noise ration. Instead, we match the mode with the image of a 20×20 pixel square for waist and divergence matching and a 100×100 square for k-vector matching.

The alignment procedure consists of two steps. The first one, a coarser alignment, consists of using telescopes and mirrors in the LO path to match the waist and divergence of the $HG_{0,0}$ mode generated at the SLM with the signal. In that step, we try to maximize the power of the HD photocurrents, which would correspond to maximum overlap.

In the second, finer alignment step, we match the propagation direction. Now, we generate modes $HG_{0,1}$ and $HG_{1,0}$ at the SLM and try to minimise their overlap with the 100×100 *alignment square* [Fig. 7.12(A)] since these modes are supposed to be orthogonal to the PSF when perfectly aligned. In this case, we have now a *null measurement*, which is significantly more sensitive. Our goal is to achieve a suppression $10\log\left(\mathcal{P}_{0,1/1,0}/\mathcal{P}_{0,0}\right)$ of at least -40 dB, where $\mathcal{P}_{m,n}$ is the photocurrent power with the $HG_{m,n}$ mode.

Because the data acquisition takes several hours, it is necessary to regularly realign the LO and signal beams with respect to each other in both transverse position and direction. To this end, every 3 hours we repeat the fine alignment step automatically, where the LO modes are now translated and reoriented via translations and diffraction gratings applied directly to the SLM holograms. The algorithm scans these parameters and stops when it reaches a suppression of -40 dB. In order to speed up the program, the cavity and crosstalk corrections are not applied to the holograms. The whole procedure usually lasts around 3 minutes, and the corresponding code can be found in Ref. [171].

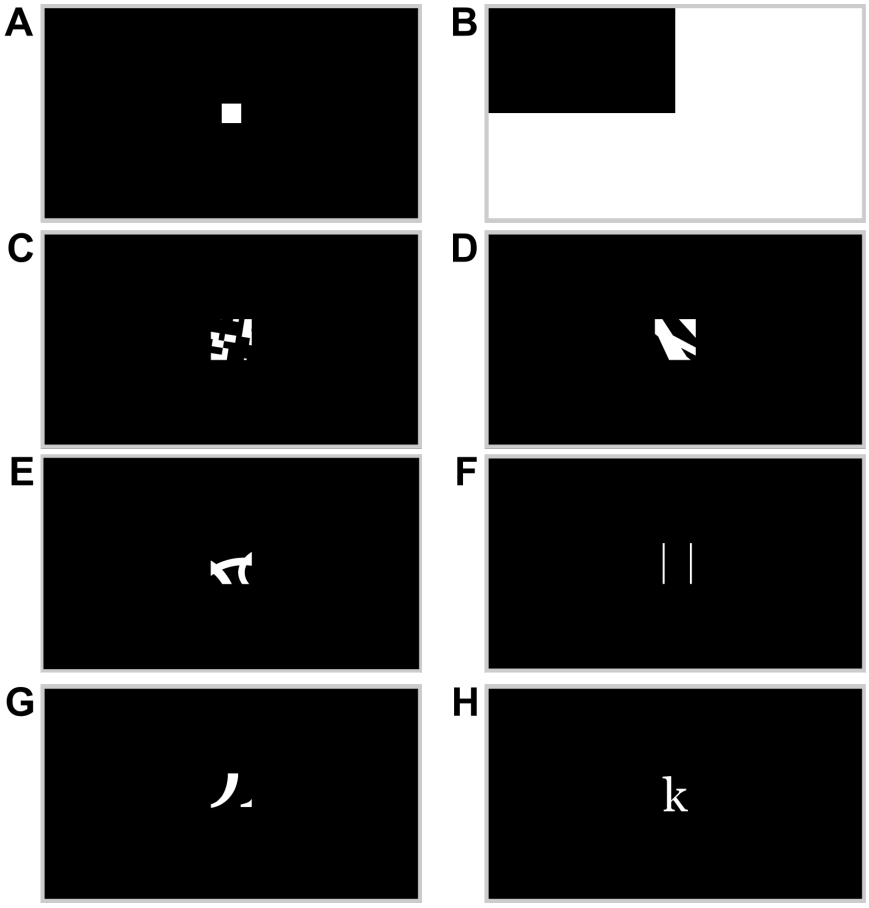


Figure 7.12: Examples of DMD frames used in the experiment: (A) alignment square; (B) phase reference frame; (C) random matrix (train dataset); (D) superposed random lines (train dataset); (E) random parts of an ellipse (train dataset); (F) pair of lines (test dataset); (G) frame from Oxford University logo (test dataset); (H) text symbol (test dataset).

Data acquisition

We set the acquisition order to minimize the overall measurement time, taking into account the SLM and DMD refresh times (~ 0.5 s and ~ 0.1 ms, respectively), and the time required to load new frames into the DMD internal memory (~ 20 s for a batch of 400 binary frames). After each batch is loaded into the DMD, it is displayed sequentially while keeping the LO mode constant. The corresponding 400 photocurrents are acquired with a digital oscilloscope in a single trace, along with the AOM driving signal to keep track of the phase. While the trace is transferred from the oscilloscope to the PC in the binary format via USB (it takes approximately 1 s to transfer 55000 data points and 4 channels), the SLM hologram is updated

to produce the next HG mode. The acquisition for one such batch and all 441 LO modes lasts about 9 minutes.

The acquired traces are processed to extract the amplitude and phase of the heterodyne detection photocurrent for each object. The relative phase between the signal and LO arms of the optical setup drifts with time. To keep track of this drift, we use the phase reference object [Fig. 7.12(B)] that is the union of the bottom and top-right quadrants of the DMD screen. Its asymmetric shape has nonzero overlap with all LO modes, so this object always generates a measurable photocurrent. The phase $\phi_{j,m,n}$ associated with the LO mode $\mathrm{HG}_{m,n}$ and object j can then be calculated according to

$$\phi_{j,m,n} = \left(\phi^{\mathrm{HD}}_{j,m,n} - \phi^{\mathrm{AOM}}_{j,m,n}\right) - \left(\phi^{\mathrm{HD}}_{\mathrm{ref},m,n} - \phi^{\mathrm{AOM}}_{\mathrm{ref},m,n}\right), \tag{7.24}$$

where each term is parentheses is the difference between the phases of the photocurrent and the AOM driving signal (Mixer 2 output) acquired by the oscilloscope, both oscillating at 92.05 MHz.

The signal-to-noise ratio of the acquired signal was limited by technical noise and amounted, dependent on the object, to 25–35 dB for lowest-order modes and 0–10 dB for highest-order modes. This ratio was typically poorer for odd modes because most objects had a dominant symmetric component, which gave rise to higher signal in even modes.

Each batch loaded in the DMD includes 398 objects to be imaged, a phase reference object and an alignment square. The latter is displayed in every batch of objects during the regular data acquisition for monitoring purposes. The Python code used for the acquisition and pre-processing can be seen in our GitHub repository [171].

7.2.3 Neural network

The HGM reconstruction procedure relies on precise knowledge of the PSF and is supremely sensitive to even the slightest experimental imperfections. The sources of errors can be manifold: imperfect HG modes, phase aberrations in both beams'

paths, the intrinsic curvature of the DMD surface, hardness and asymmetry of the aperture, and the misalignment between the HG modes and the signal beam, among others. Furthermore, as we have seen in Section 7.1.2, any realistic implementation of HGM requires some sort of regularization mechanism that reduces the error from the noise at higher-order modes. To overcome all these issues, we decided to perform the reconstructions using a supervised feedforward neural network (NN) [172]. The NN would, in a "black-box" approach, learn the input-output mapping between photocurrents (features) and objects (labels).

After playing for a while with NN architectures, we realized that forcing the NN to learn the direct mapping from photocurrents to objects in their original form was not a good idea. Firstly, the sharp discontinuities of the objects, which are binary matrices of zeros and ones, are hard to learn by neural networks. Secondly, we would be forcing the algorithm to predict something that was physically impossible, namely to function as a microscope with "infinite" resolution. In other words, we would be asking the NN to work in a resolution that the very physical system that generated the data was not able to achieve. Anytime we tried that direct mapping, the NN would seriously overfit.

An alternative would be to smooth the original binary objects, say by convolution with a Gaussian kernel. In that case, we would solve the discontinuities issue and, for a strong enough smoothing, we would be "asking" the NN for a mapping that it could actually provide. Indeed, this approach worked. However, the smoothing factor would be another hyperparameter to tune, and would not necessarily lead to the highest resolutions we could achieve.

The solution we found was to use as labels not the original objects, but the HGM-reconstructed objects, i.e., the images we would have obtained by directly applying the Eq. (7.10) in an ideal, noiseless scenario. In this case, we are asking the NN to do exactly what it can achieve, and its role would only be of regularization and systematic error correction.

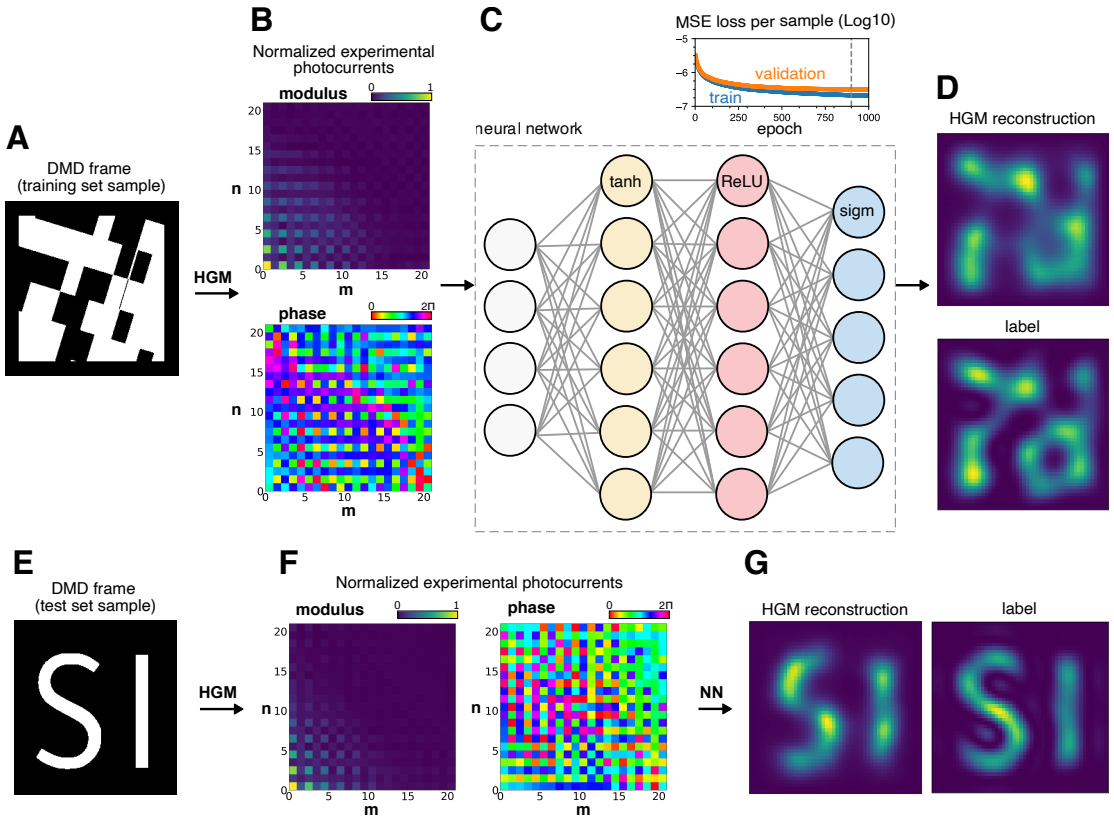


Figure 7.13: Sketch of the data acquisition and processing pipeline. We image via HGM a sample (A) and acquire its complex photocurrents (B), which are then fed to the NN (C) (the inset shows the learning curve, i.e. the training and cross-validation loss versus the training epoch). The NN is trained to predict the idealized HGM reconstruction of the sample image (D). (E-G) show the reconstruction of a sample from the test set.

Architecture

The NN, whose architecture is shown in Fig. 7.13, is a multilayer perceptron with two hidden layers of 6000 units each. The features and labels are generated as follows. A 210×210 binary matrix M is printed in the centre of the DMD. The complex matrices of 21×21 = 441 experimental photocurrents J_{mn} it generates is the features vector. These matrices are normalised, flattened and split into real and imaginary parts, becoming a feature vector of dimension 882 which is fed to the NN input layer. The ideal HGM reconstruction of M is found by numerical simulation and the resulting image is compressed to a 50×50 array, which is normalised, flattened, and used as a 2500-dimensional labels vector.

We use the hyperbolic tangent as activation function for the first layer. For

the second layer, we use a ReLU activation for train speed purposes and, after the third, a sigmoid, in order to adapt to the $[0, 1]$ range of the labels. Note that if we had objects with more than two (binary) intensity levels we would have to scrap the sigmoid and simply have a final layer with no activation function: we would be performing linear regression, as we would want to preserve the intensity levels. However, since our images are binary, we can "cheat" and use the sigmoid to increase the contrast between zeros and ones, therefore easing the NN training.

The optimization method is Adaptive Moment Estimation (Adam), with a learning rate of 10^{-4}, exponential decay rate moving parameters for the first and second moment estimates $(0.9, 0.999)$ and weight decay 0. We set the batch size to 512. We notice that a large batch size, apart from reducing the training time, allows achieving lower training and cross-validation losses. The loss function is the (per pixel) mean squared error (MSE) loss, which makes the NN effectively behave as a nonlinear regressor. The NN is trained for 900 epochs, achieving training and validation $\log_{10}$-loss of -6.674 and -6.498 per sample, respectively, as illustrated in the inset of Fig. 7.13.

Train, validation and test datasets

The training/cross-validation dataset consisted of

- 20000 frames that are tiled with black and white rectangles. The tile size is randomly chosen from 10 to 50 pixels in each dimension and the binary level (black vs white) of each tile is also random, sampled from a uniform distribution between 0.2 and 0.8; the entire matrix is then randomly rolled along both axes and randomly oriented, Fig. 7.12(C);

- 3000 random sets of 1 to 5 lines with random orientation, position and width (10 to 70 pixel), Fig. 7.12(D);

- 3000 random sets of 1 to 5 segments of elliptic rings with randomly variable eccentricity and width (minor and major axis from 40 to 120 pixel, width from 10 to 15 pixel), Fig. 7.12(E);

- 58 single squares of sizes 30 pixel and 70 pixel at various positions within the 210×210 frame;
- 441 squares of 30 pixel size at variable positions within the 210×210 frame, from which a smaller square of 10 pixel size is subtracted, with variable positions of the smaller square with respect to the larger one;
- a blank image;
- a square occupying the entire 210×210 pixel frame.

During the training time, the train/cross-validation dataset is randomly split into 90% for training and 10% for validation.

The test dataset consists of

- University of Oxford logo, divided into 10×3 frames [Fig. 7.14(A), top].
- University of Oxford coat of arms, divided into 10×12 frames [Fig. 7.14(A), bottom].
- 96 text symbols from the Latin alphabet and special ASCII characters [Fig. 7.15(A)].
- 56 centred pairs of vertical lines of 10 pixel thickness and varied spacing, used to evaluate the Rayleigh distance [Fig. 7.16(A)].

These datasets can be downloaded at [171]. Acquiring the training set (26501 DMD frames) takes about 10 hours, while the test set (Oxford logos, text symbols and line pairs) is limited to 302 DMD frames and is hence acquired as a single batch.

7.3 Results

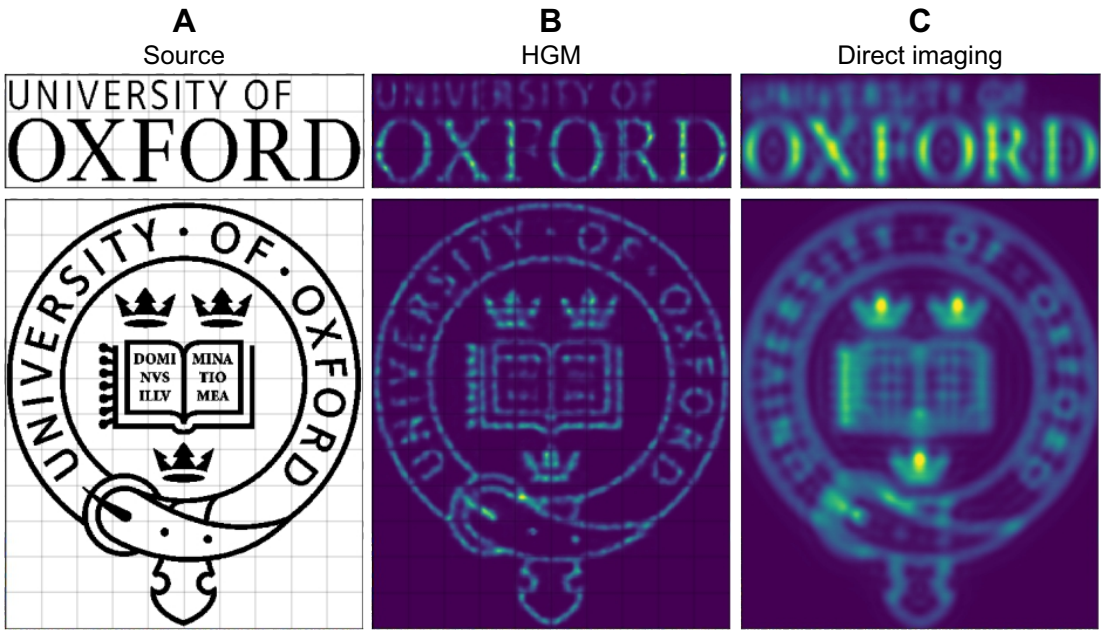


Figure 7.14: Original (A), HGM-reconstructed (B) and camera (C) images of the Oxford logo and coat-of-arms test sets. The grid shows the division of the images into square sub-images that are fed to HGM sequentially.

7.3.1 Qualitative results

The test set ground truths and HGM reconstructions are shown in Figs. 7.14(A,B), 7.15(A,B) and 7.16(A,B). The logo and coat-of-arms have been split into respectively 30 and 120 smaller square rasters of size 210×210 DMD pixels (shown by the grid in Fig. 7.14), each displayed on the DMD and fed to the NN in sequence. This procedure is equivalent to the transverse scanning of the object with the stride of $210 \times 7.56 = 1588$ μm.

In order to assess the resolution improvement of HGM with respect to conventional microscopy, we perform direct intensity measurements of the image field with a CMOS camera at the image plane of the objective lens. For the parallel lines and alphanumerical symbols, a single direct image of the 210×210 pixel object is acquired. For the two larger logos, we perform the scanning with a smaller stride (10 DMD pixels). The reason is that direct imaging is more sensitive to aberrations, and border effects occur when joining images taken with a high stride. These results are shown in Figs. 7.14(C), 7.15(C) and 7.16(C).

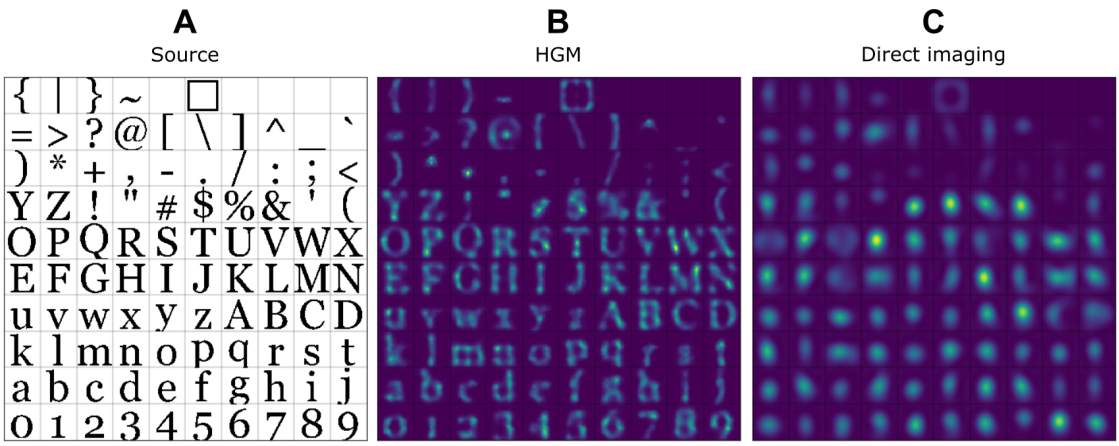


Figure 7.15: Original (A), HGM-reconstructed (B) and camera (C) images of the *text symbols* test set.

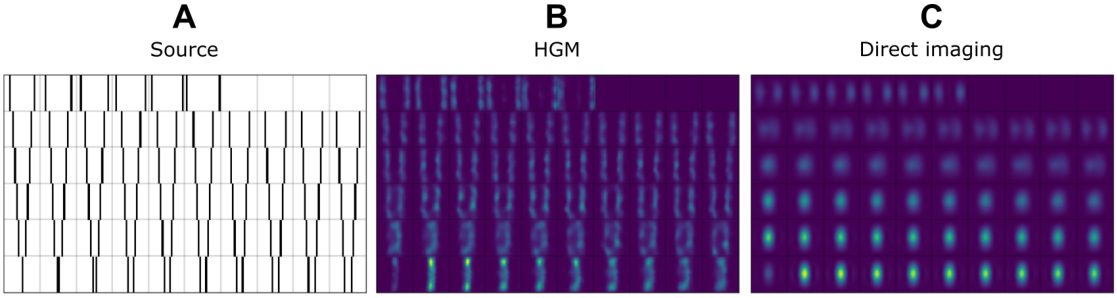


Figure 7.16: Original (A), HGM-reconstructed (B) and camera (C) images of the *pairs of lines* test set.

Qualitatively, we can see that the HGM reconstructions are much sharper than direct images and allow us to see fine details and features which otherwise could not be distinguished.

7.3.2 Quantitative results

In Fig. 7.17, we quantitatively benchmark the resolution gain by imaging pairs of parallel lines. Examples are shown in Fig. 7.17(A-C), whereas Fig. 7.17(D) plots the HGM resolution as a function of the number of HG modes used in both dimensions. To quantify the resolution, the classic Rayleigh criterion is used, i.e. that two sources are considered resolved when the intensity at their midpoint is at most 75% of the maximum intensity. We find that HGM with up to the $HG_{20,20}$ mode can resolve two sources at approximately one-half of the the diffraction

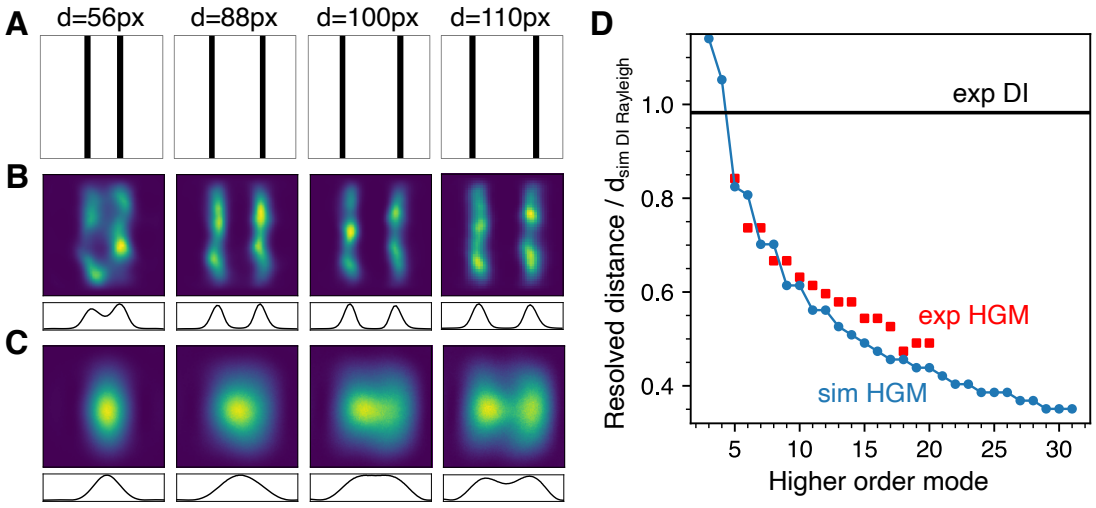


Figure 7.17: Qualitative assessment of the resolution improvements of HGM. (A) Lines pairs of increasing spacing, (B) their reconstruction via HGM using 21×21 modes and (C) camera images. (D) Rayleigh criterion for experimental (red squares) and simulated (blue connected dots) HGM vs number of used modes, normalized to the value for simulated direct imaging, together with the experimental direct imaging (continuous black line).

limit. In other words, the HGM resolution is comparable to the direct imaging performed using a lens that is twice as wide.

We can also see from Fig. 7.17(d) that our experimental results on the resolution are close to the simulations. The HGM resolution is expected to scale approximately as the inverse square root of the number of modes in each dimension. Hence the theoretically achievable resolution enhancement is significantly higher than that shown here. In practice, an important limitation is associated with generating high-order modes, which is increasingly challenging to do with high fidelity due to the limited SLM resolution.

An alternative way of estimating the resolution improvement is via the *resolution enhancement* measure, Eq. (4.92), but with the figure of merit being the mean squared error (MSE) between the reconstructed images and the original objects instead of the Fisher information. As in Eq. (4.92), we calculate the MSE $M_{\mathrm{DI}}(\mathrm{NA})$ between theoretically calculated direct camera images with varying numerical apertures and the original objects, and then compare these values with the MSE $M_{\mathrm{HGM}}(\mathrm{NA}_0)$ for the HGM reconstruction with the numerical aperture NA_0. Because HGM is generally superior to direct imaging, $M_{\mathrm{DI}}(\mathrm{NA}_0) > M_{\mathrm{HGM}}(\mathrm{NA}_0)$. However,

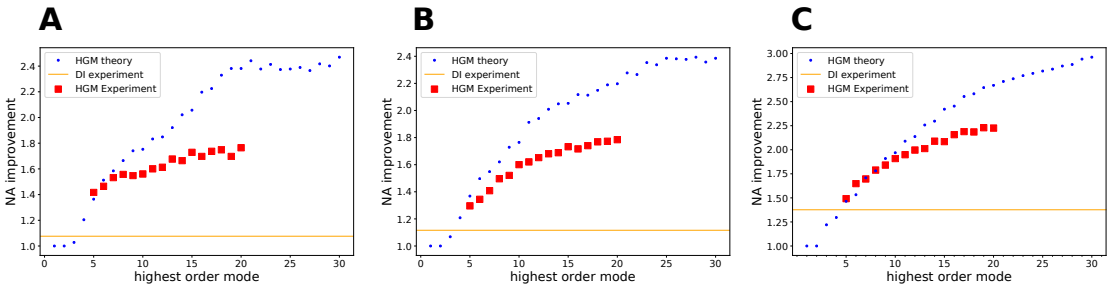


Figure 7.18: Resolution improvement for Oxford logo, Oxford coat of arms and and the text symbols test set, evaluated through the MSE loss as described in the text. For the 20×20 reconstructions, we achieved resolution improvements of 1.76, 1.78 and 2.22, respectively.

the function $M_{\mathrm{DI}}(\mathrm{NA})$ monotonically decreases with NA and becomes equal to $M_{\mathrm{HGM}}(\mathrm{NA}_0)$ at some $\mathrm{NA}_1 > \mathrm{NA}_0$: in other words, the image quality obtained via direct imaging with the numerical aperture NA_1 is similar to that with HGM and numerical aperture NA_0. The quantity $\mathrm{NA}_1/\mathrm{NA}_0$ corresponds to the resolution enhancement and is plotted, as a function of the number of modes used in HGM, in Fig. 7.18.

7.3.3 Comparison with direct imaging and deconvolution

To further evaluate the performance of HGM, we post-process the experimental camera images using one of the commonly used deconvolution algorithms, Richardson-Lucy (RL) deconvolution [71, 173]. RL is an iterative algorithm that aims to estimate the original object $O(x, y)$, given the point spread function $\psi(x, y)$ and the experimental camera image $I(x, y)$, via

$$\hat{O}_{t+1} = \hat{O}_t \cdot \left(\frac{I}{\hat{O}_t * \psi} * \psi \right) \tag{7.25}$$

with $\hat{O}_t$ being the estimate of the object at the iteration t and asterisks denote convolution.

Fig. 7.19(B-D) shows the result of RL applied to the test datasets (two University of Oxford logos and the text symbols test), after 1000 iterations and using the Airy point spread function (PSF) reported in Fig. 7.19A. The size of the Airy PSF is

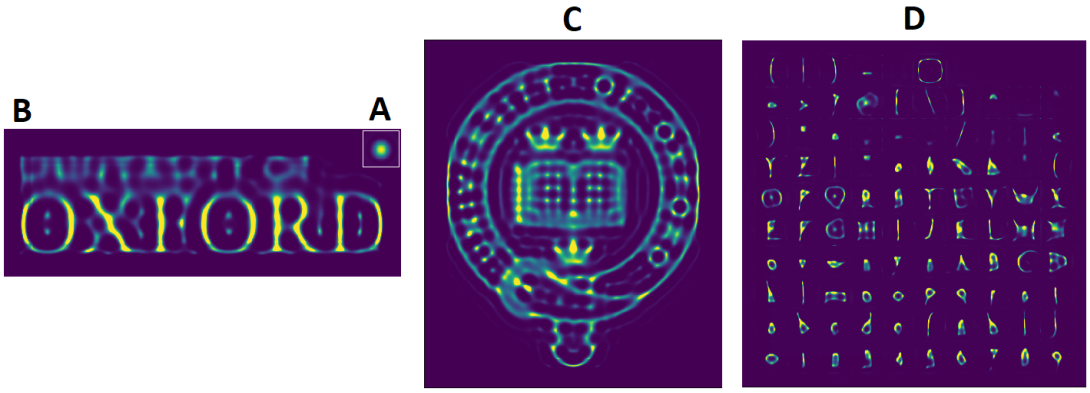


Figure 7.19: RL applied to the camera images of the test datasets using the Airy point spread function (A). Estimates of the two University of Oxford logos (B,C) and text symbols (D).

chosen to match the experimental PSF, with 45 camera pixels from its centre to the radius of the first zero. The program used to perform RL can be found in [171].

By qualitatively comparing the HGM predictions (Fig. 7.14 and Fig. 7.15) with the results of RL (Fig. 7.19), we conclude that HGM allows for better reconstruction of the original object, with a higher resolution and fewer artifacts. On a side note, we did not use the deconvolved parallel lines to estimate resolution since they were too "wiggly" for the Rayleigh criterion.

7.3.4 Shot noise behaviour

In Fig. 7.20 we assessed the NN's efficiency against shot noise. We did this in simulations, instead of in the actual experimental setup, because we didn't have means of performing photon counting. For the assessment, we simulated training and test data for different shot noise values (which corresponds to different numbers of photons in the acquisitions). We then analysed how a NN trained with a specific noise level would behave when tested with the other test sets. We observe that although the reconstruction quality degrades with the falling number of photons, we are still able to obtain meaningful results even with a very limited number of photons. Remarkably, HGM with as low as 10^3 photons per image section yields the same reconstruction quality as direct imaging in the absence of the shot noise.

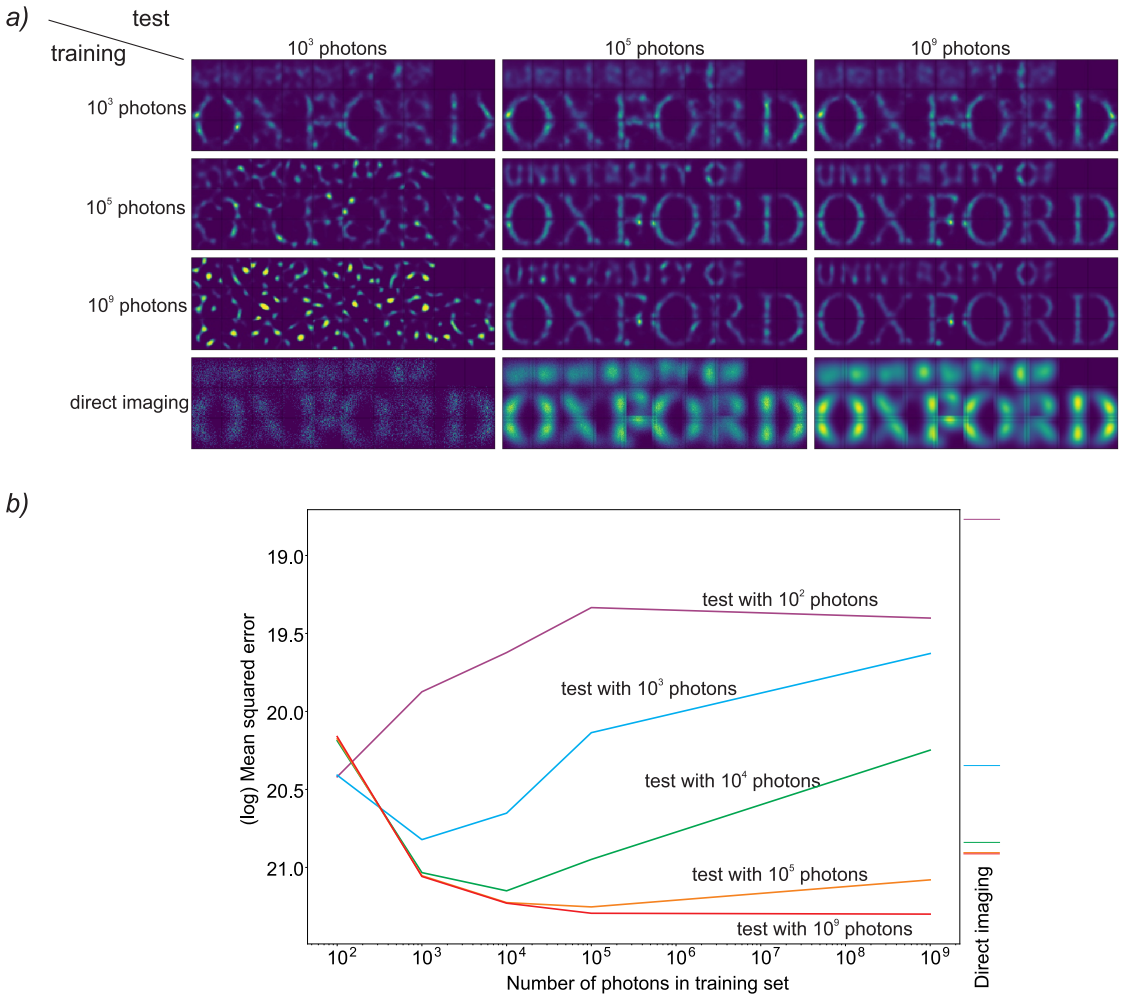


Figure 7.20: Simulated Oxford logo reconstruction via direct imaging and HGM using heterodyne detection with different photon numbers available per image section during training and test phases. a) Examples. b) (Logarithm of) mean squared error between the noisy reconstructed logo and the original object. Horizontal lines on the right correspond to the error obtained for the direct imaging reconstructions with the same photon numbers.

In Fig. 7.20, we also analyze how the reconstruction quality is affected when the NN trained on noisy data with certain number of photons available is used to reconstruct an image from data acquired with a different number of photons. For a given number of photons in the test set, the optimal reconstruction quality is obtained when training set has been acquired with the same number of photons (i.e. has the same noise level). If difference is present, the reconstruction quality is affected less significantly if the test set has a better signal-to-noise ratio than the training set, rather than the other way around.

7.4 Discussion

7.4.1 What about Fisher information?

After the long and elaborate statistical discussions in earlier chapters, it might seem odd that for our experiment we threw all that out of the window and went back to the Rayleigh criterion for the estimation of resolution. Even with the MSE measure of Fig. 7.18, we're still ignoring the particle nature of light and all statistical phenomena. However, although the Cramér-Rao bound and all that jazz are invaluable for theoretical research, in experiments the Rayleigh limit is still widely used and respected. Furthermore, the rigorous (although not as general as we might have wanted) results from Chapters 4 and 5 back up our resolution claims in the experiment.

Regardless, one of the future plans for our laboratory is to perform HGM with full photon counting, and unambiguously prove its superiority over direct imaging with statistical rigour.

It is important to emphasize that the reasons for the success of heterodyning HGM are different in the incoherent and coherent cases. For the former, it is due to the fact that mode projections can capture information in the spatial correlations, which would be otherwise discarded in conventional imaging. For the latter, it is the fact that heterodyne detection is advantageous when there are uncontrollable relative phases between the sources' points.

Finally, we should point out that the HGM reconstruction (Eq. (7.10)) is only one possible estimator for the mode sorted measurements we performed. It might be possible that, for the same measurement and data we performed, a different estimator might yield better results. We can only benchmark HGM against the fundamental limits with full photon counting.

7.4.2 What about incoherent sources?

In practice, most objects in imaging applications emit incoherent light. If the light's bandwidth is narrow enough, we can use the incoherent HGM scheme of Section 7.1.3 with the mode superpositions. For broadband radiation, heterodyne detection might be unfeasible due to the limited bandwidth of electronics. A solution is to replace it with spatial mode demultiplexing, i.e., directly decomposing the image field into its mode components and measuring their intensities with photodetectors. This scheme, although even more advantageous with respect to achievable resolution (See Chapter 4), is limited by current mode sorting technology.

Using the incoherent HGM scheme, we have preliminary experimental results that confirm HGM can achieve approximately twofold resolution enhancement over direct imaging for incoherent emitters.

7.4.3 What about microscopic samples?

In order to adapt our setup to the microscopic realm, we would need to perform imaging with a microscopic objective lens and microscopic sample objects. Since the mode sorting stage of HGM happens at the image plane (which can have any magnification we want), the rest of the setup wouldn't need very drastic adaptations.

The main issue would be with respect to the NN training, as finding a large amount of arbitrary and static microscopic samples could be hard and/or cost prohibitive. For that, one could rely on off-the-shelf calibration slides and microplates for optical microscopes [174]. A calibration slide of a few tens of μm size, containing several thousand training objects of size 0.5–1 μm, can be fabricated with a resolution of a few tens of nanometers by way of lithography or laser writing. This slide can be scanned in front of the microscope objective in various orientations to increase the training set size.

If the variety and amount of the training data are still not sufficient, one could implement data augmentation techniques [175], adapting them to the HGM NN. For incoherent imaging, a single point source scanned across a sufficiently fine grid of positions around the imaging window could be enough, as the linearity of

the problem would allow us to create any "synthetic" dataset from these points and train the NN properly. For coherent imaging, however, the situation is more complicated as the problem is no longer linear and relative phases might occur, and no simple solutions for that are known.

7.4.4 Other adaptations

HGM is a vastly simpler and cheaper alternative to many existing super-resolution methods. Furthermore, its passive nature permits universal application in a wide variety of imaging scenarios, including those not accessible by existing schemes. HGM can also be combined with other imaging techniques to further increase the resolution. This could open up a whole new direction in both the academic and industrial sides of optical imaging.

HGM can also be tweaked in a plethora of ways. It could be used in a confocal setup, as proposed in Ref.[71]. The mode sorting could also occur at the aperture plane, which means that HGM can perform lensless imaging. Since it is essentially a single-pixel imaging scheme [176, 177], it could be used in situations where a camera of photodetector array can't be used.

For incoherent imaging, we could use it with the schemes proposed in Chapter 4, such as quadrant displaced and scanning mode sorting. It could also benefit from adaptive strategies, where the mode sorting basis is continuously refined by feedback loops. In that case, the supervised NN could be replaced by a reinforcement learning algorithm.

As we can see, HGM is simply the starting point of the journey, and only time will tell where it will be able to take us.

8 | Conclusion

8.1 Final words

In the never ending tale of exploration that is optical imaging, the diffraction limit, or *Rayleigh's curse*, has presented itself as an insurmountable wall. It could be circumvented by neglecting any of its hypotheses, but this meant abdicating of generality: the explorer would be taken to a very rocky path. With the novel field of quantum-inspired super-resolution, we could finally step over this wall and continue our regular journey.

Having crossed the wall, we are barely at is feet. In this thesis we aimed to contribute at least a tiny bit to this new journey. Firstly, we hope its pedagogical style can help newcomers, specially early researchers, to understand and get into the field. Secondly, we expect our results to have filled important gaps in the literature, specially on the experimental side. Finally, we hope our discussions enlighten researchers into which next steps to take. We have presented several new roads in this already highly branched field, but we have provided only the first step.

8.2 My vision for the future

The modus operandi of quantum inspired super-resolution (QISR) has been the following. First, we derive the quantum resolution bounds of a particular problem and see if there's room for improvement over classical methods. Second, we try to find optimal measurements that achieve (or at least approach) the quantum limit. Third, we look for good estimators that achieve or approach the CRB given a dataset acquired via the optimal measurements.

Of these three steps, only the first is analytical. Finding optimal measurements and estimators is more of an art than science, and it requires domain (or even extra-domain) knowledge from the researcher. For estimators in particular, even though the maximum likelihood estimator is asymptotically optimal, it is not very useful in real experimental scenarios, and in practice we need estimators that are also robust to the many sources of uncontrollable noise that appear in real life.

In this thesis we have found quantum bounds and optimal (or near-optimal) measurements for several imaging scenarios. Furthermore, for the specific case of coherent 2D imaging, we have used the HGM equations, aided with neural networks, in order to implement a suitable estimator for heterodyning mode sorting. Although we haven't performed proper photon counting, we conjecture that our method can be applied to more photon-limited cases and still be efficient. What made our whole work feasible was the machine learning step, which provided good estimators even among many experimental imperfections.

We have also seen that quantum-optimal measurements for general imaging scenarios depend on the ground truth and, therefore, are very likely to be adaptive. Adaptive measurements can largely benefit from machine learning methods as well, as shown in recent works [178–180].

In view of this, I personally believe that the future of QISR, and maybe optical microscopy in general, is to perform computational processing before and after light detection. The latter, data post-processing, is already being done and has greatly contributed to the advancement of microscopy in the past decades [181–186]. The former, *coherent processing of light*, is what QISR brings to the table. Mode sorting is a simple special case of it, where the coherent basis is pre-determined. The ultimate QISR microscope would use machine learning to integrate both the coherent and non-coherent processing steps, taking optical microscopy to its physical limits.

Appendices

A | Density matrices review

In this appendix we review the basic postulates of quantum mechanics via the density matrix formalism [137, 187].

Every quantum system is associated with a Hilbert space (say, $\mathcal{H}$), called a *state space* [137]. Quantum states in $\mathcal{H}$ are completely described by *density operators* ρ, which are positive semi-definite operators with unit trace. Two state spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ can be composed via the tensor product operation, resulting in the state space $\mathcal{H}_1 \otimes \mathcal{H}_2$. Furthermore, a sub-state space can be discarded via the *partial trace* operation.

If ρ has rank one it can be decomposed as $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a vector in $\mathcal{H}$, and it is said to be a *pure state*. The $|\psi\rangle$ vector has unit norm and is unique up to a global phase

$$|\psi\rangle \sim e^{i\phi} |\psi\rangle , \tag{A.1}$$

which is a consequence of the unitary trace of ρ. Otherwise, ρ is said to be a *mixed state* and it can be decomposed as a *statistical mixture* of pure states

$$\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i| , \tag{A.2}$$

where $|\psi_i\rangle\langle\psi_i|$ are pures states and λ_i non-negative numbers that sum to one, which therefore can be interpreted as probabilities [137]. It is important to note that the decomposition of Eq. (A.2) is not unique.

Mixed states can be seen as pure states in a larger state space, which allows us to interpret them as states with respect to which we only have "partial knowledge". Mathematically, this means that they admit *purifications*

$$\rho = \mathrm{Tr}_A \left[|\Psi\rangle\langle\Psi| \right] , \tag{A.3}$$

where the vectors $|\Psi\rangle$ belong to some larger space $\mathcal{H}\otimes\mathcal{A}$. The "auxiliary" space $\mathcal{A}$ is usually called an *ancilla space*. Any two purifications are related by a partial isometry

$$|\Psi\rangle = (I \otimes V)\,|\Phi\rangle\,, \tag{A.4}$$

where V is a linear isometry [188], i.e., it obeys $V^\dagger V = I$. If it also obeys $VV^\dagger = I$ it becomes an unitary operation. However, since the ancilla spaces might have different dimensions, this is usually not possible.

A general quantum measurement is represented by a *positive operator-valued measurement* (POVM), which is a set of positive semi-definite operators $\Pi = \{\Pi_m\}$ that sum to the identity matrix, $\sum_m \Pi_m = I$ [137]. After performing the measurement $\{\Pi_m\}$ over ρ, we obtain the outcome m with probability

$$\Pr(m) = \mathrm{tr}\left\{\Pi_m \rho\right\}. \tag{A.5}$$

Pure quantum states evolve via unitary operations,

$$\rho' = U\rho U^\dagger. \tag{A.6}$$

General mixed states evolve via *quantum channels*

$$\rho' = \mathcal{E}(\rho), \tag{A.7}$$

which are completely-positive, trace-preserving transformations [188]. In other worlds, quantum channels take density operators to density operators.

B | Vectorization of a matrix

The *vectorization* $|M\rangle\rangle$[1] of a matrix M consists of "stacking" its columns on top of each other, in order to form a vector [136, 188]. For example, if

$$M = \begin{bmatrix} a & c \\ b & d \end{bmatrix}, \tag{B.1}$$

then $|M\rangle\rangle = (a\ b\ c\ d)^\intercal$. More formally, given H^A and H^B (finite-dimensional) Hilbert spaces, this operation defines an isomorphism between linear operators from H^A to H^B and bipartite vectors in $H^A \otimes H^B$ via

$$|M\rangle\rangle = (I_A \otimes M)\,|\Omega\rangle\,, \tag{B.2}$$

where M and $|M\rangle\rangle$ are the linear operator and bipartite vector, respectively, I_A is the identity in H_A and $|\Omega\rangle$ is the (unnormalized) Bell state in $H^A \otimes H^A$,

$$|\Omega\rangle = \sum_i |i\rangle \otimes |i\rangle\,, \tag{B.3}$$

where $\{|i\rangle\}$ is an orthonormal basis of H^A.

When dealing with vectorizations, the following two relations will be necessary [136]:

$$(I \otimes M)\,|\Omega\rangle = (M^\intercal \otimes I)\,|\Omega\rangle\,, \tag{B.4}$$

$$\langle\Omega|\, I \otimes M\,|\Omega\rangle = \operatorname{Tr} M. \tag{B.5}$$

With them, it is easy to show many property of vectorization. For example, the trace relation,

$$\begin{aligned} \langle\langle M|N\rangle\rangle &= \langle\Omega|\,(I \otimes M^\dagger)(I \otimes N)\,|\Omega\rangle \\ &= \langle\Omega|\,(I \otimes M^\dagger N)\,|\Omega\rangle = \operatorname{Tr} M^\dagger N, \end{aligned} \tag{B.6}$$

[1] Usually denoted by $\operatorname{vec}(M)$ in the literature [67, 136].

the double product,

$$\begin{aligned} |MN\rangle\rangle &= (I \otimes M)(I \otimes N)\,|\Omega\rangle \\ &= (N^{\intercal} \otimes I)(I \otimes M)\,|\Omega\rangle = (N^{\intercal} \otimes I)|\,M\rangle\rangle, \end{aligned} \tag{B.7}$$

and the triple product,

$$\begin{aligned} |MNO\rangle\rangle &= (I \otimes M)(I \otimes N)(I \otimes O)\,|\Omega\rangle \\ &= (O^{\intercal} \otimes I)(I \otimes M)|\,N\rangle\rangle = (O^{\intercal} \otimes M)|\,N\rangle\rangle. \end{aligned} \tag{B.8}$$

C | Coherent states and displacement operators

In this appendix we review a few results in quantum optics necessary for the rest of the text. They mostly concern coherent states of the electromagnetic field and derived concepts. The interested reader can find more details in the excellent Refs. [147, 189, 190].

Ladder operators

A single mode of a quantized electromagnetic field is described by *ladder operators* $\hat{a}$ and $\hat{a}^\dagger$ [189]. These operators obey the commutation relation (with normalized units)

$$[\hat{a}, \hat{a}^\dagger] = 1. \tag{C.1}$$

The operator $\hat{n} = \hat{a}^\dagger \hat{a}$ has eigenstates

$$\hat{n} \left|n\right\rangle = n \left|n\right\rangle, \tag{C.2}$$

which are called *Fock states*. The ladder operators act on the Fock states as

$$\hat{a} \left|n\right\rangle = \sqrt{n} \left|n-1\right\rangle, \tag{C.3}$$

$$\hat{a}^\dagger \left|n\right\rangle = \sqrt{n+1} \left|n+1\right\rangle. \tag{C.4}$$

Since $\hat{n}$ is a positive semi-definite operator, its eigenvalues are non-negative. This fact and Eq. (C.3) imply that the lowest eigenvalue of $\hat{n}$ is zero, corresponding to the eigenstate $\left|0\right\rangle$. Eq. (C.4) implies that the remaining eigenvalues are integers, unbounded from above, and that the eigenstates can be written as

$$\left|n\right\rangle = \frac{(\hat{a}^\dagger)^n}{\sqrt{n!}} \left|0\right\rangle. \tag{C.5}$$

This representation can be interpreted as if each application of the $\hat{a}^\dagger$ operator created an excitation of the electromagnetic field at its corresponding mode. These excitations are dubbed *photons*. Under this light, the $\hat{n}$ operators are usually called *photon-number operators* and its eigenstates $|n\rangle$ are *n-photon states*. The $|0\rangle$ state is called *0-photon state* or *vacuum state* and corresponds to no excitations in the mode. It is also sometimes written as $|\text{vac}\rangle$. Due to their action in Eqs. (C.3) and (C.4), the $\hat{a}^\dagger$ and $\hat{a}$ operators are respectively called *creation* and *annihilation* operators.

Coherent states

The eigenstates of the annihilation operator,

$$\hat{a}\,|\alpha\rangle = \alpha\,|\alpha\rangle\,, \tag{C.6}$$

are called *coherent states*. The eigenvalue α is an arbitrary complex number, called the *complex amplitude* of the state. In terms of the photon-number states, they can be expanded as

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\,, \tag{C.7}$$

which implies that the overlap between Fock and coherent states is

$$\langle n|\alpha\rangle = \frac{e^{-|\alpha|^2/2}\alpha^n}{\sqrt{n!}}. \tag{C.8}$$

Displacement operators

Coherent states can be neatly expressed as

$$|\alpha\rangle = \mathcal{D}_{\hat{a}}(\alpha)\,|\text{vac}\rangle\,, \tag{C.9}$$

where

$$\mathcal{D}_{\hat{a}}(\alpha) = e^{\alpha\hat{a}^\dagger - \alpha^*\hat{a}}. \tag{C.10}$$

is called the *displacement operator* [189], since it "displaces" the state's amplitude in the complex plane by a value of α. It can be equivalently expressed as

$$\mathcal{D}_{\hat{a}}(\alpha) = e^{-|\alpha|^2/2} e^{\alpha\hat{a}^\dagger} e^{-\alpha^*\hat{a}} \tag{C.11}$$

$$= e^{|\alpha|^2/2} e^{-\alpha^*\hat{a}} e^{\alpha\hat{a}^\dagger}. \tag{C.12}$$

Displacement operators also displace the ladder operators as in

$$\mathcal{D}_{\hat{a}}^{\dagger}(\alpha)\hat{a}\mathcal{D}_{\hat{a}}(\alpha) = \hat{a} + \alpha, \tag{C.13}$$

$$\mathcal{D}_{\hat{a}}(\alpha)\hat{a}^{\dagger}\mathcal{D}_{\hat{a}}^{\dagger}(\alpha) = \hat{a}^{\dagger} + \alpha^{*}. \tag{C.14}$$

They are unitary,

$$\mathcal{D}_{\hat{a}}^{\dagger}(\alpha)\mathcal{D}_{\hat{a}}(\alpha) = \mathcal{D}_{\hat{a}}(\alpha)\mathcal{D}_{\hat{a}}^{\dagger}(\alpha) = I, \tag{C.15}$$

and the adjoint of a displacement can be interpreted as a displacement in the opposite direction:

$$\mathcal{D}_{\hat{a}}^{\dagger}(\alpha) = \mathcal{D}_{\hat{a}}(-\alpha). \tag{C.16}$$

Different displacements do not commute, and two successive displacements α and β act as a single displacement $\alpha + \beta$ plus a complex phase:

$$\mathcal{D}_{\hat{a}}(\alpha)\mathcal{D}_{\hat{a}}(\beta) = e^{\alpha\beta^{*}-\alpha^{*}\beta}\mathcal{D}_{\hat{a}}(\beta)\mathcal{D}_{\hat{a}}(\alpha) = e^{(\alpha\beta^{*}-\alpha^{*}\beta)/2}\mathcal{D}_{\hat{a}}(\alpha+\beta) \tag{C.17}$$

P representation

Coherent states form a resolution of the identity,

$$I = \frac{1}{\pi}\int |\alpha\rangle\langle\alpha|\,\mathrm{d}^{2}\alpha, \tag{C.18}$$

where $\mathrm{d}^{2}\alpha = \mathrm{d}\,\mathrm{Re}\{\alpha\}\mathrm{d}\,\mathrm{Im}\{\alpha\}$ is an integration over the whole complex plane. This means that they form a basis for this state space. They are not, however, mutually orthogonal, as their overlap

$$\langle\alpha|\alpha'\rangle = e^{-|\alpha-\alpha'|^{2}}e^{(\alpha^{*}\alpha'-\alpha\alpha'^{*})/2} \tag{C.19}$$

is nonzero for amplitudes α and α' not very far apart from each other. This means that they form an *overcomplete basis*. Startling at first, this is actually a feature [189], since for this specific overcomplete basis[1] it allows us to express any state vector $|\psi\rangle$ as a superposition of coherent states $|\alpha\rangle$ for a finite range of amplitudes

[1] This property doesn't hold for overcomplete bases in general.

$|\alpha\rangle$ (instead of the whole complex plane). Furthermore [189], it allows us to represent any operator in a *diagonal* mixture of pure coherent operators,

$$\rho = \int P(\alpha)\,|\alpha\rangle\,\langle\alpha|\,\mathrm{d}^2\alpha, \tag{C.20}$$

Eq. (C.20) is known as the *Glauber-Sudarshan P representation* [189, 191, 192]. $P(\alpha)$ is a quasi-probability distribution: it is normalized to one,

$$\int P(\alpha)\;\mathrm{d}^2\alpha = 1, \tag{C.21}$$

but it can assume negative and/or singular values (more singular than the delta function). When that happens, it represents a nonclassical state [189].

The P representation is very useful when calculating averages of normally ordered operators. For example, for the operator $\langle(\hat{a}^\dagger)^n(\hat{a})^m\rangle$,

$$\begin{aligned}\langle(\hat{a}^\dagger)^n(\hat{a})^m\rangle &\equiv \mathrm{tr}\Big((\hat{a})^m\rho(\hat{a}^\dagger)^n\Big)\\ &= \int P(\alpha)\,\mathrm{tr}\Big((\hat{a})^m\,|\alpha\rangle\,\langle\alpha|\,(\hat{a}^\dagger)^n\Big)\mathrm{d}^2\alpha\\ &= \int P(\alpha)\alpha^m(\alpha^*)^n\mathrm{d}^2\alpha.\end{aligned} \tag{C.22}$$

The average $\langle\hat{n}\rangle = \langle\hat{a}^\dagger\hat{a}\rangle$ is called the *average photon number* of ρ.

Multimode fields

Finally, a multimode field can be described by a set of ladder operators $\hat{\boldsymbol{a}} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_S)^\intercal$ such that

$$[\hat{a}_i, \hat{a}_j] = [\hat{a}_i^\dagger, \hat{a}_j^\dagger] = 0 \tag{C.23}$$

$$[\hat{a}_i, \hat{a}_j^\dagger] = \delta_{ij}. \tag{C.24}$$

Their coherent states are represented by $|\alpha\rangle_j$ and a coherent state of the whole field is

$$|\boldsymbol{\alpha}\rangle = \bigotimes_{j=1}^{S} |\alpha_j\rangle_j\,, \tag{C.25}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_S)^\intercal$. Its P representation is

$$\rho = \int P(\boldsymbol{\alpha})\,|\boldsymbol{\alpha}\rangle\,\langle\boldsymbol{\alpha}|\,\mathrm{D}\boldsymbol{\alpha}, \tag{C.26}$$

where $\mathrm{D}\boldsymbol{\alpha} \equiv \mathrm{d}^2\alpha_1\ \mathrm{d}^2\alpha_2 \cdots \mathrm{d}^2\alpha_S$. If the field components along different modes are uncorrelated, then $P(\boldsymbol{\alpha})$ simplifies to

$$P(\boldsymbol{\alpha}) = \prod_{j=1}^{S} P_j(\alpha_j). \tag{C.27}$$

D | Quantum theory of imaging

In this appendix we develop the quantum theory of optical imaging, which underlies the quantitative works of chapters 4 and 5. It is assumed the reader is already familiar with the basics of quantum mechanics (reviewed in Appendix A) and quantum optics (reviewed in Appendix C).

D.1 Image plane operators

Recall the imaging scenario described in Fig. 3.2. The quantum version of the Huygens-Fresnel diffraction integral, Eq. (3.1), is given by simply replacing the electric fields with their operator counterparts [140],

$$\hat{E}_z(\vec{r},t) = \frac{e^{i\omega z}}{i\lambda z}\int_{\mathcal{A}_0} \mathrm{d}\vec{r'}\; e^{i\pi|\vec{r}-\vec{r'}|^2/\lambda z}\; \hat{E}_0(\vec{r'},t-z/c). \tag{D.1}$$

where the quantum-classical equivalence is determined by the averages, i.e., $E_z(\vec{r},t) = \langle \hat{E}_z(\vec{r},t)\rangle$.

With similar reasoning as in Chapter 3, it is possible to find an expression for the electric field operator at the image plane of a far field imaging system with PSF $\psi(\vec{r})$. It is more useful for our purposes, however, to see how the mode ladder operators transform from object to image plane. Thus, consider the creation operator $\hat{a}_j^\dagger$ for a field at position $\vec{r_j}$ at the object plane. At the image plane, it undergoes the transformation [33, 140]

$$\hat{a}_j^\dagger \to \sqrt{\eta}\hat{\psi}_j^\dagger + \sqrt{1-\eta}\hat{v}_j^\dagger. \tag{D.2}$$

The *PSF operator* $\hat{\psi}_j^\dagger$ is given by

$$\hat{\psi}_j^\dagger = \int \mathrm{d}\vec{r}\; \psi(\vec{r}-\vec{r}_j)\hat{a}_{\vec{r}}^\dagger, \tag{D.3}$$

where $\hat{a}^\dagger_{\vec{r}}$ is the creation operator at the image plane position $\vec{r}$. It represents the image of a point source originating at $\vec{r}_j$ at the object plane. The term η is an attenuation factor and $\hat{v}^\dagger_j$ is a creation operator for the environmental modes that come from outside the aperture and inevitably couple to the image plane [42]. The term $\sqrt{1-\eta}\hat{v}^\dagger_j$ is also necessary to yield the correct field operator commutation relations [140]. The attenuation factor reflects the loss of overall optical power between object and image plane, and will be conveniently omitted sometimes as it is simply a constant proportionality factor. The object and image plane operators here introduced are pictorially represented in Fig. D.1.

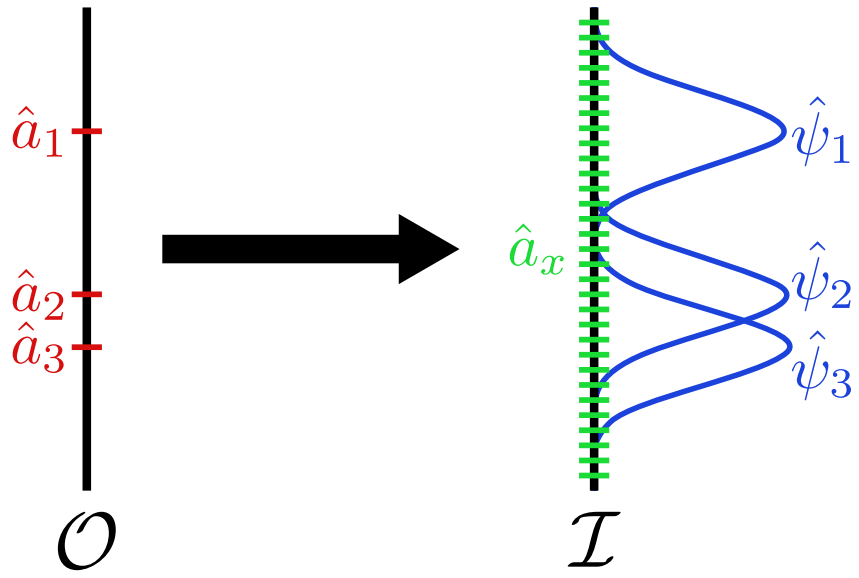


Figure D.1: Pictorial representation of the object and image plane operators.

The image plane ladder operators obey the canonical commutation relations (CCR)[1] [189]

$$[\hat{a}_{\vec{r}}, \hat{a}^\dagger_{\vec{r'}}] = \delta^2(\vec{r} - \vec{r'}), \tag{D.4}$$

which implies that the PSF operators obey

$$[\hat{\psi}_i, \hat{\psi}^\dagger_j] = \int \mathrm{d}\vec{r}\psi^*(\vec{r} - (\vec{r}_i - \vec{r}_j))\psi(\vec{r}) \equiv \Delta(\vec{r}_i - \vec{r}_j). \tag{D.5}$$

[1]The relations $[\hat{a}^\dagger_{\vec{r}}, \hat{a}^\dagger_{\vec{r'}}] = [\hat{a}_{\vec{r}}, \hat{a}_{\vec{r'}}] = 0$, which are also part of the CCR, were not included for brevity and because they are somewhat trivial.

The term $\Delta(\vec{r}_i - \vec{r}_j)$ is the overlap between the images of sources originating from $\vec{r}_i$ and $\vec{r}_j$ at the object plane, and is a quantum manifestation of the blurring caused by the imaging system's aperture. It implies that two PSF operators are not orthogonal if the image of their sources overlap, and this non-orthogonality means there is no measurement that can perfectly discriminate these sources [6, 33]. At the limit of an aperture with very large diameter or, equivalently, of a very narrow PSF, $\Delta(\vec{r}_i - \vec{r}_j) \to \delta^2(\vec{r}_i - \vec{r}_j)$ and the operators become orthogonal. The overlap term also obeys $\Delta(0) = 1$ and $\Delta^*(\vec{r}) = \Delta(-\vec{r})$ [19]. As an example, for a Gaussian PSF with waist σ we have $\Delta(\vec{r}) = e^{-\|\vec{r}\|^2/8\sigma^2}$.

The PSF operators can be used to define the displacement operators (See Appendix C)

$$\mathcal{D}_{\hat{\psi}_j}(\alpha) = e^{\alpha\hat{\psi}_j^\dagger - \alpha^*\hat{\psi}_j}, \tag{D.6}$$

which obey the commutation relations

$$\mathcal{D}_{\hat{\psi}_i}(\alpha_i)\mathcal{D}_{\hat{\psi}_j}(\alpha_j) = \mathcal{D}_{\hat{\psi}_j}(\alpha_j)\mathcal{D}_{\hat{\psi}_i}(\alpha_i)e^{-2i\,\mathrm{Im}\left\{\alpha_i^*\alpha_j\Delta(\vec{r}_i - \vec{r}_j)\right\}}. \tag{D.7}$$

As can be seen, due to the overlap term the displacement operators for different $\hat{\psi}_j$ do not commute.

D.2 P-representation

Consider S light sources in the object plane, each located at some position $\vec{r}_j$. The quantum state of light can be represented via its (multimode) Glauber-Sudarshan P representation $P_{\mathcal{O}}$ (See Appendix C) [6, 42],

$$\rho = \int P_{\mathcal{O}}(\boldsymbol{\alpha})\,|\boldsymbol{\alpha}\rangle\,\langle\boldsymbol{\alpha}|\,\mathrm{D}\boldsymbol{\alpha}, \tag{D.8}$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_S)^\intercal$ is a vector with the modes' complex amplitudes, $\mathrm{D}\boldsymbol{\alpha}$ a shorthand notation for $\mathrm{d}^2\alpha_1 \cdots \mathrm{d}^2\alpha_S$ and

$$|\boldsymbol{\alpha}\rangle = \bigotimes_{j=1}^{S} |\alpha_j\rangle_j\,, \tag{D.9}$$

where the $|\alpha\rangle_j$ are coherent states of the source modes, i.e., $\hat{a}_j |\alpha\rangle_j = \alpha |\alpha\rangle_j$ or, equivalently, $|\alpha\rangle_j \equiv e^{\alpha_j \hat{a}_j^\dagger - \alpha_j^* \hat{a}_j} |\text{vac}\rangle$ ($|\text{vac}\rangle$ is the vacuum state). An important quantity to know is the *average photon number* (within one coherence length) at the object plane,

$$\epsilon_0 = \sum_{j=1}^{S} \langle \hat{a}_j^\dagger \hat{a}_j \rangle = \sum_{j=1}^{S} \text{Tr}\left[\rho \hat{a}_j^\dagger \hat{a}_j\right]. \tag{D.10}$$

For the P representation, the averages of normally-ordered operators become amplitude averages (See Appendix C),

$$\epsilon_0 \equiv \sum_{j=1}^{S} \int P_{\mathcal{O}}(\boldsymbol{\alpha}) \langle \boldsymbol{\alpha}| \hat{a}_j^\dagger \hat{a}_j |\boldsymbol{\alpha}\rangle \, \mathrm{D}\boldsymbol{\alpha} = \int P_{\mathcal{O}}(\boldsymbol{\alpha}) \sum_j |\alpha_j|^2 \mathrm{D}\boldsymbol{\alpha}. \tag{D.11}$$

Eq. (D.11) can be rewritten as

$$\epsilon_0 = \int \mathrm{d}\vec{r} \left\langle |E_0^{\boldsymbol{\alpha}}(\vec{r})|^2 \right\rangle, \tag{D.12}$$

where

$$E_0^{\boldsymbol{\alpha}}(\vec{r}) = \sum_{j=1}^{S} \alpha_j \delta(\vec{r} - \vec{r}_j) \tag{D.13}$$

is the classical electric field at the object plane generated by a distribution of coherent sources with complex amplitude vector $\boldsymbol{\alpha}$. Therefore, the average photon number is simply the total average classical field intensity.

Proposition 1. *The quantum state in Eq.* (D.8) *becomes, at the image plane,*

$$\rho = \int P_{\mathcal{I}}(\boldsymbol{\alpha}) |\boldsymbol{\alpha}\rangle_\psi \langle \boldsymbol{\alpha}|_\psi \, \mathrm{D}\boldsymbol{\alpha}, \tag{D.14}$$

where the image plane "coherent states" are

$$|\boldsymbol{\alpha}\rangle_\psi = \prod_{j=1}^{S} e^{\alpha_j \hat{\psi}_j^\dagger - \alpha_j^* \hat{\psi}_j} |\text{vac}\rangle \tag{D.15}$$

and the P representation is simply rescaled as an effect of attenuation,

$$P_{\mathcal{I}}(\boldsymbol{\alpha}) = \frac{1}{\eta^S} P_{\mathcal{O}}\left(\frac{\boldsymbol{\alpha}}{\sqrt{\eta}}\right). \tag{D.16}$$

Proof. Note that Eq. (D.9) can be rewritten as

$$|\boldsymbol{\alpha}\rangle = \prod_{j=1}^{S} \mathcal{D}_{\hat{a}_j}(\alpha_j)\,|\mathrm{vac}\rangle\,, \tag{D.17}$$

where $\mathcal{D}_{\hat{a}}(\alpha) = e^{\alpha\hat{a}^\dagger - \alpha^*\hat{a}}$. Using the transformation in Eq. (D.2) and the change of variables $\boldsymbol{\beta} = \sqrt{\eta}\boldsymbol{\alpha},\ \boldsymbol{\gamma} = \sqrt{1-\eta}\boldsymbol{\alpha}$, we have that

$$\begin{aligned}
|\boldsymbol{\alpha}\rangle &= \prod_{j=1}^{S} e^{\alpha_j\hat{a}_j^\dagger - \alpha_j^*\hat{a}_j}\,|\mathrm{vac}\rangle \\
&\to \prod_{j=1}^{S} e^{\alpha_j\sqrt{\eta}\hat{\psi}_j^\dagger - \alpha_j^*\sqrt{\eta}\hat{\psi}_j + \alpha_j\sqrt{1-\eta}\hat{v}_j^\dagger - \alpha_j^*\sqrt{1-\eta}\hat{v}_j}\,|\mathrm{vac}\rangle \\
&= \prod_{j=1}^{S} e^{\beta_j\hat{\psi}_j^\dagger - \beta_j^*\hat{\psi}_j + \gamma_j\hat{v}_j^\dagger - \gamma_j^*}\,|\mathrm{vac}\rangle \\
&= \prod_{j=1}^{S} \mathcal{D}_{\hat{\psi}_j}(\beta_j) \prod_{j=1}^{S} \mathcal{D}_{\hat{v}_j}(\gamma_j)\,|\mathrm{vac}\rangle \equiv |\boldsymbol{\beta}\rangle_\psi \otimes |\boldsymbol{\gamma}\rangle_{\mathrm{env}}\,,
\end{aligned} \tag{D.18}$$

where

$$|\boldsymbol{\gamma}\rangle_{\mathrm{env}} = \prod_{j=1}^{S} \mathcal{D}_{\hat{v}_j}(\gamma_j)\,|\mathrm{vac}\rangle \tag{D.19}$$

and we have used the fact the the PSF and environment operators live in different Hilbert spaces and, therefore, commute. The total state in the image plane then becomes

$$\rho_{\mathcal{I}} = \int P_{\mathcal{O}}(\boldsymbol{\alpha})\,|\boldsymbol{\beta}\rangle_\psi\,\langle\boldsymbol{\beta}|_\psi \otimes |\boldsymbol{\gamma}\rangle_v\,\langle\boldsymbol{\gamma}|_v\,\mathrm{D}\boldsymbol{\alpha}. \tag{D.20}$$

The final result, Eq. (D.14), is obtained by tracing out the environment modes. □

It is important to see that Eq. (D.14) **is not a valid (multimode) P representation**, despite the striking similarity. Although the rescaling in Eq. (D.16) is a valid transformation between quasi-probability distributions, the multimode state $|\boldsymbol{\alpha}\rangle_\psi$ is not a tensor product of the single mode coherent states $|\alpha\rangle_{\psi_j} \equiv e^{\alpha_j\hat{\psi}_j^\dagger - \alpha_j^*\hat{\psi}_j}\,|\mathrm{vac}\rangle$ (because the latter do not "live" in different Hilbert spaces, due to the non-orthogonality of the PSF operators). Therefore, special attention must be paid when dealing with Eq. (D.14), as the usual machinery of P representations might not be valid.

Proposition 2. *The states* $|\boldsymbol{\alpha}\rangle_\psi$ *can be expanded as*

$$|\boldsymbol{\alpha}\rangle_\psi = \bigotimes_{\vec{r}\in\mathbb{R}^2} |E^{\boldsymbol{\alpha}}(\vec{r})\rangle_{\vec{r}}\,, \tag{D.21}$$

where $|\alpha\rangle_{\vec{r}} = \mathcal{D}_{\hat{a}_{\vec{r}}}(\alpha)\,|\text{vac}\rangle$ *is a coherent state of the* $\hat{a}_{\vec{r}}$ *operator (*$\hat{a}_{\vec{r}}\,|\alpha\rangle_{\vec{r}} = \alpha\,|\alpha\rangle_{\vec{r}}$*) and* $E^{\boldsymbol{\alpha}}(\vec{r})$ *is the classical field amplitude at the image plane position* $\vec{r}$*,*

$$E^{\boldsymbol{\alpha}}(\vec{r}) = \sum_{j=1}^{S} \alpha_j \psi(\vec{r} - \vec{r}_j). \tag{D.22}$$

Proof. Starting with the definition of $|\boldsymbol{\alpha}\rangle_\psi$, Eq. (D.15), successive applications of Eq. (D.7) lead to

$$\begin{aligned} |\boldsymbol{\alpha}\rangle_\psi &= \prod_{j=1}^{S} e^{\alpha_j \hat{\psi}_j^\dagger - \alpha_j^* \hat{\psi}_j}\,|\text{vac}\rangle \\ &= e^{i\Phi} e^{\sum_j \alpha_j \hat{\psi}_j^\dagger - \alpha_j^* \hat{\psi}_j}\,|\text{vac}\rangle\,, \end{aligned} \tag{D.23}$$

where Φ is an accumulated global phase, which can be ignored. Further, expanding the PSF operators leads to

$$|\boldsymbol{\alpha}\rangle_\psi = e^{\int \mathrm{d}\vec{r}\left(E^{\boldsymbol{\alpha}}(\vec{r})\hat{a}_{\vec{r}}^\dagger - [E^{\boldsymbol{\alpha}}(\vec{r})]^* \hat{a}_{\vec{r}}\right)}\,|\text{vac}\rangle\,. \tag{D.24}$$

Since the terms in the integrand commute for different $\vec{r}$,

$$\left[E^{\boldsymbol{\alpha}}(\vec{r})\hat{a}_{\vec{r}}^\dagger - [E^{\boldsymbol{\alpha}}(\vec{r})]^*\hat{a}_{\vec{r}}, E^{\boldsymbol{\alpha}}(\vec{r'})\hat{a}_{\vec{r'}}^\dagger - [E^{\boldsymbol{\alpha}}(\vec{r'})]^*\hat{a}_{\vec{r'}}\right] = 0, \tag{D.25}$$

we have that

$$\begin{aligned} |\boldsymbol{\alpha}\rangle_\psi &= \prod_{\vec{r}\in\mathbb{R}^2} e^{E^{\boldsymbol{\alpha}}(\vec{r})\hat{a}_{\vec{r}}^\dagger - [E^{\boldsymbol{\alpha}}(\vec{r})]^*\hat{a}_{\vec{r}}}\,|\text{vac}\rangle \\ &= \prod_{\vec{r}\in\mathbb{R}^2} \mathcal{D}_{\hat{a}_{\vec{r}}}(E^{\boldsymbol{\alpha}}(\vec{r}))\,|\text{vac}\rangle \equiv \bigotimes_{\vec{r}\in\mathbb{R}^2} |E^{\boldsymbol{\alpha}}(\vec{r})\rangle_{\vec{r}}\,. \end{aligned} \tag{D.26}$$

□

Eq. (D.21) allows an easier manipulation of $|\boldsymbol{\alpha}\rangle_\psi$, as the coherent states $|\alpha\rangle_{\vec{r}}$ lie in different Hilbert spaces for different $\vec{r}$. With it we can easily calculate the average photon number at the image plane,

$$\epsilon = \int \mathrm{d}\vec{r}\,\langle \hat{a}_{\vec{r}}^\dagger \hat{a}_{\vec{r}}\rangle = \int \mathrm{d}\vec{r}\left\langle |E^{\boldsymbol{\alpha}}(\vec{r})|^2 \right\rangle, \tag{D.27}$$

where we used that $\hat{a}_{\vec{r}}\,|\boldsymbol{\alpha}\rangle_\psi = E^{\boldsymbol{\alpha}}(\vec{r})\,|\boldsymbol{\alpha}\rangle_\psi$.

D.3 Quantum measurements

Direct imaging

Consider a photon number-resolved detector at position $\vec{r}$ in the image plane. Assuming it covers an infinitesimal area around its centre, it will detect n photons with probability

$$\Pr^{(\vec{r})}(n) = \mathrm{Tr}\left(\rho \left|n\right\rangle_{\vec{r}} \left\langle n\right|_{\vec{r}}\right), \tag{D.28}$$

where

$$\left|n\right\rangle_{\vec{r}} = \frac{\left(\hat{a}^{\dagger}_{\vec{r}}\right)^n}{\sqrt{n!}} \left|\mathrm{vac}\right\rangle \tag{D.29}$$

is the photon number state at position $\vec{r}$. Using the expansion of Eq. (D.21) and the overlap (See Appendix C)

$$\left\langle n\right|_{\vec{r}} \left|\alpha\right\rangle_{\vec{r'}} = \delta^2(\vec{r} - \vec{r'}) \frac{e^{-|\alpha|^2/2}\alpha^n}{\sqrt{n!}}, \tag{D.30}$$

we can express the probabilities as

$$\begin{aligned} \Pr^{(\vec{r})}(n) &= \int P_{\mathcal{I}}(\boldsymbol{\alpha}) \frac{e^{-|E^{\boldsymbol{\alpha}}(\vec{r})|^2} |E^{\boldsymbol{\alpha}}(\vec{r})|^{2n}}{n!} \mathrm{D}\boldsymbol{\alpha} \\ &= \left\langle \frac{e^{-|E^{\boldsymbol{\alpha}}(\vec{r})|^2} |E^{\boldsymbol{\alpha}}(\vec{r})|^{2n}}{n!} \right\rangle. \end{aligned} \tag{D.31}$$

We call *direct intensity imaging* a measurement of photon numbers at every position $\vec{r}$ of the image plane.

Eq. (D.31) can be simplified if the object sources are uncorrelated, i.e.,

$$P_{\mathcal{O}}(\boldsymbol{\alpha}) = \prod_{j=1}^{S} P_j(\alpha_j). \tag{D.32}$$

In this case, a change of variables can turn Eq. (D.31) into a single (complex plane) integral. In order to see that, consider first the case where the PSF is nonzero everywhere in the image plane, which happens, for example, with the Gaussian PSF. Then, $\psi(\vec{r} - \vec{r}_j) \neq 0$ for every source position $\vec{r}_j$ and we can perform the change of variables $\beta = E(\vec{r})$ and $\beta_j = \alpha_j \psi(\vec{r} - \vec{r}_j)$ for $j = 2, \ldots, S$, yielding

$$\Pr^{(\vec{r})}(n) = \int P_{\vec{r}}(\beta) \frac{e^{-|\beta|^2} |\beta|^{2n}}{n!} \mathrm{d}^2\beta, \tag{D.33}$$

where

$$P_{\vec{r}}(\beta) = \int \frac{1}{\eta|\psi(\vec{r}-\vec{r}_1)|} P_1\left(\frac{\beta - \sum_{j=2}^{S}\beta_j}{\sqrt{\eta}\psi(\vec{r}-\vec{r}_1)}\right) \times \left[\prod_{j=1}^{S}\frac{1}{\eta|\psi(\vec{r}-\vec{r}_j)|} P_j\left(\frac{\beta_j}{\sqrt{\eta}\psi(\vec{r}-\vec{r}_j)}\right)\right] \mathrm{d}^2\beta_2 \ldots \mathrm{d}^2\beta_S$$
$$= *_{j=1}^{S}\left[\frac{1}{\eta|\psi(\vec{r}-\vec{r}_j)|^2} P_j\left(\frac{\beta}{\sqrt{\eta}\psi(\vec{r}-\vec{r}_j)}\right)\right], \tag{D.34}$$

where $*$ denotes multiple convolution operations. Although this looks overcomplicated, in some situations it is actually simpler than carrying out a full-fledged calculation.

For the general case, if $\psi(\vec{r}-\vec{r}_{j'}) = 0$ for some j', then we perform the same change of variables, except for $\beta_{j'} = \alpha_{j'}$. All the terms related to j' will then integrate out, and we reach again Eq. (D.34) but with the j'-th term missing from the multiple convolutions. The influence of the PSF zeros on direct imaging is studied in Ref. [96].

Mode sorting

An alternative to direct imaging is to perform a *spatial mode sorting* or *spatial demultiplexing* of the image plane field prior to photodetection. We start by choosing an orthonormal basis $\{\phi_m(\vec{r})\}$ of functions in the image plane ($\mathbb{R}^2$). Orthonormality means that

$$\int \mathrm{d}\vec{r}\phi_m^*(\vec{r})\ \phi_{m'}^*(\vec{r}) = \delta_{mm'}, \tag{D.35}$$

and as a basis it must be complete [146],

$$\sum_{m=0}^{\infty}\phi_m(\vec{r})\phi_m^*(\vec{r'}) = \delta^2(\vec{r}-\vec{r'}). \tag{D.36}$$

We then define the *spatial mode operators*

$$\hat{\phi}_m^\dagger = \int \mathrm{d}\vec{r}\ \phi_m(\vec{r})\hat{a}_{\vec{r}}^\dagger. \tag{D.37}$$

Eq. (D.35) implies that the modes are also orthonormal, $[\hat{\phi}_m^\dagger, \hat{\phi}_m'] = \delta_{mm'}$, and Eq. (D.36) implies that any other operator in this space can be expanded in the $\{\hat{\phi}_m^\dagger\}$ basis. In particular,

$$\hat{\psi}_j^\dagger = \sum_{m=0}^{\infty} R_m(\vec{r}_j)\hat{\phi}_m^\dagger, \tag{D.38}$$

where

$$R_m(\vec{r}_j) = \int \mathrm{d}\vec{r}\; \phi_m^*(\vec{r})\psi(\vec{r}-\vec{r}_j) \tag{D.39}$$

is the overlap between the source image $\psi(\vec{r}-\vec{r}_j)$ and the spatial mode $\phi_m(\vec{r})$.

Mode sorting consists of decomposing the image plane light into a basis of modes and subsequently performing photodetection at each resulting component. Mathematically, it consists of a photon number measurement at each mode component,

$$\Pr^{(m)}(n) = \mathrm{Tr}\left(\rho \left|n\right\rangle_m \left\langle n\right|_m\right), \tag{D.40}$$

where

$$\left|n\right\rangle_m = \frac{\left(\hat{\phi}_m^\dagger\right)^n}{\sqrt{n!}} \left|\mathrm{vac}\right\rangle. \tag{D.41}$$

Experimental implementations of mode sorting are discussed in Chapter 6.

Akin to Eq. (D.21), the multimode state $\left|\boldsymbol{\alpha}\right\rangle_\psi$ can be decomposed in the spatial mode basis.

Proposition 3. *The state $\left|\boldsymbol{\alpha}\right\rangle_\psi$ can be expanded as*

$$\left|\boldsymbol{\alpha}\right\rangle_\psi = \bigotimes_{m=0}^{\infty} \left|E_m^{\boldsymbol{\alpha}}\right\rangle_m, \tag{D.42}$$

where $\left|\alpha\right\rangle_m = \mathcal{D}_{\hat{\phi}_m}(\alpha)\left|\mathrm{vac}\right\rangle$ is the coherent state of the spatial mode operator $\hat{\phi}_m$ and $E_m^{\boldsymbol{\alpha}}$ is the component of the image plane field at that mode,

$$E_m^{\boldsymbol{\alpha}} = \sum_{j=1}^{S} R_m(\vec{r}_j)\alpha_j = \int d\vec{r}\phi_m^*(\vec{r})E^{\boldsymbol{\alpha}}(\vec{r}) \tag{D.43}$$

Proof. Analogously to the proof of Proposition 2, we have that

$$\begin{aligned}\left|\boldsymbol{\alpha}\right\rangle_\psi &= e^{\sum_j \sum_m \left(R_m(\vec{r}_j)\alpha_j\phi_m^\dagger - R_m^*(\vec{r}_j)\alpha_j^*\phi_m\right)}\left|\mathrm{vac}\right\rangle \\ &= e^{\sum_m \left[E_m^{\boldsymbol{\alpha}}\phi_m^\dagger - [E_m^{\boldsymbol{\alpha}}]^*\phi_m\right]}\left|\mathrm{vac}\right\rangle.\end{aligned} \tag{D.44}$$

Since

$$\left[E_m^{\boldsymbol{\alpha}}\phi_m^\dagger - [E_m^{\boldsymbol{\alpha}}]^*\phi_m, E_{m'}^{\boldsymbol{\alpha}}\phi_{m'}^\dagger - [E_{m'}^{\boldsymbol{\alpha}}]^*\phi_{m'}\right] = 0, \tag{D.45}$$

we have

$$\left|\boldsymbol{\alpha}\right\rangle_\psi = \prod_{m=0}^{\infty} e^{E_m^{\boldsymbol{\alpha}}\phi_m^\dagger - [E_m^{\boldsymbol{\alpha}}]^*\phi_m}\left|\mathrm{vac}\right\rangle = \bigotimes_{m=0}^{\infty}\left|E_m^{\boldsymbol{\alpha}}\right\rangle_m. \tag{D.46}$$

□

The overlap Eq. (C.8) leads to the expression

$$\Pr^{(m)}(n) = \int P_{\mathcal{I}}(\boldsymbol{\alpha}) \frac{e^{-|E_m^{\boldsymbol{\alpha}}|^2} |E_m^{\boldsymbol{\alpha}}|^{2n}}{n!} \mathrm{D}\boldsymbol{\alpha}. \tag{D.47}$$

For uncorrelated sources, in a procedure similar to the one in the previous section, the probabilities reduce to

$$\Pr^{(m)}(n) = \int P_m(\alpha) \frac{e^{-|\alpha|^2} |\alpha|^{2n}}{n!} \mathrm{d}^2\alpha, \tag{D.48}$$

where

$$P_m(\alpha) = *_{j=1}^{S} \left[\frac{1}{\eta |R_m(\vec{r}_j)|^2} P_j \left(\frac{\alpha}{\sqrt{\eta} R_m(\vec{r}_j)} \right) \right]. \tag{D.49}$$

Note that a spatially-resolved measurement can be seen as a particular case of mode sorting in the basis $\psi_{\vec{r'}}(\vec{r}) = \psi(\vec{r} - \vec{r'})$, where the mode index $(\vec{r'})$ is now continuous. For this reason, in some instances we might use a unified notation for both spatial detection and mode sorting, where the index m will be used for both the discrete and continuous cases.

Mode sorting is usually performed in the laboratory with spatial light modulators (SLM). Mode sorting technology, however, is still in its infancy. Current schemes [193–197] either sort very few modes and/or work with low accuracy.

Quadrature measurements

In the aforementioned schemes, after decomposing the light field (either in the position basis or some mode basis) we performed photon-number measurements, i.e., intensity detection, on each component. However, we don't need to restrict ourselves to this. A common alternative to direct photodetection is to measure the quadratures of light, i.e., the real and imaginary parts of its amplitude, as in

$$\alpha = \frac{1}{\sqrt{2}}(q + ip). \tag{D.50}$$

They are eigenvalues of the quadrature operators

$$\hat{q} = \frac{\hat{a} + \hat{a}^\dagger}{\sqrt{2}}, \ \hat{p} = \frac{\hat{a} - \hat{a}^\dagger}{\sqrt{2}i}. \tag{D.51}$$

The overlap between the eigenstates $|q\rangle$ of the q quadrature operator and coherent states is

$$\langle q|\alpha\rangle = \frac{1}{\pi^{1/4}} e^{-\left(q-\sqrt{2}\,\mathrm{Re}\{\alpha\}\right)^2/2} e^{iq\sqrt{2}\,\mathrm{Im}\{\alpha\}}, \tag{D.52}$$

with analogous results for the p quadrature. Using this overlap, we can see that a single quadrature measurement leads to the probabilities

$$\Pr^{(m)}(q) = \int P_{\mathcal{I}}(\boldsymbol{\alpha}) \frac{1}{\sqrt{\pi}} e^{-\left(q-\sqrt{2}\,\mathrm{Re}\{E_m^{\boldsymbol{\alpha}}\}\right)^2} \mathrm{D}\boldsymbol{\alpha}. \tag{D.53}$$

A full quadrature measurement, which consists of performing two independent single quadrature measurements, leads to the joint probability distribution

$$\Pr^{(m)}(q,p) = \frac{1}{2} \int P_{\mathcal{I}}(\boldsymbol{\alpha}) \frac{1}{\pi} e^{-\left(q-\sqrt{2}\,\mathrm{Re}\{E_m^{\boldsymbol{\alpha}}\}\right)^2/2} e^{-\left(p-\sqrt{2}\,\mathrm{Im}\{E_m^{\boldsymbol{\alpha}}\}\right)^2/2}\, \mathrm{D}\boldsymbol{\alpha}. \tag{D.54}$$

Note the 1/2 term is a consequence of the non-commutativity of the measurements. For uncorrelated sources, these probabilities reduce to

$$\Pr^{(m)}(q) = \frac{1}{\sqrt{\pi}} \int P_m(\alpha)\; e^{-\left(q-\sqrt{2}\,\mathrm{Re}\{\alpha\}\right)^2} \mathrm{d}^2\alpha \tag{D.55}$$

and

$$\Pr^{(m)}(q,p) = \frac{1}{2\pi} \int P_m(\alpha)\; e^{-\left(q-\sqrt{2}\,\mathrm{Re}\{\alpha\}\right)^2/2} e^{-\left(p-\sqrt{2}\,\mathrm{Im}\{\alpha\}\right)^2/2}\; \mathrm{d}^2\alpha, \tag{D.56}$$

respectively.

The most common way to experimentally perform single and full quadrature measurements is via homodyne and heterodyne detection, respectively [25, 147].

D.4 Incoherent light

A single mode of thermal light has the P representation

$$P_\Gamma(\alpha) = \frac{1}{\pi\Gamma} e^{-|\alpha|^2/\Gamma}, \tag{D.57}$$

where

$$\Gamma = \frac{1}{e^{\hbar\omega/k_B T} - 1} \tag{D.58}$$

is the light strength, and $\hbar$, ω, k_B and T represent the Planck constant, the light angular frequency, the Boltzmann constant and the state's temperature, respectively [189]. For a general multimode thermal state we have

$$P_\Gamma(\boldsymbol{\alpha}) = \frac{1}{\pi|\Gamma|} e^{-\boldsymbol{\alpha}^\dagger \Gamma^{-1} \boldsymbol{\alpha}}, \tag{D.59}$$

where $|\cdot|$ denotes the determinant and $\Gamma = \langle \boldsymbol{\alpha}\boldsymbol{\alpha}^\dagger \rangle$ is the covariance matrix of the thermal state. Furthermore, its average photon number is $\epsilon = \operatorname{Tr}\Gamma$.

In this thesis we only work with thermal sources which are uncorrelated at the object plane, i.e., $P_\mathcal{O}(\boldsymbol{\alpha}) = P_{\Gamma_0}(\boldsymbol{\alpha})$ where Γ_0 is diagonal. This means that the average photon number is simply the sum of individual sources' strengths $(\Gamma_0)_j \equiv (\Gamma_0)_{jj}$,

$$\epsilon_0 = \sum_{j=1}^{S} (\Gamma_0)_j. \tag{D.60}$$

The corresponding quasi-distribution (which is no longer part of a P representation) at the image plane is

$$P_\mathcal{I}(\boldsymbol{\alpha}) = \frac{1}{\eta^S} P_\mathcal{O}\left(\frac{\boldsymbol{\alpha}}{\sqrt{\eta}}\right) = P_\Gamma(\boldsymbol{\alpha}), \tag{D.61}$$

where $\Gamma = \eta\Gamma_0$. Likewise, the average photon number is $\epsilon = \eta\epsilon_0$. Using Eqs. (D.48) and (D.49), the probabilities for intensity measurements after mode sorting are

$$\Pr^{(m)}(n) = \frac{1}{\Gamma_m + 1}\left(\frac{\Gamma_m}{\Gamma_m + 1}\right)^n, \tag{D.62}$$

where

$$\Gamma_m = \sum_{j=1}^{S} \Gamma_j |R_m(\vec{r}_j)|^2 \tag{D.63}$$

is the average photon number of each mode component. In particular, for direct imaging we have

$$\Gamma_{\vec{r}} = \sum_{j=1}^{S} \Gamma_j |\psi(\vec{r} - \vec{r}_j)|^2. \tag{D.64}$$

For single and full quadrature measurements the probabilities are, respectively,

$$\Pr^{(m)}(q) = \frac{1}{\sqrt{\pi(2\Gamma_m + 1)}} e^{-q^2/(2\Gamma_m+1)} \tag{D.65}$$

and

$$\Pr^{(m)}(q,p) = \frac{1}{2\pi(\Gamma_m + 1)} e^{-(q^2+p^2)/2(\Gamma_m+1)}. \tag{D.66}$$

D.5 Coherent light

A single mode of coherent light has the P representation

$$P_A(\alpha) = \delta^2(\alpha - A), \tag{D.67}$$

where A is the complex amplitude of the coherent state. A multimode uncorrelated state in the object plane is given by

$$P_{\mathcal{O}}(\boldsymbol{\alpha}) = P_{\boldsymbol{A_0}}(\boldsymbol{\alpha}) \equiv \prod_{j=1}^{S} \delta^2(\alpha_j - (\boldsymbol{A_0})_j), \tag{D.68}$$

where $(\boldsymbol{A_0})$ is the vector of amplitudes of each source. At the image plane it becomes $P_{\mathcal{I}}(\boldsymbol{\alpha}) = P_{\boldsymbol{A}}(\boldsymbol{\alpha})$, where $\boldsymbol{A} = \sqrt{\eta}\boldsymbol{A_0}$, so the quantum state is simply

$$\rho = |\boldsymbol{A}\rangle_\psi \langle \boldsymbol{A}|_\psi . \tag{D.69}$$

Intensity, single and full quadrature measurement probabilities can be straightforwardly found:

$$\Pr^{(m)}(n) = \frac{e^{-|E_m|^2}|E_m|^{2n}}{n!}, \tag{D.70}$$

$$\Pr^{(m)}(q) = \frac{1}{\sqrt{\pi}} e^{-\left(q - \sqrt{2}\,\mathrm{Re}\{E_m\}\right)^2}, \tag{D.71}$$

$$\Pr^{(m)}(q,p) = \frac{1}{2\pi} e^{-\left(q - \sqrt{2}\,\mathrm{Re}\{E_m\}\right)^2/2} e^{-\left(p - \sqrt{2}\,\mathrm{Im}\{E_m\}\right)^2/2}, \tag{D.72}$$

where

$$E_m = \sum_{j=1}^{S} A_j R_m(\vec{r}_j) = \int \mathrm{d}\vec{r}\; \phi_m^*(\vec{r}) E^{\boldsymbol{A}}(\vec{r}) \tag{D.73}$$

is the amplitude of the component field. For direct imaging, it reduces to

$$A_{\vec{r}} = E^{\boldsymbol{A}}(\vec{r}) * \psi(\vec{r}). \tag{D.74}$$

Not as simple is the behaviour of the average photon number. While for the object plane it is $\epsilon_0 = |\boldsymbol{A_0}|^2$, i.e., the sum of individual sources' intensities, at the image plane, due to interference effects, it becomes

$$\epsilon = \int \mathrm{d}\vec{r} \left| \sum_{j=1}^{S} A_j \psi(\vec{r} - \vec{r}_j) \right|^2 = \sum_{i,j=1}^{S} A_i^* A_j \Delta(\vec{r}_i - \vec{r}_j). \tag{D.75}$$

The dependence of ϵ on the object parameters (intensities, relative phases and positions) can lead to unexpected results and confusion, as evidenced by the small controversy it created in the literature [27, 32, 41, 45, 53, 69]. Therefore, special attention on this subject is required.

E | Weak thermal light

A general multimode thermal state is represented by the P function [189]

$$P_\Gamma(\boldsymbol{\alpha}) = \frac{1}{\pi|\Gamma|} e^{-\boldsymbol{\alpha}^\dagger \Gamma^{-1} \boldsymbol{\alpha}}, \tag{E.1}$$

where $\Gamma = \langle \boldsymbol{\alpha}\boldsymbol{\alpha}^\dagger \rangle$ is the covariance matrix of the thermal state. We assume the light at the object plane consists of uncorrelated thermal sources,

$$P_{\mathcal{O}}(\boldsymbol{\alpha}) = P_{\Gamma_0}(\boldsymbol{\alpha}), \tag{E.2}$$

where Γ_0 is a diagonal matrix. At the image plane, the quantum state of light is

$$\rho = \int P_{\mathcal{I}}(\boldsymbol{\alpha}) \left|\boldsymbol{\alpha}\right\rangle_\psi \left\langle\boldsymbol{\alpha}\right|_\psi \mathrm{D}\boldsymbol{\alpha}, \tag{E.3}$$

where

$$P_{\mathcal{I}}(\boldsymbol{\alpha}) = \frac{1}{\eta^S} P_{\mathcal{O}}\left(\frac{\boldsymbol{\alpha}}{\sqrt{\eta}}\right) = P_\Gamma(\boldsymbol{\alpha}) \tag{E.4}$$

and $\Gamma = \eta\Gamma_0$.

Proposition 4. *The thermal state in Eq.* (E.3) *is photon-number diagonal, i.e.,*

$$\rho = \sum_{n=0}^{\infty} \pi_n \rho_n, \tag{E.5}$$

where ρ_n is a n-photon multimode Fock state and π_n is the corresponding probability [6].

Proof. Consider the Fock states of the position modes,

$$\left|\{n\}\right\rangle = \left|n(\vec{r}_1)\right\rangle_{\vec{r}_1} \otimes \cdots \otimes \left|n(\vec{r}_i)\right\rangle_{\vec{r}_i} \otimes \cdots, \tag{E.6}$$

such that $\sum_j n(\vec{r}_j) = n$. Then, the matrix elements of the state in Eq. (E.3) are

$$\begin{aligned}\langle\{n\}|\,\rho\,|\{m\}\rangle &= \int P_{\mathcal{I}}(\boldsymbol{\alpha})\mathrm{D}\boldsymbol{\alpha}\prod_{\vec{r}}\langle n(\vec{r})|E^{\boldsymbol{\alpha}}(\vec{r})\rangle\,\langle E^{\boldsymbol{\alpha}}(\vec{r})|m(\vec{r})\rangle \\ &= \frac{1}{\sqrt{\prod_{\vec{r}} n(\vec{r})!m(\vec{r})!}}\int P_{\mathcal{I}}(\boldsymbol{\alpha})\mathrm{D}\boldsymbol{\alpha}\prod_{\vec{r}} e^{-|E^{\boldsymbol{\alpha}}(\vec{r})|^2}E^{\boldsymbol{\alpha}}(\vec{r})^{n(\vec{r})}\left([E^{\boldsymbol{\alpha}}]^*(\vec{r})\right)^{m(\vec{r})}.\end{aligned} \tag{E.7}$$

Now, note that

$$P_{\mathcal{I}}(\boldsymbol{\alpha})\prod_{\vec{r}} e^{-|E^{\boldsymbol{\alpha}}(\vec{r})|^2} \tag{E.8}$$

is simply another Gaussian state. Therefore,

$$\langle\{n\}|\,\rho\,|\{m\}\rangle = \left\langle\prod_{\vec{r}} E^{\boldsymbol{\alpha}}(\vec{r})^{n(\vec{r})}\left([E^{\boldsymbol{\alpha}}]^*(\vec{r})\right)^{m(\vec{r})}\right\rangle. \tag{E.9}$$

By the Gaussian moment theorem [6, 189], $\langle\{n\}|\,\rho\,|\{m\}\rangle = 0$ if $n \neq m$, i.e., if $\sum_j n(\vec{r}_j) \neq \sum_j m(\vec{r}_j)$. □

At optical frequencies, the average photon number of most thermal sources is very small [112, 142],

$$\epsilon \ll 1. \tag{E.10}$$

For example, a typical fluorophore emits less than 10^7 photons per coherence time (< 50 fs), which leads to $\epsilon < 10^6$ for a single source [6]. In line with the literature on QISR [6], light that obeys Eq. (E.10) is called *weak incoherent light*.

Proposition 5. *The quantum state of weak incoherent light can be approximated as*

$$\rho \approx (1-\epsilon)\rho_0 + \epsilon\rho_1. \tag{E.11}$$

The one photon state ρ_1 is given by

$$\rho_1 = \sum_{i=1}^{S} p_i\,|\psi_i\rangle\,\langle\psi_i|\,, \tag{E.12}$$

where

$$|\psi_i\rangle = \hat{\psi}_i^\dagger\,|\mathrm{vac}\rangle = \int \mathrm{d}\vec{r}\psi(\vec{r}-\vec{r}_i)\,|\vec{r}\rangle \tag{E.13}$$

are the one photon PSF states, $|\vec{r}\rangle \equiv \hat{a}_{\vec{r}}^\dagger|\mathrm{vac}\rangle$ are the one photon position states at the image plane and

$$p_i = \frac{\Gamma_i}{\sum_k \Gamma_k} \tag{E.14}$$

are the relative intensities of each source [6].

Proof. Recall that thermal states are photon-number diagonal, as per Eq. (E.5). The probability π_0 of the state being in the vacuum state ρ_0 is simply

$$\pi_0 = \mathrm{Tr}\left[\rho \left|\mathrm{vac}\right\rangle \left\langle\mathrm{vac}\right|\right] = \left\langle \prod_{\vec{r}} e^{-|E^{\boldsymbol{\alpha}}(\vec{r})|^2} \right\rangle = \left\langle e^{-\int \mathrm{d}\vec{r}\ |E^{\boldsymbol{\alpha}}(\vec{r})|^2} \right\rangle. \tag{E.15}$$

Taylor expanding the exponential, and noting that $\int \mathrm{d}\vec{r} \langle |E^{\boldsymbol{\alpha}}(\vec{r})|^2 \rangle = \epsilon$, leads to

$$\pi_0 = 1 - \epsilon + \mathcal{O}(\epsilon^2). \tag{E.16}$$

Likewise, the one-photon probability is

$$\pi_1 = \left\langle e^{-\int \mathrm{d}\vec{r}\ |E^{\boldsymbol{\alpha}}(\vec{r})|^2} |E^{\boldsymbol{\alpha}}(\vec{r})|^2 \right\rangle \approx \epsilon + \mathcal{O}(\epsilon^2), \tag{E.17}$$

and the multiphoton probability is

$$\sum_{n=2}^{\infty} \pi_n = 1 - \pi_0 - \pi_1 \approx \mathcal{O}(\epsilon^2). \tag{E.18}$$

Therefore, for weak thermal light ($\epsilon \ll 1$) the multiphoton events can be neglected and the state can be well approximated by

$$\rho \approx (1-\epsilon)\rho_0 + \epsilon\rho_1. \tag{E.19}$$

In order to find the form of the one-photon state ρ_1, let's start by expanding the whole state ρ in the one-photon position basis $\{|\vec{r}\rangle\}$:

$$\left\langle \vec{r}\right| \rho \left|\vec{r}'\right\rangle = \int P_{\mathcal{I}}(\boldsymbol{\alpha}) \mathrm{D}\boldsymbol{\alpha} \left\langle \vec{r}|\boldsymbol{\alpha}\right\rangle_{\psi} \left\langle \boldsymbol{\alpha}\right|_{\psi} \left|\vec{r}'\right\rangle. \tag{E.20}$$

Noting that

$$\left\langle \vec{r}|\alpha\right\rangle_{\vec{r}} = \left\langle \mathrm{vac}\right| \hat{a}_{\vec{r}} \left|\alpha\right\rangle_{\vec{r}} = \alpha \left\langle \mathrm{vac}|\alpha\right\rangle = \alpha e^{-|\alpha|^2/2}, \tag{E.21}$$

we can use the decomposition

$$\left|\boldsymbol{\alpha}\right\rangle_{\psi} = \bigotimes_{\vec{r}\in\mathbb{R}^2} \left|E^{\boldsymbol{\alpha}}(\vec{r})\right\rangle_{\vec{r}} \tag{E.22}$$

to find

$$\left\langle \vec{r}|\boldsymbol{\alpha}\right\rangle_{\psi} = e^{-\int \mathrm{d}\vec{s}\ |E^{\boldsymbol{\alpha}}(\vec{s})|^2} E^{\boldsymbol{\alpha}}(\vec{r}). \tag{E.23}$$

Then, replacing Eq. (E.23) into Eq. (E.20) leads to

$$\left\langle \vec{r}\right| \rho \left|\vec{r}'\right\rangle = \left\langle e^{-\int \mathrm{d}\vec{s}\ |E^{\boldsymbol{\alpha}}(\vec{s})|^2} E^{\boldsymbol{\alpha}}(\vec{r}) [E^{\boldsymbol{\alpha}}]^*(\vec{r}') \right\rangle = \langle E^{\boldsymbol{\alpha}}(\vec{r}) [E^{\boldsymbol{\alpha}}]^*(\vec{r}') \rangle + \mathcal{O}(\epsilon^2). \tag{E.24}$$

The electric field correlation function can be simplified to

$$\begin{aligned}\langle E^{\boldsymbol{\alpha}}(\vec{r})[E^{\boldsymbol{\alpha}}]^*(\vec{r}')\rangle &= \sum_{ij}\langle\alpha_i\alpha_j^*\rangle\psi(\vec{r}-\vec{r}_i)\psi^*(\vec{r}'-\vec{r}_j)\\ &= \sum_i \Gamma_i\psi(\vec{r}-\vec{r}_i)\psi^*(\vec{r}'-\vec{r}_i).\end{aligned} \tag{E.25}$$

Finally, we can use Eq. (E.20) and Eq. (E.11) to find the matrix elements of the one-photon state:

$$\langle\vec{r}|\,\rho\,|\vec{r}'\rangle \approx (1-\epsilon)\,\langle\vec{r}|\,\rho_0\,|\vec{r}'\rangle + \epsilon\,\langle\vec{r}|\,\rho_1\,|\vec{r}'\rangle\,. \tag{E.26}$$

Since $\langle\vec{r}|\,\rho_0\,|\vec{r}'\rangle = 0$, the one-photon state is

$$\rho_1 = \iint \mathrm{d}\vec{r}\mathrm{d}\vec{r}' \sum_i \left(\frac{\Gamma_i}{\epsilon}\right)\psi(\vec{r}-\vec{r}_i)\psi^*(\vec{r}'-\vec{r}_i)\,|\vec{r}\rangle\,\langle\vec{r}'| = \sum_{i=1}^{S} p_i\,|\psi_i\rangle\,\langle\psi_i|\,. \tag{E.27}$$

□

Eq. (E.11) means that multiphoton events can be ignored for weak incoherent light. Furthermore, since the vacuum state ρ_0 carries no information, we can safely ignore it as well and focus on the one-photon state ρ_1.

From Eq. (E.11), direct imaging probabilities are simply

$$\Pr(\vec{r}) = \sum_{i=1}^{S} p_i|\,\langle\vec{r}|\psi_i\rangle\,|^2 = \sum_{i=1}^{S} p_i|\psi(\vec{r}-\vec{r}_i)|^2, \tag{E.28}$$

and mode sorting intensity measurements are determined by

$$\Pr(m) = \sum_{i=1}^{S} p_i|\,\langle\phi_m|\psi_i\rangle\,|^2 = \sum_{i=1}^{S} p_i|R_m(\vec{r}_i)|^2, \tag{E.29}$$

where $|\phi_m\rangle \equiv \hat{\phi}_m^\dagger\,|\mathrm{vac}\rangle$.

F | Gaussian aperture properties

An apodized Airy point-spread function can be approximated by a Gaussian function

$$\psi(\vec{r}) = \frac{e^{-|\vec{r}|^2/4\sigma^2}}{(2\pi\sigma^2)^{1/2}}, \tag{F.1}$$

where $\sigma = 0.42\lambda/2\mathrm{NA}$ [143]. The corresponding aperture is another Gaussian, with waist $\omega = 1/4\pi\sigma$.

One-dimensional case

We start by studying its one dimensional version,

$$\psi(x) = \frac{e^{-x^2/4\sigma^2}}{(2\pi\sigma^2)^{1/4}}, \tag{F.2}$$

from which the 2D version follows simply as $\psi(\vec{r}) = \psi(x)\psi(y)$. The $\psi(x)$ is L^2 normalised,

$$\int_{-\infty}^{+\infty} \mathrm{d}x \; |\psi(x)|^2 = 1, \tag{F.3}$$

and inversion-symmetric, $\psi(-x) = \psi(x)$. The overlap between displaced PSFs is defined as

$$\begin{aligned} \Delta(\alpha - \beta) &\equiv \int_{-\infty}^{+\infty} \mathrm{d}x \; \psi^*(x - \alpha)\psi(x - \beta) \\ &= \int_{-\infty}^{+\infty} \mathrm{d}x \; \psi^*(x - (\alpha - \beta))\psi(x), \end{aligned} \tag{F.4}$$

and it obeys $\Delta(0) = 1$ and $\Delta^*(-x) = \Delta(x)$. For a Gaussian PSF,

$$\Delta(x) = e^{-x^2/8\sigma^2}. \tag{F.5}$$

The overlap between the PSF and its derivative is given by

$$\begin{aligned}\gamma(\alpha-\beta) &\equiv \int_{-\infty}^{+\infty} \mathrm{d}x\ \frac{\partial\psi^*(x-\alpha)}{\partial\alpha}\psi(x-\beta) \\ &= -\int_{-\infty}^{+\infty} \mathrm{d}x\ \psi^*(x-\alpha)\frac{\partial\psi(x-\beta)}{\partial\beta} \\ &= \int_{-\infty}^{+\infty} \mathrm{d}x\ \psi^*(x-(\alpha-\beta))\frac{\partial\psi(x)}{\partial x},\end{aligned} \tag{F.6}$$

and it obeys $\gamma(0)=0$ and $\gamma^*(-x)=-\gamma(x)$. For a Gaussian PSF, its derivative is

$$\frac{\partial\psi(x)}{\partial x} = -\left(\frac{x}{2\sigma^2}\right)\psi(x), \tag{F.7}$$

and the overlap term is

$$\gamma(x) = -\frac{x}{4\sigma^2}e^{-x^2/8\sigma^2}. \tag{F.8}$$

Finally, the overlap between PSF derivatives is

$$\tau(\alpha,\beta) \equiv \int_{-\infty}^{+\infty} \mathrm{d}x\ \frac{\partial\psi^*(x-\alpha)}{\partial\alpha}\frac{\partial\psi(x-\beta)}{\partial\beta}, \tag{F.9}$$

and it obeys $\tau^*(\beta,\alpha)=\tau(\alpha,\beta)$. For the Gaussian case, it depends only on the difference between its arguments, i.e., $\tau(\alpha,\beta)\equiv\tau(\alpha-\beta)$, and it is given by

$$\tau(x) = \frac{1}{4\sigma^2}\left(1-\frac{x^2}{4\sigma^2}\right)e^{-x^2/8\sigma^2}. \tag{F.10}$$

The zero-displacement overlap $\tau(0)$ corresponds to the spatial-frequency variance of the PSF,

$$\tau(0) = 4\pi^2\int_{-\infty}^{+\infty} \mathrm{d}k\ k^2|\mathcal{A}(k)|^2 \equiv 4\pi^2\langle k^2\rangle = \frac{1}{4\sigma^2}. \tag{F.11}$$

Hermite-Gaussian basis

Recall that the PSF-adapted (PAD) basis is the mode sorting basis, for a given aperture, that achieves the qCRB (for the case of estimation separation of two centred sources) with the smallest number of modes. The PAD basis is defined in the aperture plane as

$$\tilde{\phi}_m(u) = i^m g_m(u)\mathcal{A}(u). \tag{F.12}$$

The $\{g_m(u)\}$ is a basis of polynomials which are orthonormal with respect to the weight $|\mathcal{A}(u)|^2$, where

$$\mathcal{A}(u) = \frac{e^{-u^2/4\omega^2}}{(2\pi\omega^2)^{1/4}} \tag{F.13}$$

and $\omega = 1/4\pi\sigma$. Such condition is satisfied by the normalised (physicists') Hermite polynomials

$$g_m(u) = \frac{H_m\left(\frac{u}{\sqrt{2}\omega}\right)}{\sqrt{2^m m!}}, \tag{F.14}$$

which can be easily proved via the orthogonality condition [198]

$$\int_{-\infty}^{+\infty} \mathrm{d}x \; H_m(x) H_n(x) e^{-x^2} = \sqrt{\pi} 2^m m! \delta_{mn}. \tag{F.15}$$

The aperture plane modes are then

$$\tilde{\phi}_m(u) = i^m \frac{H_m\left(\frac{u}{\sqrt{2}\omega}\right)}{\sqrt{2^m m!}} \frac{e^{-u^2/4\omega^2}}{(2\pi\omega^2)^{1/4}}, \tag{F.16}$$

which are Hermite-Gaussian modes waist-matched to the aperture function and with a complex global phase i^m. Since HG modes are eigenmodes of the Fourier transform, via [198]

$$\mathcal{F}\left[e^{-k/2} H_m(k)\right](k) = \sqrt{2\pi}(-i)^m H_m(2\pi x) e^{-2\pi^2 x^2}, \tag{F.17}$$

then the PAD basis is also composed of HG modes,

$$\phi_m(x) = \frac{H_m\left(\frac{x}{\sqrt{2}\sigma}\right)}{\sqrt{2^m m!}} \frac{e^{-x^2/4\sigma^2}}{(2\pi\sigma^2)^{1/4}}, \tag{F.18}$$

which are now waist-matched to the PSF. Hermite-Gaussian modes are orthonormal, complete, and with well-defined parity as

$$\phi_m(-x) = (-1)^m \phi_m(x). \tag{F.19}$$

Their overlap with the PSF is the Weierstrass transform of Hermite polynomials [89],

$$R_m(x) = \int_{-\infty}^{+\infty} \mathrm{d}x' \; \phi_m^*(x') \psi(x' - x) = \frac{1}{\sqrt{m!}} \left(\frac{x}{2\sigma}\right)^m e^{-x^2/8\sigma^2}. \tag{F.20}$$

Two-dimensional case

In two dimensions, all the equivalent properties are similar, due to the separability in x and y of the PSF. For example, the PAD basis is simply

$$\phi_{mn}(x,y) = \phi_m(x)\phi_n(y). \tag{F.21}$$

We denote the overlap terms more generally as

$$\Delta_{ij}^{\mu,\nu} \equiv \int \mathrm{d}\vec{r}\, \frac{\partial\psi^*(\vec{r}-\vec{r}_i)}{\partial\mu}\frac{\partial\psi(\vec{r}-\vec{r}_j)}{\partial\nu}. \tag{F.22}$$

A missing superscript means that no derivation is performed, as in

$$\Delta_{ij}^{\mu,} \equiv \int \mathrm{d}\vec{r}\, \frac{\partial\psi^*(\vec{r}-\vec{r}_i)}{\partial\mu}\psi(\vec{r}-\vec{r}_j). \tag{F.23}$$

The main overlap terms can be expressed in terms of the one dimensional ones:

$$\Delta_{ij} = \Delta(x_i-x_j)\Delta(y_i-y_j), \tag{F.24}$$

$$\Delta_{ij}^{x_i,x_j} = \tau(x_i-x_j)\Delta(y_i-y_j), \tag{F.25}$$

$$\Delta_{ij}^{y_i,y_j} = \Delta(x_i-x_j)\tau(y_i-y_j), \tag{F.26}$$

$$\Delta_{ij}^{x_i,} = \gamma(x_i-x_j)\Delta(y_i-y_j) = -\Delta_{ij}^{,x_j}, \tag{F.27}$$

$$\Delta_{ij}^{y_i,} = \Delta(x_i-x_j)\gamma(y_i-y_j) = -\Delta_{ij}^{,y_j}, \tag{F.28}$$

$$\Delta_{ij}^{x_i,y_j} = -\gamma(x_i-x_j)\gamma(y_i-y_j) = \Delta_{ij}^{y_i,x_j}. \tag{F.29}$$

Due to the dependence only on the difference $\vec{r}_i - \vec{r}_j$, we can also denote these terms as $\Delta_{ij}^{\mu,\nu} \equiv \Delta^{\mu,\nu}(\vec{r})$ where $\vec{r} = \vec{r}_i - \vec{r}_j$.

G | Circular aperture properties

A circular aperture is given by

$$\mathcal{A}(\vec{r}) = \frac{2}{\sqrt{\pi D^2}} \mathrm{circ}\left(\frac{2|\vec{r}|}{D}\right), \tag{G.1}$$

where $D = 2\mathrm{NA}/\lambda$ is its diameter[1] and $\mathrm{circ}(r)$ is unity for $r \leq 1$ and zero elsewhere. The point-spread function is its Fourier transform,

$$\begin{aligned}\psi(r,\phi) &= \int \mathrm{d}\vec{\rho}\, \mathcal{A}(\vec{\rho}) e^{-i2\pi\vec{\rho}\cdot\vec{r}} \\ &= \frac{2}{\sqrt{\pi D^2}} \int_0^{D/2} \rho\, \mathrm{d}\rho \int_0^{2\pi} \mathrm{d}\theta\; e^{-i2\pi\rho r\cos(\theta-\phi)}.\end{aligned} \tag{G.2}$$

Using the integral representation of the Bessel function [198]

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\theta\; e^{-ix\cos(\theta-\phi)}, \tag{G.3}$$

the recurrence relation $xJ_n'(x) + nJ_n(x) = xJ_{n-1}(x)$ and the integral

$$xJ_1(x) = \int_0^x \mathrm{d}x' x' J_0(x'), \tag{G.4}$$

the PSF can be shown to be the Airy function

$$\psi(\vec{r}) = \frac{1}{\sqrt{\pi}} \frac{J_1(\pi D|\vec{r}|)}{|\vec{r}|}, \tag{G.5}$$

which is, as its aperture, inversion symmetric, $\psi(-\vec{r}) = \psi(\vec{r})$, and L^2 normalized.

[1] This is due to the rescaling of the aperture plane coordinates, which we performed in Chapter 3.

Top-hat aperture

The one dimensional version of the circular aperture is the top-hat

$$\mathcal{A}(x) = \frac{1}{\sqrt{D}} \operatorname{rect}\left(\frac{x}{D}\right), \tag{G.6}$$

where $\operatorname{rect}(x)$ is unity for $|x| \leq 1/2$ and zero elsewhere. The corresponding PSF is

$$\psi(x) = \sqrt{D}\operatorname{sinc}(Dx), \tag{G.7}$$

where $\operatorname{sinc}(x) = \sin(\pi x)/\pi x$ is the normalised sinc function. Its derivative is

$$\frac{\partial\psi(x)}{\partial x} = \sqrt{D}\,\frac{\cos(\pi Dx) - \operatorname{sinc}(Dx)}{x}. \tag{G.8}$$

The overlap between PSFs

$$\Delta(x) = \int_{-\infty}^{+\infty} \mathrm{d}x'\ \psi^*(x'-x)\psi(x'), \tag{G.9}$$

is more easily evaluated via the convolution theorem, using the aperture function:

$$\Delta(x) = \int_{-\infty}^{+\infty} \mathrm{d}u\ e^{-i2\pi ux}|\mathcal{A}(u)|^2 = \operatorname{sinc}(Dx). \tag{G.10}$$

Similarly, the other overlap terms are

$$\begin{aligned}\gamma(x) &\equiv \int_{-\infty}^{+\infty} \mathrm{d}x'\ \psi^*(x'-x)\frac{\partial\psi(x')}{\partial x'} \\ &= -i2\pi\int_{-\infty}^{+\infty} \mathrm{d}u\ ue^{-i2\pi ux}|\mathcal{A}(u)|^2 \\ &= \frac{1}{x}\left(\cos(\pi Dx) - \operatorname{sinc}(Dx)\right)\end{aligned} \tag{G.11}$$

and

$$\begin{aligned}\tau(x) &= 4\pi^2\int_{-\infty}^{+\infty} \mathrm{d}u\ u^2e^{-i2\pi ux}|\mathcal{A}(u)|^2 \\ &= \frac{1}{x^2}\left[((\pi Dx)^2 - 2)\operatorname{sinc}(Dx) + 2\cos(\pi Dx)\right].\end{aligned} \tag{G.12}$$

Finally, The zero-displacement overlap $\tau(0)$, which corresponds to the spatial-frequency variance of the PSF, is

$$\tau(0) = \frac{\pi^2D^2}{3}. \tag{G.13}$$

Bessel basis

The PAD basis for the top-hat aperture is

$$\tilde{\phi}_m(u) = i^m g_m(u)\mathcal{A}(u), \tag{G.14}$$

where the $\{g_m(u)\}$ is a basis of polynomials which are orthonormal with respect to the weight $|\mathcal{A}(u)|^2$. Such a condition is satisfied by the normalised Legendre polynomials [34]

$$g_m(u) = \sqrt{2m+1}P_n\left(\frac{2u}{D}\right), \tag{G.15}$$

since [198]

$$\int_{-1}^{+1} \mathrm{d}x\ P_m(x)P_n(x) = \frac{2}{2m+1}\delta_{mn}. \tag{G.16}$$

By using the relation

$$\int_{-1}^{+1} \mathrm{d}u\ P_m(u)e^{iux} = i^m\sqrt{\frac{2\pi}{x}}J_{n+1/2}(x), \tag{G.17}$$

where $J_\alpha(x)$ is the α-th order Bessel function of the first kind, the PAD basis can be shown to be the Bessel basis

$$\phi_m(x) = \sqrt{m+1/2}\frac{J_{n+1/2}(\pi Dx)}{\sqrt{x}}. \tag{G.18}$$

The Bessel basis is orthonormal, complete and with well-defined parity, $\phi_m(-x) = (-1)^m\phi_m(x)$. The overlap with the PSF is

$$\begin{aligned} R_m(x) &= \phi_m^*(x) * \psi(x) \\ &= \mathcal{F}^{-1}\left[(-i)^m g_m(u)|\mathcal{A}(u)|^2\right](x) \\ &= \sqrt{m+1/2}\frac{J_{n+1/2}(\pi Dx)}{\sqrt{Dx}}. \end{aligned} \tag{G.19}$$

Since $J_{1/2}(x) = \sqrt{2/\pi x}\sin(x)$, the 0-th order mode is simply the PSF, $\phi_0(x) = \psi(x)$.

Fourier-Zernike basis

In the two-dimensional case, the PAD basis in the aperture plane is

$$\tilde{\phi}_{mn}(\vec{\rho}) = i^m g_{mn}(\vec{\rho})\mathcal{A}(\vec{\rho}), \tag{G.20}$$

where

$$\int \mathrm{d}\vec{\rho}\; g^*_{m,n}(\vec{\rho}) g_{m',n'}(\vec{\rho})|\mathcal{A}(\vec{\rho})|^2 = \delta_{m,m'}\delta_{n,n'}. \tag{G.21}$$

Using the polar coordinates (ρ, θ) of $\vec{\rho}$ and Eq. (G.1), Eq. (G.21) becomes

$$\frac{1}{\pi}\int_0^1 \mathrm{d}\rho\; \rho \int_0^{2\pi} \mathrm{d}\theta\; g^*_{m,n}(D\rho/2,\theta) g_{m',n'}(D\rho/2,\theta) = \delta_{m,m'}\delta_{n,n'}. \tag{G.22}$$

This condition is satisfied by the Zernike Polynomials $Z_n^m(\rho,\theta)$ [40, 198]:

$$\int_0^1 \mathrm{d}\rho\; \rho \int_0^{2\pi} \mathrm{d}\theta\; Z_n^m(\rho,\theta) Z_{n'}^{m'}(\rho,\theta) = \frac{(1+\delta_{m,0})}{2(n+1)}\pi\delta_{m,m'}\delta_{n,n'}. \tag{G.23}$$

The polynomials are then

$$g_{m,n}(\rho,\theta) = \sqrt{\frac{2(n+1)}{1+\delta_{m,0}}} Z_n^m(2\rho/D,\theta). \tag{G.24}$$

Therefore, in the aperture plane the Zernike basis is optimal. It is also convenient, as the modes are strictly restricted to the aperture disk, where light is present (unlike the HG modes for the Gaussian aperture, which extend way beyond where there is light). Zernike polynomials have some rules regarding their indices: the index n can be an arbitrary non-negative integer; the index m can be either positive or negative, but it must have the same parity as n (i.e., both must be either even or odd) and obey $|m| \leq n$. Furthermore, negative values of m lead to a change in orientation. Finally, their parity is defined as

$$Z_n^m(-\vec{r}) = (-1)^m Z_n^m(\vec{r}) = (-1)^n Z_n^m(\vec{r}), \tag{G.25}$$

since m and n are always both even or odd.

In the image plane, the corresponding PAD basis is [40, 71]

$$\phi_{mn}(\vec{r}) = \sqrt{\frac{2(n+1)}{1+\delta_{m,0}}} \frac{J_{n+1}(\pi D r)}{r} \sin\left(m\phi + \frac{\pi}{2}\mathcal{H}(m)\right) \tag{G.26}$$

where (r, ϕ) are the polar coordinates of $\vec{r}$ and $\mathcal{H}(m) = 1$ for $m \geq 0$ and $\mathcal{H}(m) = 0$ for $m < 0$. The PAD basis for the Airy PSF consists of Fourier transforms of Zernike polynomials, which we call the *Fourier-Zernike* (FZ) basis.

The overlap terms in two dimensions,

$$\Delta_{ij}^{\mu,\nu} \equiv \int \mathrm{d}\vec{r}\, \frac{\partial\psi^*(\vec{r}-\vec{r}_i)}{\partial\mu}\frac{\partial\psi(\vec{r}-\vec{r}_j)}{\partial\nu}, \tag{G.27}$$

can also be evaluated via Fourier transforms. For simplicity we use the notation $\Delta^{\mu,\nu}(\vec{r}) \equiv \Delta_{ij}^{\mu,\nu}$ where $\vec{r} = \vec{r}_i - \vec{r}_j$. The main terms are

$$\Delta(\vec{r}) = \int \mathrm{d}\vec{\rho}\, |\mathcal{A}(\vec{\rho})|^2 e^{-i2\pi\vec{\rho}\cdot\vec{r}}, \tag{G.28}$$

$$\Delta^{x_i,x_j}(\vec{r}) = 4\pi^2 \int \mathrm{d}\vec{\rho}\, \rho^2 \cos^2(\theta)|\mathcal{A}(\vec{\rho})|^2 e^{-i2\pi\vec{\rho}\cdot\vec{r}}, \tag{G.29}$$

$$\Delta^{y_i,y_j}(\vec{r}) = 4\pi^2 \int \mathrm{d}\vec{\rho}\, \rho^2 \sin^2(\theta)|\mathcal{A}(\vec{\rho})|^2 e^{-i2\pi\vec{\rho}\cdot\vec{r}}, \tag{G.30}$$

$$\Delta^{x_i,}(\vec{r}) = -i2\pi \int \mathrm{d}\vec{\rho}\, \rho\cos(\theta)|\mathcal{A}(\vec{\rho})|^2 e^{-i2\pi\vec{\rho}\cdot\vec{r}} = -\Delta^{,x_j}(\vec{r}), \tag{G.31}$$

$$\Delta^{y_i,}(\vec{r}) = -i2\pi \int \mathrm{d}\vec{\rho}\, \rho\sin(\theta)|\mathcal{A}(\vec{\rho})|^2 e^{-i2\pi\vec{\rho}\cdot\vec{r}} = -\Delta^{,y_j}(\vec{r}), \tag{G.32}$$

$$\Delta^{x_i,y_j}(\vec{r}) = 4\pi^2 \int \mathrm{d}\vec{\rho}\, \rho^2 \sin(\theta)\cos(\theta)|\mathcal{A}(\vec{\rho})|^2 e^{-i2\pi\vec{\rho}\cdot\vec{r}} = \Delta^{y_i,x_j}(\vec{r}). \tag{G.33}$$

The $\Delta(\vec{r})$ term is simply proportional to the PSF,

$$\Delta(\vec{r}) = \frac{2}{\pi D}\frac{J_1(\pi D|\vec{r}|)}{|\vec{r}|}. \tag{G.34}$$

For the $\Delta^{x_i,}(\vec{r})$ term, we need the relations [40]

$$\int_0^{2\pi} \mathrm{d}\theta e^{\pm ix\cos(\theta-\phi)}\cos(n\theta) = (\mp i)^n 2\pi\cos(n\phi)J_n(x), \tag{G.35}$$

$$\int_0^{2\pi} \mathrm{d}\theta e^{\pm ix\cos(\theta-\phi)}\sin(n\theta) = (\mp i)^n 2\pi\sin(n\phi)J_n(x). \tag{G.36}$$

With Eq. (G.35), it can be rewritten as

$$\Delta^{x_i,}(\vec{r}) = -\frac{16\pi}{D^2}\cos(\phi)\int_0^{D/2} \mathrm{d}\rho\, \rho^2 J_1(2\pi r\rho). \tag{G.37}$$

The Bessel function relation [198]

$$x^n J_{n-1}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left[x^n J_n(x)\right] \tag{G.38}$$

can be used in the integral, and the final form is

$$\Delta^{x_i,}(\vec{r}) = -2\cos(\phi)\frac{J_2(\pi Dr)}{r}. \tag{G.39}$$

Likewise,

$$\Delta^{y_i,}(\vec{r}) = -2\sin(\phi)\frac{J_2(\pi Dr)}{r}. \tag{G.40}$$

The $\Delta^{x_i,x_j}(\vec{r})$ term is

$$\Delta^{x_i,x_j}(\vec{r}) = -\frac{16\pi^2}{D^2}\cos(\phi)\int_0^{D/2} \mathrm{d}\rho\ \rho^3\left[J_2(2\pi r\rho)\cos(2\phi) - J_0(2\pi r\rho)\right]. \tag{G.41}$$

Again, using Eq. (G.38) leads to

$$\Delta^{x_i,x_j}(\vec{r}) = -\pi D(\cos(2\phi)+1)\frac{J_3(\pi Dr)}{r} + 2\frac{J_2(\pi Dr)}{r^2}. \tag{G.42}$$

Analogously,

$$\Delta^{y_i,y_j}(\vec{r}) = -\pi D(-\cos(2\phi)+1)\frac{J_3(\pi Dr)}{r} + 2\frac{J_2(\pi Dr)}{r^2}. \tag{G.43}$$

Finally,

$$\Delta^{x_i,y_j}(\vec{r}) = -\pi D\sin(2\pi)\frac{J_3(\pi Dr)}{r}. \tag{G.44}$$

H | Single mode Fisher information

Proposition 6. *Consider a thermal state with average photon number* Γ*. The Fisher information related to the estimation of* Γ *is, after an intensity measurement,*

$$\mathcal{F}^{(n)}(\Gamma) = \frac{1}{\Gamma\left(\Gamma+1\right)}. \tag{H.1}$$

For a single quadrature measurement,

$$\mathcal{F}^{(q)}(\Gamma) = \frac{2}{\left(2\Gamma+1\right)^2}, \tag{H.2}$$

and for a full quadrature measurement,

$$\mathcal{F}^{(q,p)}(\Gamma) = \frac{1}{\left(\Gamma+1\right)^2}. \tag{H.3}$$

Proof. The probability for detecting n photons in an intensity measurement is

$$\Pr(n) = \frac{1}{\Gamma+1}\left(\frac{\Gamma}{\Gamma+1}\right)^n. \tag{H.4}$$

Its natural logarithm is

$$\ln\Pr(n) = n\ln\Gamma - (n+1)\ln(\Gamma+1), \tag{H.5}$$

whose derivative is

$$\frac{\partial \ln\Pr(n)}{\partial\ \Gamma} = \frac{n-\Gamma}{\Gamma(\Gamma+1)}. \tag{H.6}$$

The Fisher information is then

$$\begin{aligned}\mathcal{F}^{(n)}(\Gamma) &= \sum_{n=0}^{\infty}\left[\frac{\partial}{\partial\Gamma}\ln\Pr\left(n|\Gamma\right)\right]^2\Pr\left(n|\Gamma\right)\\ &= \frac{1}{(\Gamma)^2(\Gamma+1)^3}\sum_{n=0}^{\infty}\left(n-\Gamma\right)^2\left(1+\frac{1}{\Gamma}\right)^{-n} = \frac{1}{\Gamma\left(\Gamma+1\right)},\end{aligned} \tag{H.7}$$

where we used

$$\sum_{n=0}^{\infty} \frac{(n-\Gamma)^2}{\left(1+\frac{1}{\Gamma}\right)^n} = \Gamma(\Gamma+1)^2. \tag{H.8}$$

The probability of measuring a quadrature value q is

$$\Pr(q) = \frac{e^{-q^2/(2\Gamma+1)}}{\sqrt{\pi(2\Gamma+1)}}. \tag{H.9}$$

The derivative term in the Fisher information is

$$\frac{\partial \ln \Pr(q)}{\partial\, \Gamma} = \frac{2q^2-(2\Gamma+1)}{(2\Gamma+1)^2}. \tag{H.10}$$

The FI is then

$$\begin{aligned}
\mathcal{F}^{(q)}(\Gamma) &= \int_{-\infty}^{+\infty} \mathrm{d}q \ \left[\frac{\partial}{\partial\Gamma} \ln \Pr\left(q|\Gamma\right)\right]^2 \Pr\left(q|\Gamma\right) \\
&= \frac{1}{\sqrt{\pi(2\Gamma+1)^5}} \int_{-\infty}^{+\infty} \mathrm{d}q \ \left(\frac{2q^2}{2\Gamma+1}-1\right)^2 e^{-q^2/(2\Gamma+1)} \\
&= \frac{1}{\sqrt{\pi}} \frac{1}{(2\Gamma+1)^2} \int_{-\infty}^{+\infty} \mathrm{d}y \ \left(2y^2-1\right)^2 e^{-y^2} = \frac{2}{(2\Gamma+1)^2}.
\end{aligned} \tag{H.11}$$

Finally, the probability of a full quadrature measurement (q,p) is

$$\Pr(q,p) = \frac{e^{-(q^2+p^2)/2(\Gamma+1)}}{2\pi(\Gamma+1)}, \tag{H.12}$$

the derivative is

$$\frac{\partial \ln \Pr(q,p)}{\partial\, \Gamma} = \frac{1}{\Gamma+1}\left[\frac{q^2+p^2}{2(\Gamma+1)}\right], \tag{H.13}$$

And the FI is

$$\begin{aligned}
\mathcal{F}^{(q,p)}(\Gamma) &= \iint_{-\infty}^{+\infty} \mathrm{d}q\mathrm{d}p \left[\frac{\partial}{\partial\Gamma} \ln \Pr\left(q,p|\Gamma\right)\right]^2 \Pr\left(q,p|\Gamma\right) \\
&= \iint_{-\infty}^{+\infty} \mathrm{d}q\mathrm{d}p \frac{1}{(\Gamma+1)^2}\left[\frac{q^2+p^2}{2(\Gamma+1)}\right]^2 \frac{e^{-(q^2+p^2)/2(\Gamma+1)}}{2\pi(\Gamma+1)} \\
&= \frac{1}{(\Gamma+1)^2}.
\end{aligned} \tag{H.14}$$

□

Proposition 7. *Consider a coherent state with complex amplitude A. For an intensity measurement, the Fisher information related to the estimation of the absolute value of A is*

$$\mathcal{F}^{(n)}(|A|) = 4. \tag{H.15}$$

For a single quadrature measurement, the FI related to the estimation of the real part of A is

$$\mathcal{F}^{(q)}(\mathrm{Re}\{A\}) = 4. \tag{H.16}$$

For a full quadrature measurement, the FI related to the estimation of any two parameters θ_i and θ_j is

$$\mathcal{F}_{ij}^{(q,p)} = 2\,\mathrm{Re}\left\{\frac{\partial A^*}{\partial \theta_i}\frac{\partial A}{\partial \theta_j}\right\}. \tag{H.17}$$

Proof. The probability of detecting n photons in a photon counting measurement is

$$\Pr(n) = \frac{e^{-|A|^2}|A|^{2n}}{n!}. \tag{H.18}$$

Note that intensity measurements are insensitive to the phase of A, and therefore only its absolute value can be estimated. Consider the derivative term

$$\frac{\partial \ln \Pr(n)}{\partial |A|^2} = \frac{n}{|A|^2} - 1. \tag{H.19}$$

It can be used in the FI for $|A|^2$,

$$\begin{aligned}\mathcal{F}^{(n)}(|A|^2) &= \sum_{n=0}^{\infty}\left[\frac{\partial}{\partial |A|^2}\ln \Pr\left(n||A|^2\right)\right]^2 \Pr\left(n||A|^2\right) \\ &= \frac{e^{-|A|^2}}{|A|^4}\sum_{n=0}^{\infty}\frac{|A|^{2n}}{n!}\left(n-|A|^2\right)^2 = \frac{1}{|A|^2}.\end{aligned} \tag{H.20}$$

Then, by reparametrization, the FI related to $|A|$ is

$$\mathcal{F}^{(n)}(|A|) = \left(\frac{\partial |A|^2}{\partial |A|}\right)^2 \mathcal{F}^{(n)}(|A|^2) = 4. \tag{H.21}$$

For quadrature measurements, the probability is

$$\Pr(q) = \frac{1}{\sqrt{\pi}} e^{-\left(q-\sqrt{2}\,\mathrm{Re}\{A\}\right)^2}. \tag{H.22}$$

Now the measurement is insensitive to the imaginary part of A, which cannot be estimated. The derivative term is

$$\frac{\partial \ln \Pr(q)}{\partial \operatorname{Re}\{A\}} = 2\sqrt{2}(q - \sqrt{2}\operatorname{Re}\{A\}). \tag{H.23}$$

The FI is then

$$\begin{aligned}\mathcal{F}^{(q)}(\operatorname{Re}\{A\}) &= \int_{-\infty}^{+\infty} \mathrm{d}q \left[\frac{\partial}{\partial \operatorname{Re}\{A\}} \ln \Pr\left(q|\operatorname{Re}\{A\}\right)\right]^2 \Pr\left(q|\operatorname{Re}\{A\}\right) \\ &= \frac{8}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \mathrm{d}q \left(q - \sqrt{2}\operatorname{Re}\{|A|\}\right)^2 e^{-\left(q-\sqrt{2}\operatorname{Re}\{|A|\}\right)^2} = 4. \end{aligned} \tag{H.24}$$

A full quadrature measurement has probabilities

$$\Pr(q,p) = \frac{1}{2\pi} e^{-\left(q-\sqrt{2}\operatorname{Re}\{A\}\right)^2/2} e^{-\left(p-\sqrt{2}\operatorname{Im}\{A\}\right)^2/2}. \tag{H.25}$$

Now, the measurement is fully sensitive to A. The derivative terms are, via the chain rule,

$$\frac{\partial \ln \Pr(q)}{\partial \theta_i} = \frac{\partial \ln \Pr(q)}{\partial \operatorname{Re}\{A\}} \frac{\partial \operatorname{Re}\{A\}}{\partial \theta_i} + \frac{\partial \ln \Pr(q)}{\partial \operatorname{Im}\{A\}} \frac{\partial \operatorname{Im}\{A\}}{\partial \theta_i}, \tag{H.26}$$

where

$$\frac{\partial \ln \Pr(q)}{\partial \operatorname{Re}\{A\}} = \sqrt{2}(q - \sqrt{2}\operatorname{Re}\{A\}), \tag{H.27}$$

$$\frac{\partial \ln \Pr(q)}{\partial \operatorname{Im}\{A\}} = \sqrt{2}(q - \sqrt{2}\operatorname{Im}\{A\}). \tag{H.28}$$

After reparametrization, the Fisher information can be shown to be

$$\begin{aligned}\mathcal{F}_{ij}^{(q,p)} =& \frac{\partial \operatorname{Re}\{A\}}{\partial \theta_i} \frac{\partial \operatorname{Re}\{A\}}{\partial \theta_j} \mathcal{F}^{(q,p)}(\operatorname{Re}\{A\}) + \frac{\partial \operatorname{Im}\{A\}}{\partial \theta_i} \frac{\partial \operatorname{Im}\{A\}}{\partial \theta_j} \mathcal{F}^{(q,p)}(\operatorname{Im}\{A\}) \\ &+ \left(\frac{\partial \operatorname{Re}\{A\}}{\partial \theta_i} \frac{\partial \operatorname{Im}\{A\}}{\partial \theta_j} + \frac{\partial \operatorname{Im}\{A\}}{\partial \theta_i} \frac{\partial \operatorname{Re}\{A\}}{\partial \theta_j}\right) \mathcal{F}^{(q,p)}(\operatorname{Re}\{A\}, \operatorname{Im}\{A\}), \end{aligned} \tag{H.29}$$

where the other FI terms are $\mathcal{F}^{(q,p)}(\operatorname{Re}\{A\}) = \mathcal{F}^{(q,p)}(\operatorname{Im}\{A\}) = 2$, the FI for estimation of the real and imaginary parts, and the cross-term $\mathcal{F}^{(q,p)}(\operatorname{Re}\{A\}, \operatorname{Im}\{A\}) = 0$. Then, the FI can be simplified to

$$\begin{aligned}\mathcal{F}_{ij}^{(q,p)} &= 2\left(\frac{\partial \operatorname{Re}\{A\}}{\partial \theta_i} \frac{\partial \operatorname{Re}\{A\}}{\partial \theta_j} + \frac{\partial \operatorname{Im}\{A\}}{\partial \theta_i} \frac{\partial \operatorname{Im}\{A\}}{\partial \theta_j}\right) \\ &= 2\operatorname{Re}\left\{\frac{\partial A^*}{\partial \theta_i} \frac{\partial A}{\partial \theta_j}\right\}. \end{aligned} \tag{H.30}$$

□

I | Numerical evaluation of the QFI

Here we evaluate the Quantum Fisher information for the simultaneous estimation of all intensities and positions of a two-dimensional distribution of weak incoherent point sources. We use the method of non-orthogonal bases [135], outlined in Section 2.3.2.

Consider S weak incoherent sources at the positions $\{\vec{r}_i = (x_i, y_i)\}$ on the object plane, each with relative intensity p_i. The quantum state of the system at the image plane is

$$\rho = \sum_{i=1}^{S} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|, \tag{I.1}$$

where

$$\left|\psi_i\right\rangle = \int \mathrm{d}\vec{r}\psi(\vec{r} - \vec{r}_i) \left|\vec{r}\right\rangle, \tag{I.2}$$

and $\left|\vec{r}\right\rangle \equiv \hat{a}^\dagger_{\vec{r}} \left|\mathrm{vac}\right\rangle$ is the one-photon state at position $\vec{r}$. We are interested in estimating the $3S$-dimensional vector of parameters $\boldsymbol{\theta} = (p_1, \ldots, p_S, x_1, \ldots x_S, y_1, \ldots, y_S)$.

In order to apply the method of non-orthogonal bases, we need to find a linearly independent set of vectors that covers the support of ρ and its derivatives $\partial_i \rho$. For ρ, its support is covered by the $\{\left|\psi_i\right\rangle\}_{i=1}^{S}$. For the derivatives, let's first take a look at them. The derivative with respect to intensities is simply

$$\frac{\partial \rho}{\partial p_i} = \left|\psi_i\right\rangle \left\langle\psi_i\right|, \tag{I.3}$$

which is already inside the support of ρ. The position derivatives are

$$\frac{\partial \rho}{\partial x_i} = p_i \left(\left|\partial_{x_i}\psi_i\right\rangle \left\langle\psi_i\right| + \left|\psi_i\right\rangle \left\langle\partial_{x_i}\psi_i\right|\right), \tag{I.4}$$

and analogously for y_i. Now, the states

$$|\partial_{x_i}\psi_i\rangle = \int \mathrm{d}\vec{r}\, \frac{\partial\psi(\vec{r}-\vec{r}_i)}{\partial x_i}\,|\vec{r}\rangle \tag{I.5}$$

are outside the support of ρ, and we need to add them to our basis. It is easy to see that the set $\{|\psi_i\rangle, |\partial_{x_i}\psi_i\rangle, |\partial_{y_i}\psi_i\rangle\}_{i=1}^{S}$ covers the support of the states and its derivatives, so we're halfway through. It remains to show that it is linearly independent, which happens when the determinant of its Gramian matrix is nonzero. We we will use this piece of notation from Appendix F, where

$$\Delta_{ij}^{\mu,\nu} \equiv \langle\partial_\mu\psi_i|\partial_\nu\psi_j\rangle = \int \mathrm{d}\vec{r}\, \frac{\partial\psi^*(\vec{r}-\vec{r}_i)}{\partial\mu}\frac{\partial\psi(\vec{r}-\vec{r}_j)}{\partial\nu}. \tag{I.6}$$

A missing superscript means that no derivation is performed, as in

$$\Delta_{ij}^{\mu,} = \langle\partial_\mu\psi_i|\psi_j\rangle = \int \mathrm{d}\vec{r}\, \frac{\partial\psi^*(\vec{r}-\vec{r}_i)}{\partial\mu}\psi(\vec{r}-\vec{r}_j). \tag{I.7}$$

Then, the Gramian matrix is

$$\Upsilon = \begin{bmatrix} \Delta & \Delta^{,x} & \Delta^{,y} \\ \Delta^{x,} & \Delta^{x,x} & \Delta^{x,y} \\ \Delta^{y,} & \Delta^{y,x} & \Delta^{y,y} \end{bmatrix}, \tag{I.8}$$

where each element represents a $S \times S$ submatrix. In Appendixes F and G we evaluate the Gramian matrix terms for Gaussian and circular apertures, respectively. If there are no degenerate sources, i.e., sources exactly on top of each other, the Gramian determinant is always nonzero.

We are now ready to find the P and $D^{(i)}$ matrices necessary for the QFI evaluation. The P matrix, which represents the state ρ in the new non-orthogonal basis, is simply

$$P = \begin{bmatrix} Q & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{I.9}$$

where $Q_{ij} = \delta_{ij}p_i$ is a diagonal matrix containing the relative intensities. The $D^{(i)}$, which are the representation of the derivatives $\partial_i\rho$, are

$$D^{(p_i)} = \begin{bmatrix} \chi & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ D^{(x_i)} = p_i \begin{bmatrix} 0 & \chi_i & 0 \\ \chi_i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ D^{(y_i)} = p_i \begin{bmatrix} 0 & 0 & \chi_i \\ 0 & 0 & 0 \\ \chi_i & 0 & 0 \end{bmatrix}, \tag{I.10}$$

where $(\chi_i)_{\alpha\beta} = \delta_{\alpha i}\delta_{\beta i}$, i.e., the matrix χ_i is one on the i-th diagonal element and zero elsewhere. We use these matrices to find the SLD matrices via

$$\Big|D^{(i)}\Big\rangle\Big\rangle = \mathbb{M}\Big|L^{(i)}\Big\rangle\Big\rangle, \tag{I.11}$$

where $L^{(i)}$ is the representation of the SLD matrix $\mathcal{L}_i$ and

$$\mathbb{M} = \frac{1}{2}\left[(\Upsilon P)^\intercal \otimes I + I \otimes P\Upsilon\right]. \tag{I.12}$$

The matrix $\mathbb{M}$ is non-invertible, since the P matrices are not full rank. However, it still works [136] if we use its pseudo-inverse instead, so

$$\Big|L^{(i)}\Big\rangle\Big\rangle = \mathbb{M}^{+}\Big|D^{(i)}\Big\rangle\Big\rangle. \tag{I.13}$$

Then, we can simply "un-vectorise" (i.e., "un-stack" the columns and put them back in matrix form) $\Big|L^{(i)}\Big\rangle\Big\rangle$ and use it to get

$$\mathcal{H}_{ij} = \mathrm{Tr}\left[P\Upsilon L^{(i)}\Upsilon L^{(j)}\Upsilon\right], \tag{I.14}$$

whose real and imaginary parts yield the QFI and compatibility matrices, respectively.

J | An optical Eratosthenes' sieve for large prime numbers

We report the first experimental demonstration of a prime number sieve via linear optics [199]. The prime numbers distribution is encoded in the intensity zeros of the far field produced by a spatial light modulator hologram, which comprises a set of diffraction gratings whose periods correspond to all prime numbers below 149. To overcome the limited far field illumination window and the discretization error introduced by the spatial light modulator finite spatial resolution, we rely on additional diffraction gratings and sequential recordings of the far field. This strategy allows us to optically sieve all prime numbers below $149^2 = 22201$.

Introduction

The study of prime numbers, although millenia old, is still a trending topic in mathematics. On the theoretical side, there are relevant fundamental problems which remain open, such as the Riemann hypothesis and the primes' distribution [200, 201]. On a more applied note, prime factorization, which is the cornerstone of modern cryptography [202, 203], motivates a search for more efficient algorithms.

In the past decades, physics emerged as a fruitful venue for progress on prime numbers research. Most efforts were focused on factorization, culminating in the celebrated Shor's algorithm [204] and the current quest for practical quantum computers. However, the application of analogue physical tools for more fundamental mathematical problems is also being investigated, for example with respect to the Riemann hypothesis [205–208].

A new and relatively unexplored testbed for the connection between physics and prime numbers is classical optics. On the factorization front, interferometric [209, 210] and optoelectronically assisted algorithms [211] schemes have been proposed, and experimental demonstrations were carried out based on the Talbot effect [212, 213] and polychromatic interference in a multipath interferometer [214]. On the fundamental side, Berry [215, 216] proposed the reading of Riemann function's zeros from radiation patterns using the property of the far field propagation to represent the Fourier transform.

Inspired by Berry's proposal, Petersen et al. [217] have recently proposed a physical prime numbers sieve based on the superposition of identical waves. The scheme is based on the sieve of Eratosthenes [218], one of the oldest algorithms for finding prime numbers. Given the sequence of integers n from 2 to M, Eratosthenes' sieve iteratively discards their multiples, such that all remaining integers lower than M^2 are necessarily primes. Ref. [217] proposes to realize the sieving by gratings of equally spaced n point-like sources: in the far field profile, each grating interferes constructively at locations corresponding to the multiples of n, and destructively at other integers. The pattern resulting from all these gratings has the prime numbers encoded in its intensity zeros, and locating them permits the realization of the sieve.

In this paper, we implement and further develop the scheme devised in [217] by means of a phase-only spatial light modulator (SLM), where the apertures are realized by phase masks illuminated by a monochromatic laser. The far field intensity profile is measured by a CMOS camera, and the optical sieve corresponds to intensity minima of that profile. By making use of additional diffraction gratings and sequential camera acquisitions we can extend the sieving range beyond the pixel resolution of the SLM and the camera. The only remaining limitation associated with the SLM pixel resolution is the number of gratings that can be fitted in the phase mask. Capped by this limitation, we demonstrate an optical sieve capable of identifying all prime numbers below $149^2 = 22201$.

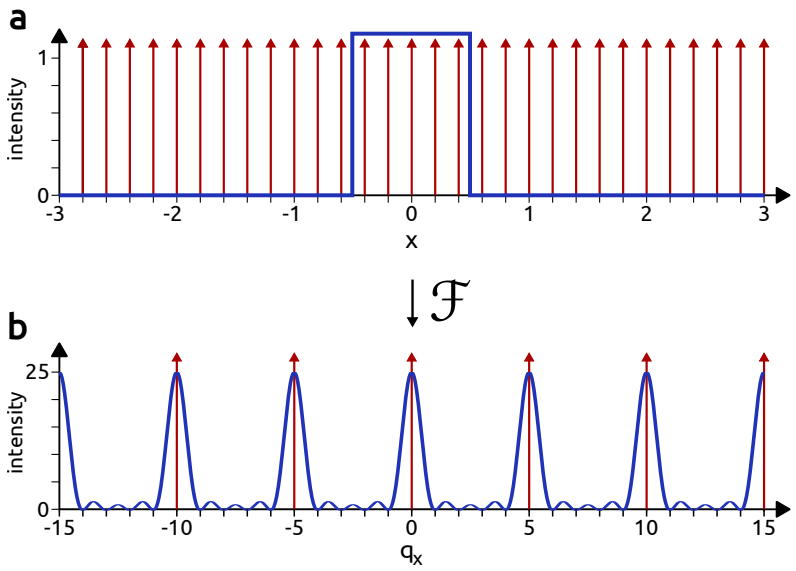


Figure J.1: Light intensity profiles at near (a) and far (b) fields, for $n = 5$ pinholes per unit length. Red arrows represent an infinite array of pinholes and the corresponding effect on the far field. When this array is restricted to five pinholes [which corresponds to multiplication by the blue envelope in (a)], the far field pattern (which is the result of a Fourier transformation $\mathcal{F}$ of the near field) becomes convolved with a sinc function, but the intensity at all integers except multiples of 5 remains at zero.

Theory

Sieving mask.

To illustrate the optical sieve proposed in [217], consider an amplitude grating of n equally spaced point-like sources, aligned along the horizontal direction x and with spatial frequency $1/n$:

$$g_n(x) = \sum_{i=0}^{n-1} \delta(x - i/n), \tag{J.1}$$

where $\delta(x)$ is the Dirac delta function. This can equivalently be seen as an infinite grating of pinholes illuminated by a plane monochromatic wave over a unit length window, as shown in Fig. J.1. Since far field propagation corresponds to taking a Fourier transform

$$\tilde{g}_n(q_x) = \int_{-\infty}^{\infty} g_n(x) e^{-2\pi i x q_x} dx, \tag{J.2}$$

where q_x is the horizontal coordinate in the far field plane, this configuration

generates a field

$$\tilde{g}_n(q_x) = \left(n \sum_{j=-\infty}^{\infty} \delta(q_x - nj) \right) * \mathrm{sinc}(q_x), \tag{J.3}$$

where $\mathrm{sinc}(t) = \sin(\pi t)/\pi t$, of equally spaced peaks (due to the pinholes) convolved with a sinc envelope (due to the illumination window). The resulting intensity pattern has zeros at all integers except the multiples of n.

Superposing, in the object plane, these gratings g_n for $n \in \{2, 3, 5, .., M\}$ leads to a far field interference pattern where, akin to the sieve of Erastothenes, the multiples of the n's are "discarded" (by becoming intensity peaks) and the remaining zeros correspond to q_x being a prime number. Therefore, via the mapping of the far field zeros, the scheme sieves the prime numbers in the range $M' \leq q_x < M'^2$, where M' is the next prime after M.

The superposition of gratings (which we refer to as the *sieving mask*) can be implemented by displaying each of them at a different vertical position, as shown in Fig. J.2 (left column). In the far field plane, Fig. J.2 (centre column), the waves from different gratings interfere, but the intensity at the primes' q_x positions remains zero for all q_y. Therefore, a horizontal profile of the far field [Fig. J.2 (right column)], obtained by integration along q_y, allows sieving of the prime numbers as described before.

Displacement grating.

The above treatment assumes ideal point-like pinholes. A more realistic model would consider finite-size rectangular apertures, as depicted in Fig. J.2(b), which are mathematically represented by the convolution

$$E_n(x) = g_n(x) * T_w(x), \tag{J.4}$$

where $T_w(x)$ is the top-hat function and w is the aperture width. The resulting far field pattern will be modulated by a sinc envelope $\mathrm{sinc}(wq_x)$. This envelope affects the pattern's visibility, which progressively degrades for higher values of $|q_x|$, and

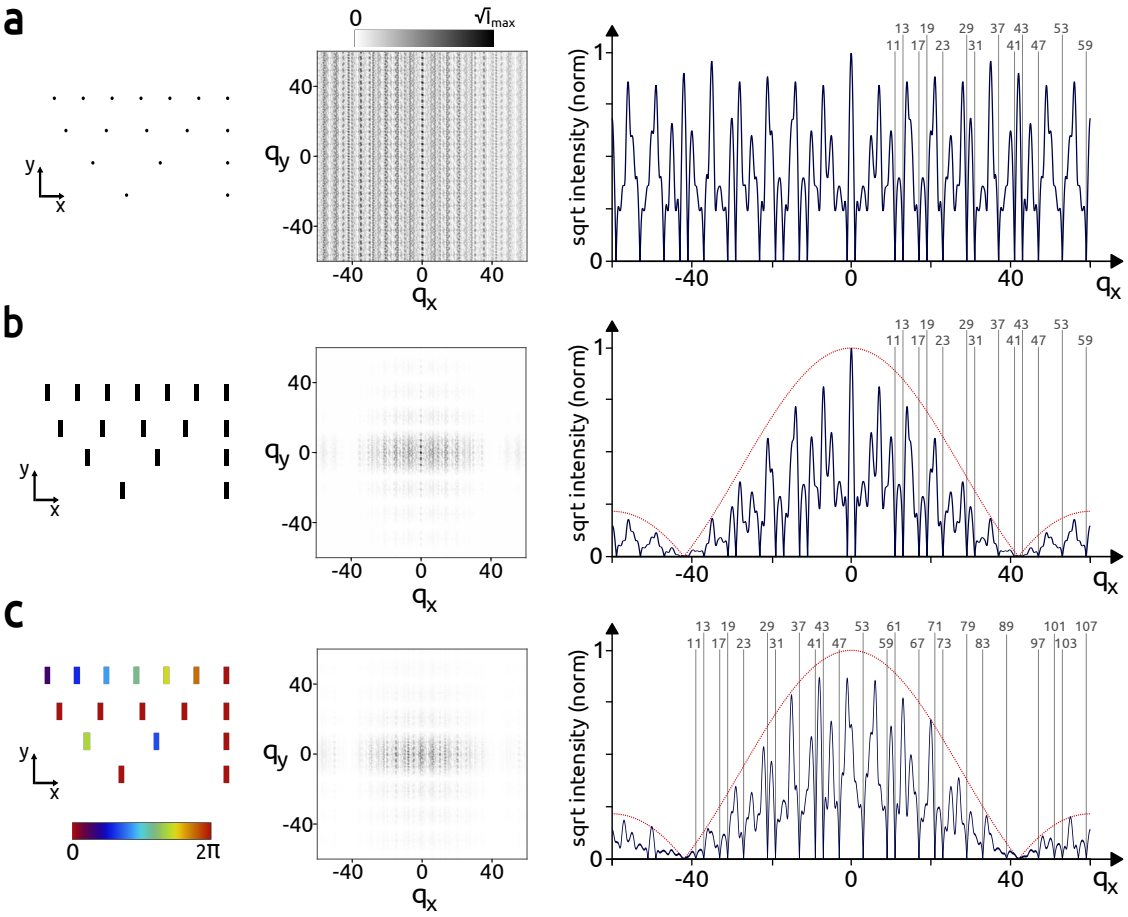


Figure J.2: Optical sieve diffraction mask (left) in the case of $\{n\} = \{2, 3, 5, 7\}$. The square root of the corresponding far field intensity profile (centre) and its horizontal marginal (integral over q_y, right), which reveals the prime numbers in the intensity zeros, highlighted by vertical grey lines. (a) Sieving mask proposed in [217], composed of point-like apertures. (b) Sieving mask with finite-size rectangular apertures of width $w = 1/42$. (c) Sieving mask superposed with displacement grating, which shifts the primes distribution by $d_x = 50$.

also introduces additional zeros, which do not correspond to prime numbers. In practice, it limits the sieve's working range to the central lobe of the sinc.

This issue is addressed by adding a position-dependent linear function, dubbed here *displacement grating*, to the phase to each aperture [Fig. J.2(c)]:

$$g'_n(x) = g_n(x)e^{i2\pi x d_x}. \tag{J.5}$$

It results in a spatial shift of the intensity zeros by d_x in the q_x coordinate of the far field,

$$\tilde{g}'_n(q_x) = \tilde{g}_n(q_x - d_x) \tag{J.6}$$

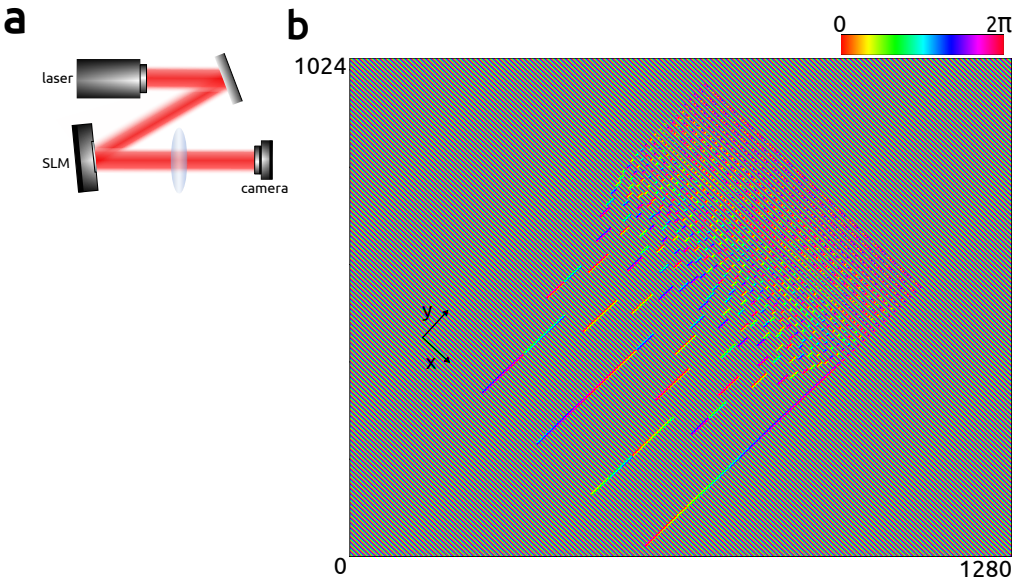


Figure J.3: Experimental implementation of the optical prime sieve. a) Scheme of the experiment. b) Phase mask hologram as displayed on the SLM screen.

so that

$$\tilde{E}'_n(x) = \tilde{g}'_n(q_x) \cdot w\,\mathrm{sinc}(wq_x). \tag{J.7}$$

Because we keep the phase inside each aperture constant, the modulation envelope $\mathrm{sinc}(wq_x)$ remains undisplaced. By progressively increasing the slope d_x of the displacement grating we are able to scan different sections of the Eratosthenes' sieve over the central lobe of the sinc envelope, therefore covering the full sieving range.

Experiment

As illustrated in the scheme of the experimental setup in Fig. J.3(a), the optical prime sieve consists of a phase-only liquid crystal on silicon (LCoS) SLM illuminated almost uniformly by a continuous-wave laser (Eagleyard EYP-DFB-0785 with $\lambda = 785$ nm) beam, along with a camera placed in its far field.

The spatial profile of the emitted laser beam is filtered by a single mode fibre and magnified by a telescope before illuminating the SLM screen. At the SLM, the laser power is approximately 0.1mW. The phase-only LCoS-SLM (Hamamatsu X13138-02) has a spatial resolution of 1280×1024 pixels, 12.5um pixel pitch, 60Hz refresh rate and 256 input signal levels.

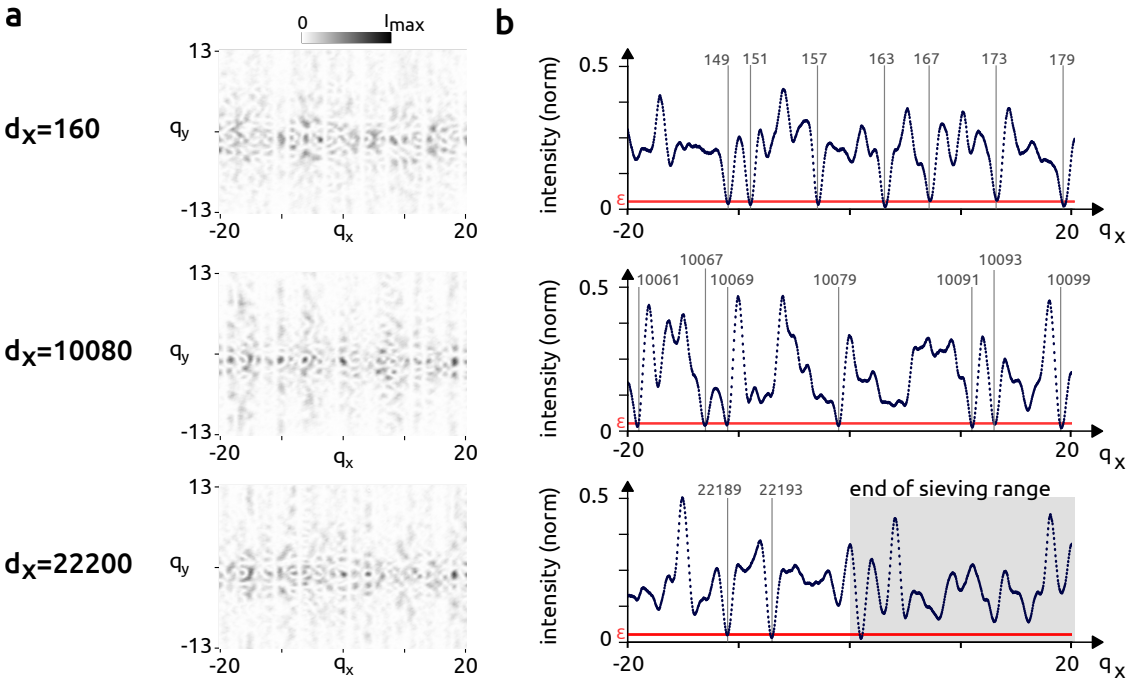


Figure J.4: Experimental results. a) Far field intensity images as acquired by the camera for different displacement grating slopes d_x, rescaled into the units of q_x, q_y. b) Integrals of the acquired intensity images along q_y. All minima below the intensity threshold ϵ (horizontal red line) are classified as prime numbers (vertical grey lines).

A set of optical lenses ($f_1 = 250$mm, $f_2 = 75$mm, $f_3 = 150$mm) produces and magnifies the far field pattern. A monochrome CMOS sensor (uEye UI-3270 CP-M-NO-R2), of pixel pitch 3.45um, aligned with the SLM diagonal and anti-diagonal axis, acquires the far field intensity profile. The camera spatial resolution (2056×1542 pixels) and the chosen optics allow a proper spatial sampling rate, with approximately 10 camera pixels corresponding to $\Delta q_x = 1$. The camera's digitiser is set to 12-bit depth and the exposure time to 3.15 ms, to maximize intensity resolution and avoid saturation.

The sieving mask is implemented by phase holograms in the SLM screen, as exemplified in Fig. J.3(b). The background outside the "apertures" is filled with a blazed grating whose lines are oriented along the x axis. The function of this grating is to diffract the light away from the sieving area, so only the light reflected from the apertures is captured by the camera. The mask is arranged diagonally at 45°, in order to prevent the light scattered from both the SLM screen borders and the space between pixels, in the horizontal and vertical directions, from contaminating

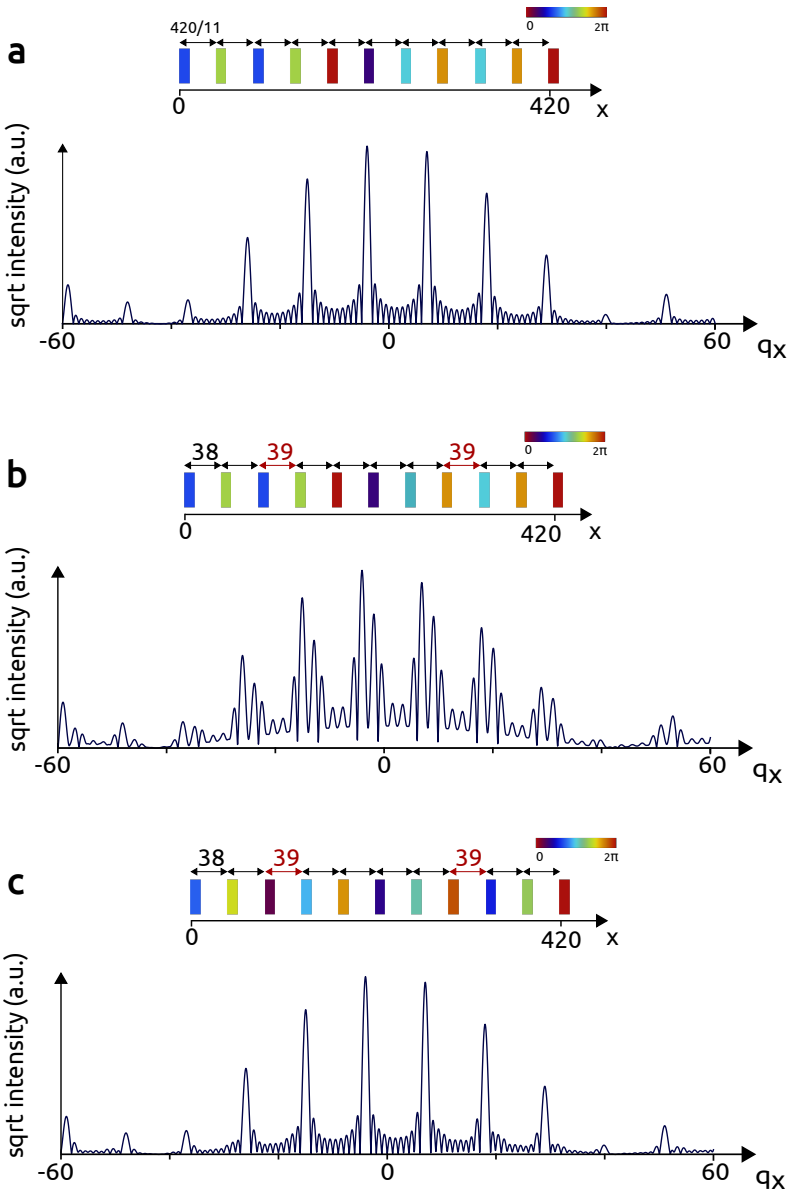


Figure J.5: Effect of the SLM's spatial resolution on the produced far field intensity profile for $n = 11$ and a displacement grating of slope $d_x = 180$ (theoretical model). (a) Evenly spaced apertures (infinite resolution); (b) Discrete pixels with a total pattern width of $W = 420$ pixels: the apertures are spaced by either 38 or 39 pixels. The integers not multiples of 11 are no longer zeros in the far field intensity profile. (c) $W = 420$ with a displacement grating of slope $d_x^{(11)} = 180 \bmod 11 = 4$: the desired structure of the far field is closely restored.

the image (in the laboratory reference frame, the SLM is mounted upright and the camera is physically tilted). Additionally, we apply a linear phase displacement to the apertures, i.e. we multiply $E'_n(x)$ by $e^{i2\pi x d_0}$, which displaces the entire far field pattern by d_0, thereby separating it from the SLM screen's specular reflection.

The patterns printed on the SLM extend over 420 pixels along x, accommodating apertures which are 3 pixels wide, corresponding to the effective width $w = 1/140$. These patterns include gratings corresponding to prime numbers up to $M = 139$, allowing a sieving range of $139 < q_x < 149^2 = 22201$, 149 being the next prime number after 139. The displacement gratings' slopes d_x are varied from zero to 22180 in increments of $R = 40$, covering the whole sieving range. While the size of the central lobe of the sinc function is $2/w = 280$, we chose a smaller increment value for higher contrast.

To improve the contrast, we set the slits' length along y to round$(200/n)$ pixels, so that each grating g_n contributes equally to the far field intensity pattern. We substitute g_2 and g_3 by pairs of gratings with $n = 4$ and $n = 6$, respectively, in order to reduce the space required by the grating with the longest slits. Additional displacement gratings are applied to one element of each pair so the collective action of the pair of elements is equivalent to that of g_2 and g_3. Further, we extend the total length of g_4 by a factor of 1.6, since it is located close to the SLM screen border and the laser power distribution, which is not perfectly uniform, is slightly lower there.

A key requirement on the optical prime sieve is the even spacing of its apertures, combined with their precise positioning. This is difficult to achieve in practice given a large number of gratings that need to be printed, as well as the finite SLM pixel resolution. If the apertures' spacing is rounded to the closest integer, the far field pattern becomes severely distorted and can no longer be used for sieving [Fig. J.5(a,b)].

This distortion is significant when the function $\tilde{g}_n(\cdot)$ has a large argument, i.e., according to Eq. (J.6), for large displacement grating slopes d_x. We address this issue by recalling that each function $\tilde{g}_n(\cdot)$ is periodic with zeros at the integers not multiples of n. Hence displacing it by d_x is equivalent to displacing it by the remainder of the division of d_x by n. Therefore the slope of each displacement grating can be modified according to $d_x^{(n)} = d_x \bmod n$ without any effect on the sieve, albeit at the cost of a different displacement slope for each line g_n. The effect

of this correction is illustrated in Fig. J.5(c). This compensation allows us to keep the argument of $\tilde{g}_n(\cdot)$ low for all g_n, thereby achieving correct sieving.

The aberrations of the setup, due to both the lenses along the optical path and the intrinsic SLM curvature, distort the wavefront of the diffracted beam, reducing its spatial coherence and thus the interference visibility. This is corrected by adding a global phase profile to the SLM mask, which is measured via the calibration method proposed in [156].

Fig. J.4(a) shows the recorded far field intensity image for the SLM holograms of Fig. J.3(b). To maximize the contrast between primes zeros and composites minima, each recorded frame is integrated over 600 pixels along q_y, generating an average intensity plot $I(q_x)$ [Fig. J.4(b)]. During this operation, we also subtract the background noise, recorded after applying to the SLM a hologram without the apertures. To cover the whole sieving range, we need to acquire $22000/40 = 550$ frames. When the intensity value $I(q_x)$ is lower than the threshold ϵ, q_x is labeled as a prime.

We find that an intensity threshold that is in between 4.8% and 5.9% of the global maximum allows us to correctly sieve all the primes (2436 numbers) and discard all the composites (19626 numbers) within the sieving range (Fig. J.6). This range quantifies the robustness of our scheme.

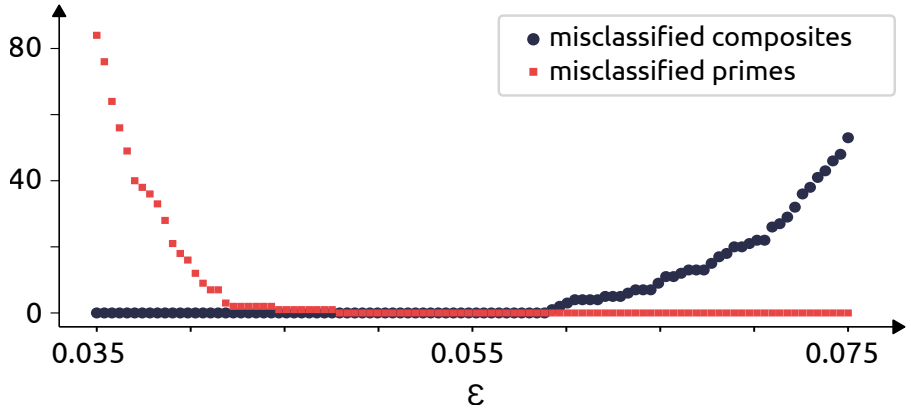


Figure J.6: Number of misclassified primes (red squares) and composites (blue dots) within the measured sieving range $139 < q_x < 22201$ as a function of the intensity threshold ϵ. By choosing $0.048 < \epsilon < 0.059$, the optical sieve correctly sieves all primes and discards composites.

Discussion

This work is the first experimental realization of an optical prime number sieve. By generating high resolution and on-demand phase masks via an SLM, we prove the effectiveness of the sieve by correctly classifying integers up to 22201 as either primes or composites.

The scalability of the sieving range, dictated by the number of apertures that can be fitted in a phase mask, is ultimately limited by the SLM spatial resolution. It is also limited by the SLM refresh rate, since a sequence of images has to be displayed in order to cover the whole sieving range. As SLM technology improves with respect to this two features, we expect better optical sieves to be implemented. The principally achievable range of the sieve can be estimated as the square of the SLM pixel resolution.

Note, however, that calculating the corrected displacement grating for each displacement d_x involves multiple modulo operations, which is equivalent, in terms of computational complexity, to checking the primality of d_x. Therefore our technique does not completely eliminate this type of calculations, but only reduces it by a factor equal to the increment value R. This value is limited by the width $2/w$ of the central lobe of the function $\mathrm{sinc}(wq_x)$, and hence cannot exceed the doubled SLM resolution.

Our method can be straightforwardly adapted to other kinds of sieves, such as Gaussian or twin primes sieves [217]. More generally, it demonstrates the potential of free-space linear optics in tasks that transcend imaging or communication, such as computation and solving mathematical problems. We can leverage the spatial coherence and superposition properties of light to achieve parallelism in computing.

Bibliography

[1] E. Abbe. “Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung”. *Archiv für Mikroskopische Anatomie* 9.1 (Dec. 1873), pp. 413–468.

[2] E. Abbe. “The Relation of Aperture and Power in the Microscope”. *Journal of the Royal Microscopical Society* 2.3 (June 1882), pp. 300–309.

[3] Rayleigh. “XXXI. Investigations in optics, with special reference to the spectroscope”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 8.49 (Oct. 1879), pp. 261–274.

[4] S. Ram, E. S. Ward, and R. J. Ober. “Beyond Rayleigh’s criterion: A resolution measure with application to single-molecule microscopy”. *Proceedings of the National Academy of Sciences* 103.12 (Mar. 2006), pp. 4457–4462.

[5] J. Chao, E. S. Ward, and R. J. Ober. “Fisher information theory for parameter estimation in single molecule microscopy: tutorial”. *J. Opt. Soc. Am. A* 33.7 (2016), B36–B57.

[6] M. Tsang, R. Nair, and X.-M. Lu. “Quantum Theory of Superresolution for Two Incoherent Optical Point Sources”. *Phys. Rev. X* 6 (3 2016), p. 031033.

[7] G. de Villiers and E. R. Pike. *The Limits of Resolution (Series in Optics and Optoelectronics)*. CRC Press, July 2016.

[8] S. W. Hell. “Far-Field Optical Nanoscopy”. *Science* 316.5828 (2007), pp. 1153–1158.

[9] A. Schropp et al. “Hard x-ray scanning microscopy with coherent radiation: Beyond the resolution of conventional x-ray microscopes”. *Applied Physics Letters* 100.25 (2012), p. 253112.

[10] U. Dürig, D. W. Pohl, and F. Rohner. “Near-field optical-scanning microscopy”. *Journal of Applied Physics* 59.10 (1986), pp. 3318–3327.

[11] V. Giovannetti, S. Lloyd, and L. Maccone. “Advances in quantum metrology”. *Nature Photonics* 5.4 (Mar. 2011), pp. 222–229.

[12] S. W. Hell and J. Wichmann. “Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy”. *Opt. Lett.* 19.11 (1994), pp. 780–782.

[13] E. Betzig et al. “Imaging Intracellular Fluorescent Proteins at Nanometer Resolution”. *Science* 313.5793 (2006), pp. 1642–1645.

[14] M. J. Rust, M. Bates, and X. Zhuang. "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)". *Nature Methods* 3.10 (Aug. 2006), pp. 793–796.

[15] R. M. Dickson et al. "On/off blinking and switching behaviour of single molecules of green fluorescent protein". *Nature* 388.6640 (July 1997), pp. 355–358.

[16] M. Tsang. "Quantum limits to optical point-source localization". *Optica* 2.7 (2015), pp. 646–653.

[17] R. Nair and M. Tsang. "Interferometric superlocalization of two incoherent optical point sources". *Opt. Express* 24.4 (2016), pp. 3684–3701.

[18] M. Tsang, R. Nair, and X.-M. Lu. "Quantum information for semiclassical optics". *Quantum and Nonlinear Optics IV*. Ed. by Q. Gong, G.-C. Guo, and B. S. Ham. Vol. 10029. International Society for Optics and Photonics. SPIE, 2016, pp. 6 –12.

[19] R. Nair and M. Tsang. "Far-Field Superresolution of Thermal Electromagnetic Sources at the Quantum Limit". *Phys. Rev. Lett.* 117 (19 2016), p. 190801.

[20] M. Tsang. "Conservative classical and quantum resolution limits for incoherent imaging". *Journal of Modern Optics* 65.11 (2018), pp. 1385–1391.

[21] S. Z. Ang, R. Nair, and M. Tsang. "Quantum limit for two-dimensional resolution of two incoherent optical point sources". *Phys. Rev. A* 95 (6 2017), p. 063847.

[22] M. Tsang. "Subdiffraction incoherent optical imaging via spatial-mode demultiplexing". *New Journal of Physics* 19.2 (2017), p. 023054.

[23] X.-M. Lu et al. "Quantum-optimal detection of one-versus-two incoherent optical sources with arbitrary separation". *npj Quantum Information* 4.1 (Dec. 2018).

[24] M. Tsang. "Subdiffraction incoherent optical imaging via spatial-mode demultiplexing: Semiclassical treatment". *Phys. Rev. A* 97 (2 2018), p. 023830.

[25] F. Yang et al. "Fisher information for far-field linear optical superresolution via homodyne or heterodyne detection in a higher-order local oscillator mode". *Phys. Rev. A* 96 (6 2017), p. 063829.

[26] M. Tsang. "Quantum limit to subdiffraction incoherent optical imaging". *Phys. Rev. A* 99 (1 2019), p. 012305.

[27] M. Tsang and R. Nair. "Resurgence of Rayleigh's curse in the presence of partial coherence: comment". *Optica* 6.4 (2019), pp. 400–401.

[28] M. Tsang. "Resolving starlight: a quantum perspective". *Contemporary Physics* 60.4 (Oct. 2019), pp. 279–298.

[29] M. Tsang. "Semiparametric estimation for incoherent optical imaging". *Phys. Rev. Research* 1 (3 2019), p. 033006.

[30] M. Tsang. "Quantum limit to subdiffraction incoherent optical imaging. II. A parametric-submodel approach". *Physical Review A* 104.5 (Nov. 2021).

[31] M. Tsang. *Efficient superoscillation measurement for incoherent optical imaging*. 2020. eprint: `arXiv:2010.11084`.

[32] M. Tsang. "Poisson Quantum Information". *Quantum* 5 (Aug. 2021), p. 527.

[33] C. Lupo and S. Pirandola. "Ultimate Precision Bound of Quantum and Subwavelength Imaging". *Phys. Rev. Lett.* 117 (19 2016), p. 190802.

[34] J. Rehacek et al. "Optimal measurements for resolution beyond the Rayleigh limit". *Opt. Lett.* 42.2 (2017), pp. 231–234.

[35] R. Kerviche, S. Guha, and A. Ashok. "Fundamental limit of resolving two point sources limited by an arbitrary point spread function". *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, June 2017.

[36] A. Chrostowski et al. "On super-resolution imaging as a multiparameter estimation problem". *International Journal of Quantum Information* 15.08 (2017), p. 1740005.

[37] J. Řehaček et al. "Multiparameter quantum metrology of incoherent point sources: Towards realistic superresolution". *Physical Review A* 96.6 (Dec. 2017).

[38] J. Řeháček et al. "Optimal measurements for quantum spatial superresolution". *Physical Review A* 98.1 (July 2018).

[39] M. P. Backlund, Y. Shechtman, and R. L. Walsworth. "Fundamental Precision Bounds for Three-Dimensional Optical Localization Microscopy with Poisson Statistics". *Phys. Rev. Lett.* 121 (2 2018), p. 023904.

[40] Z. Yu and S. Prasad. "Quantum Limited Superresolution of an Incoherent Source Pair in Three Dimensions". *Phys. Rev. Lett.* 121 (18 2018), p. 180504.

[41] W. Larson and B. E. A. Saleh. "Resurgence of Rayleigh's curse in the presence of partial coherence". *Optica* 5.11 (2018), pp. 1382–1389.

[42] S. Zhou and L. Jiang. "Modern description of Rayleigh's criterion". *Phys. Rev. A* 99 (1 2019), p. 013808.

[43] S. Prasad and Z. Yu. "Quantum-limited superlocalization and superresolution of a source pair in three dimensions". *Physical Review A* 99.2 (Feb. 2019).

[44] Z. Dutton et al. "Attaining the quantum limit of superresolution in imaging an object's length via predetection spatial-mode sorting". *Phys. Rev. A* 99 (3 2019), p. 033847.

[45] W. Larson and B. E. A. Saleh. "Resurgence of Rayleigh's curse in the presence of partial coherence: reply". *Optica* 6.4 (2019), pp. 402–403.

[46] C. Napoli et al. "Towards Superresolution Surface Metrology: Quantum Estimation of Angular and Axial Separations". *Phys. Rev. Lett.* 122 (14 2019), p. 140505.

[47] K. A. G. Bonsma-Fisher et al. "Realistic sub-Rayleigh imaging with phase-sensitive measurements". *New Journal of Physics* 21.9 (Sept. 2019), p. 093010.

[48] E. Bisketzi, D. Branford, and A. Datta. "Quantum limits of localisation microscopy". *New Journal of Physics* 21.12 (Dec. 2019), p. 123032.

[49] M. R. Grace et al. "Approaching quantum-limited imaging resolution without prior knowledge of the object location". *Journal of the Optical Society of America A* 37.8 (July 2020), p. 1288.

[50] S. Prasad. "Quantum limited super-resolution of an unequal-brightness source pair in three dimensions". *Physica Scripta* 95.5 (Mar. 2020), p. 054004.

[51] C. Lupo, Z. Huang, and P. Kok. "Quantum Limits to Incoherent Imaging are Achieved by Linear Interferometry". *Phys. Rev. Lett.* 124 (8 2020), p. 080503.

[52] T. Gefen, A. Rotem, and A. Retzker. "Overcoming resolution limits with quantum sensing". *Nature Communications* 10.1 (Nov. 2019).

[53] Z. Hradil et al. "Quantum Fisher information with coherence". *Optica* 6.11 (2019), pp. 1437–1440.

[54] Y. L. Len et al. "Resolution limits of spatial mode demultiplexing with noisy detection". *International Journal of Quantum Information* (Jan. 2020), p. 1941015.

[55] C. Lupo. "Subwavelength quantum imaging with noisy detectors". *Phys. Rev. A* 101 (2 2020), p. 022323.

[56] J. O. de Almeida et al. "Discrimination and estimation of incoherent sources under misalignment". *Physical Review A* 103.2 (Feb. 2021).

[57] O. Zhang and M. D. Lew. "Quantum limits for precisely estimating the orientation and wobble of dipole emitters". *Physical Review Research* 2.3 (July 2020).

[58] M. Gessner, C. Fabre, and N. Treps. "Superresolution Limits from Measurement Crosstalk". *Physical Review Letters* 125.10 (Aug. 2020).

[59] C. Datta et al. "Sub-Rayleigh resolution of two incoherent sources by array homodyning". *Physical Review A* 102.6 (Dec. 2020).

[60] S. Prasad. "Quantum limited source localization and pair superresolution in two dimensions under finite-emission bandwidth". *Physical Review A* 102.3 (Sept. 2020).

[61] S. Prasad. "Quantum limited superresolution of extended sources in one and two dimensions". *Physical Review A* 102.6 (Dec. 2020).

[62] C. Oh et al. "Quantum Limits of Superresolution in a Noisy Environment". *Physical Review Letters* 126.12 (Mar. 2021).

[63] O. Zhang and M. D. Lew. "Single-molecule orientation localization microscopy I: fundamental limits". *Journal of the Optical Society of America A* 38.2 (Feb. 2021), p. 277.

[64] O. Zhang and M. D. Lew. "Single-molecule orientation localization microscopy II: a performance comparison". *Journal of the Optical Society of America A* 38.2 (Feb. 2021), p. 288.

[65] Z. Huang, C. Lupo, and P. Kok. "Quantum-Limited Estimation of Range and Velocity". *PRX Quantum* 2.3 (July 2021).

[66] L. Peng and X.-M. Lu. "Generalization of Rayleigh's criterion on parameter estimation with incoherent sources". *Physical Review A* 103.4 (Apr. 2021).

[67] L. J. Fiderer et al. "General Expressions for the Quantum Fisher Information Matrix with Applications to Discrete Quantum Imaging". *PRX Quantum* 2.2 (Apr. 2021).

[68] B. Wang et al. "Quantum-limited localization and resolution in three dimensions". *Photonics Research* 9.8 (July 2021), p. 1522.

[69] K. Liang, S. A. Wadood, and A. N. Vamivakas. "Coherence effects on estimating two-point separation". *Optica* 8.2 (Feb. 2021), p. 243.

[70] Y. Wang, Y. Zhang, and V. O. Lorenz. "Superresolution in interferometric imaging of strong thermal sources". *Physical Review A* 104.2 (Aug. 2021).

[71] K. K. M. Bearne et al. "Confocal super-resolution microscopy based on a spatial mode sorter". *Optics Express* 29.8 (Mar. 2021), p. 11784.

[72] A. Sajjad et al. "Attaining quantum limited precision of localizing an object in passive imaging". *Physical Review A* 104.2 (Aug. 2021).

[73] G. Sorelli et al. "Optimal Observables and Estimators for Practical Super-resolution Imaging". 127.12 (Sept. 2021).

[74] D Koutný et al. "Axial superlocalization with vortex beams". *Quantum Science and Technology* 6.2 (Mar. 2021), p. 025021.

[75] S. Kurdzialek. *Back to sources – the role of coherence in super-resolution imaging revisited*. 2021. eprint: `arXiv:2103.12096`.

[76] Z. Hradil, D. Koutný, and J. Řeháček. "Exploring the ultimate limits: super-resolution enhanced by partial coherence". *Optics Letters* 46.7 (Mar. 2021), p. 1728.

[77] M. Shah and L. Fan. "Frequency Superresolution with Spectrotemporal Shaping of Photons". *Physical Review Applied* 15.3 (Mar. 2021).

[78] F. Bao et al. "Quantum-accelerated imaging of N stars". *Optics Letters* 46.13 (June 2021), p. 3045.

[79] J. Xin, Y. Li, and X.-M. Lu. "Optical super-resolution for two unequally bright point sources based on the fractional Hilbert transform". *Physical Review A* 103.5 (May 2021).

[80] K. Liang, S. A. Wadood, and A. N. Vamivakas. "Coherence effects on estimating general sub-Rayleigh object distribution moments". *Physical Review A* 104.2 (Aug. 2021).

[81] G. Sorelli et al. "Moment-based superresolution: Formalism and applications". *Physical Review A* 104.3 (Sept. 2021).

[82] Z. Huang and C. Lupo. "Quantum Hypothesis Testing for Exoplanet Detection". *Physical Review Letters* 127.13 (Sept. 2021).

[83] C. Datta et al. "Sub-Rayleigh characterization of a binary source by spatially demultiplexed coherent detection". *Optics Express* 29.22 (Oct. 2021), p. 35592.

[84] J. Dong et al. "Fundamental bounds on the precision of iSCAT, COBRI and dark-field microscopy for 3D localization and mass photometry". *Journal of Physics D: Applied Physics* 54.39 (July 2021), p. 394002.

[85] E. F. Matlin and L. J. Zipp. *Adaptive Imaging of Arbitrary Thermal Source Distributions with Near Quantum-Limited Resolution*. 2021. eprint: `arXiv:2106.13332`.

[86] M. R. Grace and S. Guha. *Quantum-Optimal Object Discrimination in Sub-Diffraction Incoherent Imaging*. 2021. eprint: `arXiv:2107.00673`.

[87] Z. S. Tang, K. Durak, and A. Ling. "Fault-tolerant and finite-error localization for point emitters within the diffraction limit". *Opt. Express* 24.19 (2016), pp. 22004–22012.

[88] M. Paúr et al. "Achieving the ultimate optical resolution". *Optica* 3.10 (2016), pp. 1144–1147.

[89] F. Yang et al. "Far-field linear optical superresolution via heterodyne detection in a higher-order local oscillator mode". *Optica* 3.10 (2016), pp. 1148–1152.

[90] W.-K. Tham, H. Ferretti, and A. M. Steinberg. "Beating Rayleigh's Curse by Imaging Using Phase Information". *Phys. Rev. Lett.* 118 (7 2017), p. 070801.

[91] J. M. Donohue et al. "Quantum-Limited Time-Frequency Estimation through Mode-Selective Photon Measurement". *Phys. Rev. Lett.* 121 (9 2018), p. 090501.

[92] M. Parniak et al. "Beating the Rayleigh Limit Using Two-Photon Interference". *Phys. Rev. Lett.* 121 (25 2018), p. 250503.

[93] M. Paúr et al. "Tempering Rayleigh's curse with PSF shaping". *Optica* 5.10 (2018), pp. 1177–1180.

[94] J. Hassett et al. "Sub-Rayleigh Limit Localization with a Spatial Mode Analyzer". *Frontiers in Optics / Laser Science*. OSA, 2018.

[95] Y. Zhou et al. "Quantum-limited estimation of the axial separation of two incoherent point sources". *Optica* 6.5 (2019), pp. 534–541.

[96] M. Paúr et al. "Reading out Fisher information from the zeros of the point spread function". *Opt. Lett.* 44.12 (2019), pp. 3114–3117.

[97] . "Intensity-Based Axial Localization at the Quantum Limit". *Phys. Rev. Lett.* 123 (19 2019), p. 193601.

[98] M. Salit, J. Klein, and L. Lust. "Experimental characterization of a mode-separating photonic lantern for imaging applications". *Applied Optics* 59.17 (19 June 2020), p. 5319.

[99] H. Zhang, S. Kumar, and Y.-P. Huang. "Super-resolution optical classifier with high photon efficiency". *Optics Letters* 45.18 (19 Sept. 2020), p. 4968.

[100] P. Boucher et al. "Spatial optical mode demultiplexing as a practical tool for optimal transverse distance estimation". *Optica* 7.11 (19 Nov. 2020), p. 1621.

[101] V. Ansari et al. "Achieving the Ultimate Quantum Timing Resolution". *PRX Quantum* 2.1 (19 Jan. 2021), p. 193601.

[102] S. A. Wadood et al. "Experimental demonstration of superresolution of partially coherent light sources using parity sorting". *Optics Express* 29.14 (19 June 2021), p. 22034.

[103] S. L. Mouradian et al. "Quantum sensing of intermittent stochastic signals". *Physical Review A* 103.3 (19 Mar. 2021), p. 193601.

[104] S. De et al. "Effects of coherence on temporal resolution". *Physical Review Research* 3.3 (19 July 2021), p. 193601.

[105] K. Santra et al. "Localization of Nonblinking Point Sources Using Higher-Order-Mode Detection and Optical Heterodyning: Developing a Strategy for Extending the Scope of Molecular, Super-resolution Imaging". *The Journal of Physical Chemistry B* 125.12 (19 Mar. 2021), pp. 3092–3104.

[106] W. Larson and B. E. A. Saleh. "Two-lens anisotropic image-inversion system for interferometric information processing". *Optics Express* 29.10 (19 May 2021), p. 15403.

[107] M. Mazelanik, A. Leszczyński, and M. Parniak. *Optical-domain spectral super-resolution enabled by a quantum memory*. 2021. eprint: `arXiv:2106.04450`.

[108] J. I. Costa-Filho and A. Lvovsky. *Quantum-inspired super-resolution of incoherent point sources*. 2021.

[109] J. I. Costa-Filho and A. Lvovsky. *Quantum precision bounds of coherent light imaging*. 2021.

[110] A. A. Pushkina et al. "Comprehensive model and performance optimization of phase-only spatial light modulators". *Measurement Science and Technology* 31.12 (19 2020), p. 125202.

[111] A. A. Pushkina et al. *Super-resolution linear optical imaging in the far field*. 2021. eprint: `arXiv:2105.01743`.

[112] C. Helstrom. *Quantum detection and estimation theory*. Vol. 123. New York: Academic Press, 1976, p. 193601.

[113] A. Holevo. *Probabilistic and statistical aspects of quantum theory*. Vol. 123. Pisa Italy: Edizioni della Normale, 2011, p. 193601.

[114] S. L. Braunstein and C. M. Caves. "Statistical distance and the geometry of quantum states". *Physical Review Letters* 72.22 (19 May 1994), pp. 3439–3443.

[115] M. G. A. PARIS. "Quantum estimation for quantum technology". *International Journal of Quantum Information* 07 (19 2009), pp. 125–137.

[116] G. Tóth and I. Apellaniz. "Quantum metrology from a quantum information science perspective". *Journal of Physics A: Mathematical and Theoretical* 47.42 (19 Oct. 2014), p. 424006.

[117] E. L. Lehmann. *Theory of point estimation*. Vol. 123. New York: Springer, 1998, p. 193601.

[118] H. Trees. *Detection, estimation, and modulation theory*. Vol. 123. New York Chichester: Wiley, 2001, p. 193601.

[119] E. Kreyszig. *Introductory functional analysis with applications*. Vol. 123. New York: Wiley, 1978, p. 193601.

[120] Y. C. Eldar. "Rethinking Biased Estimation: Improving Maximum Likelihood and the Cramér–Rao Bound". *Foundations and Trends in Signal Processing* 1.4 (19 2007), pp. 305–449.

[121] S. ichi Amari. "Information geometry". *Japanese Journal of Mathematics* 16.1 (19 Jan. 2021), pp. 1–48.

[122] J. Rubio and J. Dunningham. "Quantum metrology in the presence of limited data". *New Journal of Physics* 21.4 (19 Apr. 2019), p. 043037.

[123] J. Liu et al. "Quantum Fisher information matrix and multiparameter estimation". *Journal of Physics A: Mathematical and Theoretical* 53.2 (19 Dec. 2019), p. 023001.

[124] L Mirsky. *An introduction to linear algebra*. Vol. 123. New York: Dover, 1990, p. 193601.

[125] J. S. Sidhu and P. Kok. "Geometric perspective on quantum parameter estimation". *AVS Quantum Science* 2.1 (19 Feb. 2020), p. 014701.

[126] K. C. Tan and H. Jeong. "Nonclassical light and metrological power: An introductory review". *AVS Quantum Science* 1.1 (19 Dec. 2019), p. 014701.

[127] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański. "Compatibility in multiparameter quantum metrology". *Physical Review A* 94.5 (19 Nov. 2016), p. 193601.

[128] R. Demkowicz-Dobrzański, W. Górecki, and M. Guţă. "Multi-parameter estimation beyond quantum Fisher information". *Journal of Physics A: Mathematical and Theoretical* 53.36 (19 Aug. 2020), p. 363001.

[129] J. Yang et al. "Optimal measurements for quantum multiparameter estimation with general states". *Physical Review A* 100.3 (19 Sept. 2019), p. 193601.

[130] H. M. Wiseman. "Adaptive Phase Measurements of Optical Modes: Going Beyond the MarginalQDistribution". *Physical Review Letters* 75.25 (19 Dec. 1995), pp. 4587–4590.

[131] E. Polino et al. "Photonic quantum metrology". *AVS Quantum Science* 2.2 (19 June 2020), p. 024703.

[132] O. E. Barndorff-Nielsen and R. D. Gill. "Fisher information in quantum statistics". *Journal of Physics A: Mathematical and General* 33.24 (19 June 2000), pp. 4481–4490.

[133] A. Hentschel and B. C. Sanders. "Machine Learning for Precise Quantum Measurement". *Physical Review Letters* 104.6 (19 Feb. 2010), p. 193601.

[134] A. Lumino et al. "Experimental Phase Estimation Enhanced by Machine Learning". *Physical Review Applied* 10.4 (19 Oct. 2018), p. 193601.

[135] M. G. Genoni and T. Tufarelli. "Non-orthogonal bases for quantum metrology". *Journal of Physics A: Mathematical and Theoretical* 52.43 (19 Oct. 2019), p. 434002.

[136] D. Šafránek. "Simple expression for the quantum Fisher information matrix". *Physical Review A* 97.4 (19 Apr. 2018), p. 193601.

[137] M. Nielsen. *Quantum computation and quantum information*. Vol. 123. Cambridge New York: Cambridge University Press, 2010, p. 193601.

[138] H. Yuan and C.-H. F. Fung. "Quantum metrology matrix". *Physical Review A* 96.1 (19 July 2017), p. 193601.

[139] M. Born. *Principles of optics : electromagnetic theory of propagation, interference and diffraction of light*. Vol. 123. Cambridge New York: Cambridge University Press, 1999, p. 193601.

[140] J. H. Shapiro. "The Quantum Theory of Optical Communications". *IEEE Journal of Selected Topics in Quantum Electronics* 15.6 (19 2009), pp. 1547–1569.

[141] J. Goodman. *Introduction to Fourier optics*. Vol. 123. Englewood, Colo: Roberts & Co, 2005, p. 193601.

[142] J. Goodman. *Statistical optics*. Vol. 123. Hoboken, New Jersey: John Wiley & Sons Inc, 2015, p. 193601.

[143] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin. "Gaussian approximations of fluorescence microscope point-spread function models". *Appl. Opt.* 46.10 (19 2007), pp. 1819–1829.

[144] F. Zernike. "The concept of degree of coherence and its application to optical problems". *Physica* 5.8 (19 Aug. 1938), pp. 785–795.

[145] P. van Cittert. "Die Wahrscheinliche Schwingungsverteilung in Einer von Einer Lichtquelle Direkt Oder Mittels Einer Linse Beleuchteten Ebene". *Physica* 1.1-6 (19 Jan. 1934), pp. 201–210.

[146] H Davis. *Linear algebra and linear operators in engineering : with applications in Mathematica*. Vol. 123. San Diego: Academic Press, 2000, p. 193601.

[147] U. Leonhardt. *Measuring the quantum state of light*. Vol. 123. Cambridge, UK New York: Cambridge University Press, 1997, p. 193601.

[148] S. Gull and J. Skilling. "Maximum entropy method in image processing". *IEE Proceedings F Communications, Radar and Signal Processing* 131.6 (19 1984), p. 646.

[149] M. Bertero. "Linear Inverse and III-Posed Problems". *Advances in Electronics and Electron Physics Volume 75*. Vol. 123. Elsevier, 1989, pp. 1–120.

[150] P. Milanfar, W. Karl, and A. Willsky. "A moment-based variational approach to tomographic reconstruction". *IEEE Transactions on Image Processing* 5.3 (19 1996), pp. 459–470.

[151] Y. de Castro and F. Gamboa. "Exact reconstruction using Beurling minimal extrapolation". *Journal of Mathematical Analysis and Applications* 395.1 (19 Nov. 2012), pp. 336–354.

[152] L. Zhu and J. Wang. "Arbitrary manipulation of spatial amplitude and phase using phase-only spatial light modulators". *Scientific reports* 4 (19 2014), p. 7441.

[153] E. Bolduc et al. "Exact solution to simultaneous intensity and phase encryption with a single phase-only hologram". *Optics letters* 38.18 (19 2013), pp. 3546–3549.

[154] G. Lazarev et al. "Beyond the display: Phase-only liquid crystal on Silicon devices and their applications in photonics". *Optics express* 27.11 (19 2019), pp. 16206–16249.

[155] J. L. Harriman, A. Linnenberger, and S. A. Serati. "Improving spatial light modulator performance through phase compensation". *Advanced Wavefront Control: Methods, Devices, and Applications II*. Vol. 5553. International Society for Optics and Photonics. American Physical Society, 2004, pp. 58–67.

[156] T. Čižmár, M. Mazilu, and K. Dholakia. "In situ wavefront correction and its application to micromanipulation". *Nature Photonics* 4.6 (19 2010), p. 388.

[157] Z. Zhang et al. "Diffraction based phase compensation method for phase-only liquid crystal on silicon devices in operation". *Applied optics* 51.17 (19 2012), pp. 3837–3846.

[158] X. Xun and R. W. Cohn. "Phase calibration of spatially nonuniform spatial light modulators". *Applied optics* 43.35 (19 2004), pp. 6400–6406.

[159] J. L. Martínez et al. "Analysis of multiple internal reflections in a parallel aligned liquid crystal on silicon SLM". *Optics express* 22.21 (19 2014), pp. 25866–25879.

[160] J. H. Kulick et al. "Electrostatic and diffraction analysis of a liquid-crystal device utilizing fringing fields: applications to three-dimensional displays". *Applied optics* 34.11 (19 1995), pp. 1901–1922.

[161] B. Apter, U. Efron, and E. Bahat-Treidel. "On the fringing-field effect in liquid-crystal beam-steering devices". *Applied optics* 43.1 (19 2004), pp. 11–19.

[162] M. Persson, D. Engström, and M. Goksör. "Reducing the effect of pixel crosstalk in phase only spatial light modulators". *Optics express* 20.20 (19 2012), pp. 22334–22343.

[163] P. Gemayel et al. "Cross-talk compensation of a spatial light modulator for iterative phase retrieval applications". *Applied optics* 55.4 (19 2016), pp. 802–810.

[164] E. Hällstig et al. "Fringing fields in a liquid crystal spatial light modulator for beam steering". *Journal of modern optics* 51.8 (19 2004), pp. 1233–1247.

[165] M. Carbonell-Leal and O. Mendoza-Yero. "Effects of mitigation of pixel cross-talk in the encoding of complex fields using the double-phase method". *Optical Engineering* 59.04 (19 Nov. 2019), p. 1.

[166] S. Moser, M. Ritsch-Marte, and G. Thalhammer. "Model-based compensation of pixel crosstalk in liquid crystal spatial light modulators". *Optics express* 27.18 (19 2019), pp. 25046–25063.

[167] A. Pushkina et al. *Python program to generate phase-only SLM holograms*. 2020.

[168] A. Lvovsky. "Iterative maximum-likelihood reconstruction in quantum homodyne tomography". *Journal of Optics B: Quantum and Semiclassical Optics* 6.6 (19 2004), S556.

[169] R. Kumar et al. "Versatile wideband balanced detector for quantum optical homodyne tomography". *Optics Communications* 285.24 (19 2012), pp. 5259 –5267.

[170] A. Masalov, A Kuzhamuratov, and A. Lvovsky. "Noise spectra in balanced optical detectors based on transimpedance amplifiers". *Review of Scientific Instruments* 88.11 (19 2017), p. 113109.

[171] G. Maltese et al. *Hermite-Gaussian Microscopy Github repository*. `https://github.com/giomalt/HGM`. 2021.

[172] A. M. Palmieri et al. "Experimental neural network enhanced quantum tomography". *npj Quantum Information* 6.1 (19 Feb. 2020), p. 193601.

[173] L. B. Lucy. "An iterative technique for the rectification of observed distributions". *The Astronomical Journal* 79 (19 June 1974), p. 745.

[174] A. D. Corbett et al. "Microscope calibration using laser written fluorescence". *Optics express* 26.17 (19 2018), pp. 21887–21899.

[175] C. Shorten and T. M. Khoshgoftaar. "A survey on image data augmentation for deep learning". *Journal of Big Data* 6.1 (19 2019), pp. 1–48.

[176] M. P. Edgar, G. M. Gibson, and M. J. Padgett. "Principles and prospects for single-pixel imaging". en. *Nat. Photonics* 13.1 (19 Jan. 2019), pp. 13–20.

[177] G. M. Gibson, S. D. Johnson, and M. J. Padgett. "Single-pixel imaging 12 years on: a review". en. *Opt. Express* 28.19 (19 Sept. 2020), pp. 28190–28208.

[178] A. Hentschel and B. C. Sanders. "Machine Learning for Adaptive Quantum Measurement". *2010 Seventh International Conference on Information Technology: New Generations*. Vol. 123. IEEE, 2010, p. 193601.

[179] Y. Peng and H. Fan. "Feedback ansatz for adaptive-feedback quantum metrology training with machine learning". *Physical Review A* 101.2 (19 Feb. 2020), p. 193601.

[180] Y. Ban et al. "Neural-network-based parameter estimation for quantum detection". *Quantum Science and Technology* 6.4 (19 Aug. 2021), p. 045012.

[181] E. Moen et al. "Deep learning for cellular image analysis". *Nature Methods* 16.12 (19 May 2019), pp. 1233–1246.

[182] L. von Chamier, R. F. Laine, and R. Henriques. "Artificial intelligence for microscopy: what you should know". *Biochemical Society Transactions* 47.4 (19 July 2019), pp. 1029–1040.

[183] M. Weigert et al. "Content-aware image restoration: pushing the limits of fluorescence microscopy". *Nature Methods* 15.12 (19 Nov. 2018), pp. 1090–1097.

[184] H. Wang et al. "Deep learning enables cross-modality super-resolution in fluorescence microscopy". *Nature Methods* 16.1 (19 Dec. 2018), pp. 103–110.

[185] A. Durand et al. "A machine learning approach for online automated optimization of super-resolution optical microscopy". *Nature Communications* 9.1 (19 Dec. 2018), p. 193601.

[186] M. G. Haberl et al. "CDeep3M—Plug-and-Play cloud-based deep learning for image segmentation". *Nature Methods* 15.9 (19 Aug. 2018), pp. 677–680.

[187] L. Ballentine. *Quantum mechanics : a modern development*. Vol. 123. Singapore River Edge, NJ: World Scientific, 2000, p. 193601.

[188] J. Watrous. *The theory of quantum information*. Vol. 123. Cambridge, United Kingdom: Cambridge University Press, 2018, p. 193601.

[189] L. Mandel and E. Wolf. *Optical Coherence and Quantum Optics*. Vol. 123. Cambridge University Press, 1995, p. 193601.

[190] R. Loudon. *The quantum theory of light*. Vol. 123. Oxford New York: Oxford University Press, 2000, p. 193601.

[191] R. J. Glauber. "Coherent and Incoherent States of the Radiation Field". *Phys. Rev.* 131 (6 1963), pp. 2766–2788.

[192] E. C. G. Sudarshan. "Equivalence of Semiclassical and Quantum Mechanical Descriptions of Statistical Light Beams". *Phys. Rev. Lett.* 10 (7 1963), pp. 277–279.

[193] N. K. Fontaine et al. "Scalable mode sorter supporting 210 Hermite-Gaussian modes". *Optical Fiber Communication Conference Postdeadline Papers*. Vol. 123. Optical Society of America, 2018, Th4B.4.

[194] L. Martin et al. "Basis-neutral Hilbert-space analyzers". *Scientific Reports* 7.1 (19 Mar. 2017), p. 193601.

[195] Y. Zhou et al. "Hermite-Gaussian mode sorter". *Opt. Lett.* 43.21 (19 2018), pp. 5263–5266.

[196] M. Hiekkamäki, S. Prabhakar, and R. Fickler. "Near-perfect measuring of full-field transverse-spatial modes of light". *Opt. Express* 27.22 (19 2019), pp. 31456–31464.

[197] N. K. Fontaine et al. "Laguerre-Gaussian mode sorter". *Nature Communications* 10.1 (19 Apr. 2019), p. 193601.

[198] G. Szego. *Orthogonal polynomials*. Vol. 123. New York city: American mathematical Society, 1939, p. 193601.

[199] B. Li et al. "Optical Eratosthenes' sieve for large prime numbers". *Opt. Express* 28.8 (19 2020), pp. 11965–11973.

[200] A. Ingham. *The distribution of prime numbers*. Vol. 123. 30. Cambridge University Press, 1990, p. 193601.

[201] H. M. Edwards. *Riemann's Zeta Function*. Vol. 123. Dover Publications, 2001, p. 193601.

[202] R. L. Rivest, A. Shamir, and L. Adleman. "A method for obtaining digital signatures and public-key cryptosystems". *Communications of the ACM* 21.2 (19 Feb. 1978), pp. 120–126.

[203] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone. *Handbook of Applied Cryptography (Discrete Mathematics and Its Applications)*. Vol. 123. CRC Press, 1996, p. 193601.

[204] P. Shor. "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer". *SIAM J. Comput.* 26.5 (19 1997), pp. 1484–1509.

[205] D. Schumayer and D. Hutchinson. "Colloquium: Physics of the Riemann hypothesis". *Rev. Mod. Phys.* 83.2 (19 2011), p. 307.

[206] C. Bender, D. Brody, and M. Müller. "Hamiltonian for the zeros of the Riemann zeta function". *Phys. Rev. Lett.* 118.13 (19 2017), p. 130201.

[207] L. A. Bunimovich and C. P. Dettmann. "Open Circular Billiards and the Riemann Hypothesis". *Phys. Rev. Lett.* 94 (10 2005), p. 100201.

[208] M. V. Berry and J. P. Keating. "The Riemann Zeros and Eigenvalue Asymptotics". *SIAM Review* 41.2 (19 Jan. 1999), pp. 236–266.

[209] A. A. Rangelov. "Factorizing numbers with classical interference: several implementations in optics". *Journal of Physics B: Atomic, Molecular and Optical Physics* 42.2 (19 Jan. 2009), p. 021002.

[210] J. Clauser and J. Dowling. "Factoring integers with Young's N-slit interferometer". *Phys. Rev. A* 53.6 (19 1996), p. 4587.

[211] A. Shamir. "Factoring large numbers with the TWINKLE device". *International Workshop on Cryptographic Hardware and Embedded Systems*. Vol. 123. Springer. American Physical Society, 1999, pp. 2–12.

[212] D. Bigourd et al. "Factorization of Numbers with the Temporal Talbot Effect: Optical Implementation by a Sequence of Shaped Ultrashort Pulses". *Phys. Rev. Lett.* 100 (3 2008), p. 030202.

[213] K. Pelka et al. “Prime number decomposition using the Talbot effect”. *Opt. Express* 26.12 (19 2018), pp. 15009–15014.

[214] V. Tamma et al. “Factoring numbers with a single interferogram”. *Phys. Rev. A* 83.2 (19 2011), p. 020304.

[215] M. Berry. “Riemann zeros in radiation patterns”. *J. Phys. A* 45.30 (19 2012), p. 302001.

[216] M. Berry. “Riemann zeros in radiation patterns: II. Fourier transforms of zeta”. *J. Phys. A* 48.38 (19 2015), p. 385203.

[217] T. Petersen et al. “Simple wave-optical superpositions as prime number sieves”. *Phys. Rev. Lett.* 122.9 (19 2019), p. 090201.

[218] R. Crandall and C. Pomerance. *Prime numbers: a computational perspective*. Vol. 182. Springer Science & Business Media, 2006, p. 193601.