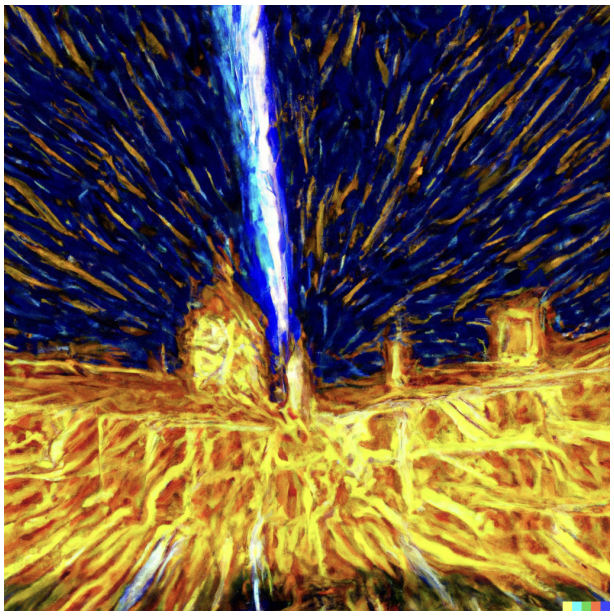
Single-shot Spatio-temporal Vector Field Measurements of Petawatt Laser Pulses



Sunny Howard

Wadham College

University of Oxford

This dissertation is submitted for the degree of

Doctor of Philosophy in Atomic and Laser Physics

Hilary Term

A second passes.

A second, second.

A third.

But they’re three seconds only for us.

— *Wisława Szymborska, “View with a Grain of Sand"*

Abstract

The field structure of ultra-intense laser pulses plays a critical role in their applications. Spatiotemporal couplings (STCs), though undetectable by conventional diagnostics, can significantly impact the pulse's properties. For example, undesirable STCs reduce the peak focused intensity, while deliberate STCs, such as those in the "flying focus", underpin advancements in structured light. Furthermore, as evidenced by a growing body of simulation-based literature, the polarisation state of the pulse can also add a valuable dimension of control and flexibility. A key barrier to the experimental realisation of structuring ultra-intense light is the fact that no method currently exists to characterise the spatiotemporal vector field of individual ultra-intense pulses. This thesis presents the development and realisation of a robust method for the single-shot characterisation of the spatiotemporal vector field, which also provides uncertainty estimates. Its efficacy is demonstrated by characterising the ATLAS-3000 PW laser and vectorial pulses such as circularly polarised optical vortices. It is hoped that the technique will be pivotal for advancing structured light applications in ultra-intense laser physics.

Table of contents

1 Introduction **6**

1.1 Structure of Thesis 12

1.2 List of Selected Peer Reviewed Publications 13

2 Theory **14**

2.1 Light 14

2.2 Numerical Propagation of Light 19

2.3 Ultra-Intense Laser Pulses 26

2.4 Ultra-Intense Laser Metrology 32

2.5 Inverse problems 62

2.6 Machine Learning 67

3 Hyperspectral Compressive Wavefront Sensing **78**

3.1 Wavefront Sensing 79

3.2 Snapshot Compressive Imaging 82

3.3 Network Architecture 88

3.4 Simulation Results 91

3.5 Multi View Deep Unrolling 94

3.6 Experimental Implementation of Diagnostic 98

3.7 Summary 104

4 Image Deconvolution and Over-complete Dictionaries 105

4.1 Image Denoising 106

4.2 Wavefront Reconstruction Using an Over-complete Dictionary 125

5 Single-shot Vector Field Measurement of Ultra-Intense Laser Pulses 142

5.1 Technique 145

5.2 Measurements 164

5.3 Discussion 182

6 Conclusion 185

References 188

Appendix A Sparse Optimization 206

A.1 Equivalency of the Minimization of the L0 and L1 Norm 206

A.2 LASSO Regression 208

Appendix B Energy Distribution in Airy Rings 209

Appendix C Additional details of the RAVEN Technique 213

C.1 Unique Encoding of Wavefront and Spectrum 213

C.2 Simulation of Polarization effects 215

Appendix D Pictures from RAVEN Experiments 217

Acknowledgements

Well, it is possible that few people will ever read this thesis, but I can at least use it as an opportunity to say thank you. Throughout the last years I have been constantly amazed by the generosity of people, and their willingness to go further than they would ever be expected to help. I feel so blessed for all the people in my life, of which there are sadly too many to list. So, thank you to;

Peter, for taking a chance on me when I had already quit one PhD, and providing constant guidance and warmth. Thank you, Andreas, for never failing to inspire me with your brilliant ideas and just always being there. Not only have you been the best supervisors imaginable, but I am happy to leave with two true friends.

Everyone at Living Optics; especially Robin Wang and Steve Chappell - it would've been simply impossible to get through these years without your support. Also to the DAAD, whose scholarship allowed me to have an amazing year in Germany, and for generally promoting the message of bringing people together. Thanks to all the lab partners from both of my groups, and especially to the Germans (Jannik, Jakob, Nils, Chris and all the others) who made me feel so welcome. It was always fun being around you all, whether in the lab or elsewhere! I also owe a lot to the kindness of my previous PhD supervisor, Giuliana.

Thanks to my parents, who have given me the ultimate gift of unconditional love, and always allow me to be whatever I wish. To my brothers, who are two of my best friends and are both much smarter than me (meaning they are often the first people I will come to with problems). And to my grandparents, who I love very much.

My friends, who also feel like family; here it really is impossible to add everyone that I would like to. To list some, those who I knew from before, Yann, Marti, Luke, Matt, Zeynep, Rosh, Zosh, Gabby, Mike, Jana and James, and some that I met along the way; Amey, Andrew, Ed, Ana, everyone from Buddhist Society, Zali, Vreni, Andy... I am 100% sure that I have missed many people, but I hope you know how much I appreciate you all.

Last, but certainly not least, thanks to Caroline, Kathy, Hugo and Roger for all the love, and for teaching me the things that I consider the most important.

Abbreviations

Adam	Adaptive Moment Estimation
AdaGrad	Adaptive Gradient Updates
ADMM	Alternating Direction Method of Multipliers
AOPDF	Acousto-Optic Dispersive Filter
BELLA	Berkeley Lab Laser Accelerator
BM	Birefringent Medium
CALA	Center for Advanced Laser Applications
CASSI	Coded Aperture Snapshot Compressive Imaging
CNN	Convolutional Neural Network
CPA	Chirped Pulse Amplification
DG	Diffraction Grating
FALCON	Fast Acquisition of Laser Couplings using Narrowband Filters
FCL	Fully Connected Layer
FROG	Frequency-Resolved Optical Gating
FTS	Fourier Transform Spectroscopy
GPU	Graphics Processing Unit
GS	Gerchberg-Saxton
HWP	Half-Wave Plate
ICF	Inertial Confinement Fusion
IPM	Interior Point Methods

ISTA	Iterative Shrinkage-Thresholding Algorithms
J-KAREN-P	Japan-Kansai Advanced Relativistic Engineering
LASSO	Least Absolute Shrinkage and Selection Operator
LC	Liquid Crystal
LCL	Locally-Connected Layer
LCN	Locally-Connected Network
LHC	Large Hadron Collider
LINAC	Linear Accelerator
LMU	Ludwig-Maximilians-Universität
LSI	Lateral Shearing Interferometry
LWFA	Laser Wakefield Accelerator
MLA	Microlens Array
MSE	Mean Squared Error
NIF	National Ignition Facility
NLL	Negative Log Likelihood
OAM	Orbital Angular Momentum
PA	Paraxial Approximation
PAC	Pixel-Adaptive Convolution
PFA	Polarisation Filter Array
PFC	Pulse Front Curvature
PFT	Pulse Front Tilt
PROD	Phase Reconstruction using an Over-complete Dictionary
PSF	Point Spread Function
PSNR	Peak Signal-to-Noise Ratio
PW	Petawatt (10^{15} W)
QED	Quantum Electrodynamics
QWP	Quarter-Wave Plate

QWLSI	Quadriwave Lateral Shearing Interferometer
RAVEN	Real-time Acquisition of Vectorial Electromagnetic Near-fields
ReLU	Rectified Linear Unit
ResNet	Residual Network
RS	Rayleigh–Sommerfeld
SCI	Snapshot Compressive Imaging
SGD	Stochastic Gradient Descent
SH	Shack-Hartmann
SHG	Second Harmonic Generation
SRSI	Self-Referenced Spectral Interferometry
STC	Spatiotemporal Couplings
STRIPED-FISH	Spatially and Temporally Resolved Intensity and Phase Evaluation Device: Full Information from a Single Hologram
TERMITES	Total E-field Reconstruction using a Michelson Interferometer Temporal Scan
Ti:Sa	Titanium Sapphire
TIE	Transport of Intensity Equation
TV	Total Variation
ULP	Ultra-intense Laser Pulse
ViT	Vision Transformer
XPW	Cross-Polarised Wave Generation

Chapter 1

Introduction

Part of humanity's nature is to explore. While this is rooted in a past evolutional advantage, in modern times this expresses itself as a desire to understand the world, and to investigate previously unexplored phenomena. It is at the extremities of science that a childlike wonder is often satisfied, where trivial questions regarding what is beyond boundaries are answered. Whether inquiries into the ultra-cold, ultra-small, ultra-fast or the ultra-distant, this natural curiosity has also contributed to the development of a plethora of new technology and applications that have been of benefit to the world; for example atomic clocks, quantum computing, nanotechnology, femtochemistry, the global satellite network and many others. Another instance of this is the field of ultra-intense laser pulses (ULPs), which represent some of the most radical environments that humanity has ever created [1–4]. To begin to intuit the level of extremity of these phenomena, one can make some comparisons. The peak power of these lasers is ~ 3 orders of magnitude larger than the total average global power consumption [5, 6], and when a pulse is subsequently focused down to a $1.1\mu m^2$ spot, its resultant peak intensity of $10^{23}\mathrm{W\,cm^{-2}}$ is larger than if the entire luminosity of the sun passed through a typical car's windscreen [7, 8]. The pulse achieves these extreme conditions despite having an energy of ~ 100J, approximately the same as a human expends in a single second, hinting at the brevity of its duration of tens of femtoseconds (10^{-15} s), which is to a second what a second is to 3 million years.

With any technology, it is important to be mindful of its applications, whether realised or potential, as they are principally the reasons for funding into its development. Applications which utilise ULPs, directly or indirectly, cover a wide range of science.

A large proportion of these involve the acceleration of particles in order to generate secondary sources [9], and thus ULPs inherit the benefits of these phenomena. Traditionally particles have been accelerated to high energies using facilities such as linear accelerators (LINACs) and synchrotrons. The former consists of a series of tubes with increasing length, where alternating current radio frequency electrodes are used to accelerate charged particles. Due to the high speed of these particles and the number of acceleration stages, LINACs end up being very large facilities. Synchrotrons accelerate particles in a loop using magnetic fields. The loop's radius is made as large as practically possible in order to minimise the energy loss due to synchrotron radiation – the most famous facility is the Large Hadron Collider (LHC) with its 27 km ring. The scale and cost of these facilities means that their applications are limited, and wide-spread use is impossible.

Due to the intense fields of an ultra-intense laser, particles can be accelerated to high speeds over a much smaller distance. A renowned example of this is the laser wakefield accelerator (LWFA) [10, 11]. Directing the laser into a plasma, the strong transverse fields cause the displacement of the charged particles. The lower inertia of electrons cause them to be accelerated faster than the ions, leaving an area of strongly positive charge trailing the laser. Electrons then rush back into this gap, forming a wake. Additional electrons that get caught in the region behind the laser pulse can be accelerated to relativistic velocities. Other particles such as ions can be accelerated by other means, such as the ablation of a material with a typically tightly-focused ULP (to maximise intensity) [12].

Accelerated electrons, protons, ions and neutrons are used in many ways. The FLASH radiotherapy proposal [13] delivers high-energy radiation at ultra-high dose rates to enable precise, non-invasive treatment of affected regions (such as tumours) [14, 15]. Its key benefit lies in the significant reduction of damage to surrounding healthy tissue compared to conventional radiotherapy, despite delivering the same physical dose – a phenomenon

whose underlying mechanisms remain an active area of research. This has been realised in mice by an ULP-produced proton beam [16]. Accelerated electrons can be used in microscopy to analyse the structures of materials [17], and can also be oscillated in order to produce X-rays, whose small wavelength allow for highly detailed imaging and deep penetration depth – both desirable in biomedical imaging [18]. Complementary to X-ray imaging is neutron imaging [19]. Due to its zero charge, the neutron interacts weakly with many common materials and has a deep penetration depth. Furthermore, it interacts predominantly with nucleus of atoms rather than the electrons; a property that means accelerated neutrons can be used to identify elements and isotopes. They have been used in applications such as cargo scans at airports [20] and to map the refractive index of materials [21].

There are various proposals for how this technology can be utilised in the energy sector; for example, in the fast ignition scheme of inertial confinement fusion (ICF) [22–25]. ICF has received increased recent publicity due to experiments at the National Ignition Facility (NIF) that surpassed 'breakeven'; that is, the energy that emerged from the reaction was larger than the energy supplied to the target [26, 27]. Here, fusion fuel is irradiated from all sides by a large number of lasers (sometimes with the aid of a hohlrahm), heating the fuel while its inertia keeps it compressed. The fast ignition scheme then proposes to blast the fuel with an ultra-intense laser. While this is not currently the typically used technique at large scale facilities such as NIF, it is still an active research area [24]. Furthermore, the previously mentioned secondary sources are involved in proposals by a number of the recently created start-ups in this space [28, 29], and there is interest in the possibility of using auxiliary heating of the fusion fuel by an electron beam [30]. Their use in ICF diagnostics is also widespread, such as in proton radiography [31]. Finally, there exists proposals for using this technology for the disposal of waste from fission reactors, via nuclear transmutation [5].

One can also perform basic research and laboratory astrophysics with ULPs [32–34]. As the maximal-achievable intensity increases, new regimes of physics are made experimentally

accessible. A current frontier is the Schwinger limit (10^{27}W cm^{-2}), where the vacuum becomes polarised [35]. This non-classical process is close to being experimentally verifiable.

Structure of Ultra-Intense Laser Pulses

The structure of the laser pulse – across space, time and polarisation – proves to be crucial, universally across all applications. Beginning with the most obvious case, applications using this type of laser are principally concerned with the focused intensity of the laser; a value that critically depends on the structure. With current technology, measuring and shaping the laser along one of these domains is relatively trivial. For example, in many high-power facilities the spatial profile of the wavefront is measured with a Shack Hartmann sensor [36] and is corrected with a deformable mirror in an adaptive feedback loop [37–39].

However, when there exists couplings across dimensions, for example spatiotemporal couplings (STCs), metrology becomes significantly more challenging. While being invisible to measurement techniques that focus solely on either space or time, STCs can reduce the peak intensity significantly [40], and can be introduced in the generation of the laser pulse [41]; for example in the final compressor with a misalignment of gratings [42], or a mismatch in the groove densities [43].

Conversely, the field of *structured light* describes deliberate attempts to manipulate light's structure in beneficial ways [44, 45]; efforts of which have spread to the ultra-intense regime [46]. A notable example of this is the 'flying focus', where a region of intensity is able to travel at tunable speed (including super-luminally) [47–49]. It has also been implemented using a range of equipment, where in each case the required STC is different; for example pulse-front curvature when using an axiparabola [50]. This effect is particularly useful in LWFA, where a classic impediment to acceleration is de-phasing due to the electrons 'outrunning' the laser [51], and there are also schemes investigating more complex couplings [52]. Structured ultra-intense light is also present in proposed

fundamental science experiments, such as gravitational wave generation and detection [53], and nonlinear QED [54]. Returning to ICF, the symmetry and uniformity of the system is important, to ensure the efficient compression and ignition of the fuel. This also applies to the laser pulse's structure, which has been investigated for the main lasers [55], and would naturally also apply to an ULP used in fast ignition.

Another dimension that can be used in the structuring of light is polarisation, and a growing amount of recent simulation work has demonstrated the importance of this domain for the properties of subsequently accelerated particles, in terms of the beam divergence and energy. This has been seen across all species of accelerated particle; for electrons [56–59], positrons [60, 61] and ions [62–64]. There are also investigations into the interaction between polarised light and other types of matter, such as plasma mirrors [65].

Experimental application of structured light techniques to ULPs remain limited [46]. The systems where structured light methods have generally been developed are low power and high repetition rate, representing a large divergence from the typical ultra-intense facility. One problem revolves around the question of how one can practically structure ultra-intense light, where more expensive and hard-to-obtain optics are required. However, an equally crucial problem is that there is no existing way to efficiently measure structured ULPs.

Characterisation of Ultra-Intense Laser Pulses

The characterisation of the laser is crucial for a variety of reasons, such as assessing if it is achieving its theoretically-capable peak intensity (and if not, understanding the reason), or for the continued amalgamation of structured light methods into ULPs. Another important point, is that due to the fact that beam-time at these facilities is limited and expensive, most prospective experiments are initially investigated extensively with simulations. When characterised, the information captured about a specific laser system can be incorporated

into these simulations and machine learning algorithms, thereby closing the reality-gap between experiment and theoretical models [66–68].

In the past decade, there have been methods developed to perform the full characterisation of the laser [40, 69–72]; this is in part what has inspired the recent flux of proposals to utilise structured light in ultra-intense experiments. The remaining restriction however, is that these techniques require multiple shots of the laser in order to perform the measurement; deemed necessary to capture the high dimensional data on a sensor that has, at most, two dimensions. This is problematic for a number of reasons. Firstly, these techniques typically take on the order of 100-1000 shots, which with typical ultra-intense facility repetition rates of 1 Hz, equates to many minutes of precious beam-time. This long measurement time is restrictive in another way; as described above in the case of the spatial profile, the shaping of a laser is often performed in an adaptive loop of iterative measurements and corrections. If one wishes to spread these adaptive optics principles to the spatiotemporal setting, a diagnostic needs to characterise the laser in as few pulses as possible, or time will become prohibitive. Finally, critical to the performance of many downstream applications is the shot-to-shot fluctuations of the laser, which can be significant. With these multi-shot methods, one only gains information about the average pulse, and thus these fluctuations are invisible.

The work of this thesis solves these problems with the development of a device capable of capturing the complete field structure of an ultra-intense laser pulse generated by a single shot of a laser, whilst providing uncertainty estimates. While this has been developed in the context of ultra-intense lasers, in order to mitigate the problems provided by low repetition rates, it can also provide benefits in any situation where individual pulses need to be studied.

1.1 Structure of Thesis

The following section details the relevant theory and existing work, required to understand the scientific progress that was achieved during this project. Predominantly, what is presented in this thesis is the development of a method to capture the complete vector field of an ultra-intense laser pulse. Chapter 3 begins by describing the initial concept to achieve this, and some proof-of-concept simulation-based results. It also discusses an attempt to implement this experimentally, and reasons why this initial approach was abandoned. Chapter 4 goes on to discuss various techniques that were developed during this project, and that were used as part of the method presented in the Chapter 5. This final chapter details the final measurement device, along with measurements on a petawatt-class laser at the Center for Advanced Laser Applications (CALA) in Munich.

The initial concept in Chapter 3 – to bring together snapshot compressive imaging and a wavefront sensor – is credited to Dr. Andreas Döpp. Aside from this, the work presented here is my own, aided by discussions with my supervisors, Dr. Döpp and Professor Peter Norreys. Similarly, Chapter 4 is solely my own work, with the exception of the experimental data used in Section 4.2.4 which was collected by Mr. Nils Weiße and Mr. Jakob Schröder. This chapter also benefited from regular discussions with my supervisors. The experiments of Chapter 5 were performed in collaboration with Dr. Döpp and students from his group, in particular Mr. Nils Weiße and Mr. Jannik Esslinger. In addition, Mr. Esslinger aided in the building of the prototype of the diagnostic while I was away from Munich. All the calibration procedures, software development (including the reconstruction algorithm) and data analysis presented in Chapter 5 was performed by myself.

1.2 List of Selected Peer Reviewed Publications

1. **S. Howard** *et al.* Single-shot Spatio-temporal Vector Field Measurements of Petawatt Laser Pulses. Nature Photonics (accepted). 2025
   Contributed to, and was present for, every part of this project including the conceptual theory, experiments and implementation of computational reconstruction algorithm.

2. **S. Howard** *et al.* Sparse Reconstruction of Wavefronts using an Over-Complete Phase Dictionary. Optics Express, 33(6):12939–12952. 2025
   Contributed to the theory and computational implementation. The experimental data was obtained by colleagues from LMU (Mr. Nils Weiße and Mr. Jakob Schröder) on a visit to Salamanca.

3. **S. Howard** *et al.* CoordGate: Efficiently Computing Spatially-Varying Convolutions in Convolutional Neural Networks. British Machine Vision Conference. 2023
   Was involved in all parts of this project, from the theory to the implementation.

4. **S. Howard** *et al.* Hyperspectral Compressive Wavefront Sensing. High Power Laser Science and Engineering, 11:e32. 2023
   The initial concept of bringing together a wavefront sensor and snapshot compressive imaging to measure the wavefront of a laser pulse was from Dr. Andreas Döpp. My part was the implementation, including the simulation and reconstruction algorithm.

5. A. Döpp, C. Eberle, **S. Howard** *et al.* Data-driven science and machine learning methods in laser–plasma physics. High Power Laser Science and Engineering, 11:e32. 2023
   Contributed generally, and especially to the chapter on Inverse Problems.

6. N. Weiße, J. Esslinger, **S. Howard** *et al.* Measuring spatio-temporal couplings using modal spatio-spectral wavefront retrieval. Optics Express, 31(12):19733–19745. 2023
   Contributed to conceptual discussions.

Chapter 2

Theory

This section details the relevant physics and previous work that is necessary to understand the content of the thesis. The multidisciplinary nature of this project means that there is a broad range of concepts to introduce.

2.1 Light

Discussions on the nature of light have a detailed history, involving some of the most famous physicists and philosophers (Newton, Huygens, Faraday, Gauss). The fundamental description of light is encapsulated in Maxwell's equations; in the absence of free charge, these equations take the form:

$$\nabla \times \vec{E}(\vec{r},t) = -\mu \frac{\partial \vec{H}(\vec{r},t)}{\partial t}, \tag{2.1}$$

$$\nabla \times \vec{H}(\vec{r},t) = \epsilon \frac{\partial \vec{E}(\vec{r},t)}{\partial t}, \tag{2.2}$$

$$\nabla \cdot \epsilon \vec{E}(\vec{r},t) = 0, \tag{2.3}$$

$$\nabla \cdot \mu \vec{H}(\vec{r},t) = 0, \tag{2.4}$$

where $\vec{E}$ and $\vec{H}$ are the electric and magnetic fields, and μ and ϵ are the magnetic permeability and electric permittivity respectively.

Faraday's law and Ampère's law (Eq. (2.1) and Eq. (2.2)) intuitively explain the propagation of electromagnetic waves: a changing magnetic field continuously induces an electric field, which in turn generates a changing magnetic field. This self-sustaining cycle allows the wave to propagate through space. Additionally, the presence of the curl operator in these equations ensures that the electric and magnetic fields are always perpendicular to each other. Combining these equations, one can derive the wave equation for these fields. In homogeneous free space ($\epsilon(x, y, z) = \epsilon_0$; $\mu(x, y, z) = \mu_0$), one can use the vector identity $\nabla \times \nabla \times \vec{E} = \nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E}$ to arrive at,

$$\nabla^2 \vec{E}(\vec{r}, t) - \frac{1}{c^2} \frac{\partial^2 \vec{E}(\vec{r}, t)}{\partial t^2} = 0, \tag{2.5}$$

where $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$ is the speed of light, and a similar equation can be found for the magnetic field. A solution is that of the plane wave,

$$\vec{E} = \vec{E}_0 e^{i\left(\vec{k} \cdot \vec{r} - \omega t + \phi_0\right)}, \tag{2.6}$$

where $\vec{k}$ is the wavenumber ($|\vec{k}| = \frac{2\pi}{\lambda}$), ω is the angular frequency and ϕ_0 is some phase term. Substituting this into the wave equation gives the dispersion relation of light in the vacuum, $\omega = ck$; i.e. there is no dispersion. Furthermore, this substitution yields the Helmholtz equation,

$$\nabla^2 \vec{E} + k^2 \vec{E} = 0. \tag{2.7}$$

In the presented case of propagation in homogeneous free space, $\epsilon(x, y, z) = \epsilon_0$, the wave equation is separable into each of the components of the vector, i.e. this equation applies to $U \in [E_x, E_y, \ldots, H_z]$. It therefore can be applied to each scalar component separately,

and is referred to as the scalar wave equation,

$$\nabla^2 U + k^2 U = 0 \tag{2.8}$$

While Chapter 5 details a technique to measure the vector field of ultra-intense laser pulses, the propagation of these pulses is always assumed to be through a vacuum, and thus this scalar treatment can be applied to each polarisation component individually. It is not difficult to see that this is not the case when the permittivity is inhomogeneous. For example, the vector identity then requires the use of the product rule,

$$\nabla \cdot \vec{E} = \frac{\nabla \cdot \left(\epsilon \vec{E}\right) - \vec{E} \cdot \nabla \epsilon}{\epsilon} = \frac{-\vec{E} \nabla \epsilon}{\epsilon} = -\vec{E} \cdot \nabla \ln(\epsilon), \tag{2.9}$$

to arrive at the wave equation,

$$\nabla^2 \vec{E}(\vec{r}, t) + \nabla \left(\vec{E} \cdot \nabla \ln(\epsilon)\right) - \frac{n^2}{c^2} \frac{\partial^2 \vec{E}(\vec{r}, t)}{\partial t^2} = 0, \tag{2.10}$$

where $n = \sqrt{\frac{\epsilon}{\epsilon_0}}$. This equation is vectorial, in the sense that depending on the variability of ϵ, each component of $\vec{E}$ experiences different dynamics.

Gauss' law for both the electric and magnetic fields (Eq. (2.3) & Eq. (2.4)) gives further insight into the nature of light. Considering the plane wave solution of Eq. (2.6), one can apply Gauss' law,

$$\nabla \cdot \vec{E} = i(\vec{E} \cdot \vec{k}) = 0. \tag{2.11}$$

Therefore, one sees that $\vec{E}$ and $\vec{H}$ are always perpendicular to the direction of propagation – light is a transverse wave. A diagram of light is shown in Fig. 2.1.

A further point about the wave equation is that it is linear – a fact that gives rise to the interference of light. Formally, if $\vec{E_1}$ and $\vec{E_2}$ are both solutions to the wave equation, then so is $\vec{E} = \vec{E_1} + \vec{E_2}$. This is a point that is fundamental to ultra-intense laser pulses described in Section 2.3.

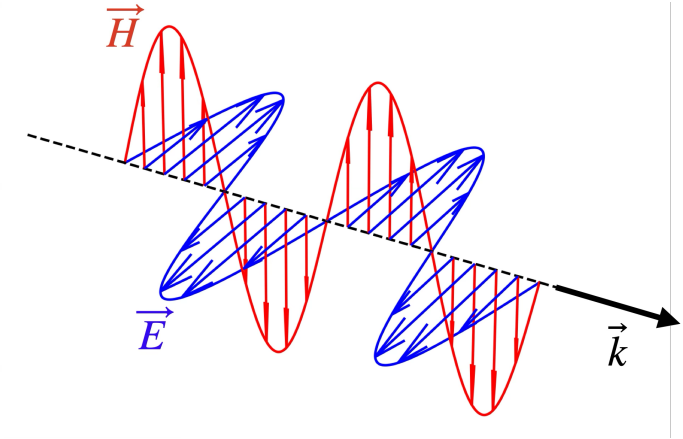


Fig. 2.1 A diagram of the propagation of light, with $\vec{E}$, $\vec{H}$ and the direction of propagation $\vec{k}$ all orthogonal to each other.

A final quantity of interest is the Poynting vector, which describes the power flux density of an electromagnetic field. It is given by,

$$\vec{N} = \vec{E} \times \vec{H}, \tag{2.12}$$

and has units of W m^{-2}. Since the Poynting vector oscillates at the optical frequency, what is typically referred to as the intensity is the cycle-averaged Poynting vector, $\langle N \rangle$.

2.1.1 Polarisation

The polarisation structure of light has particular relevance in Chapter 5. In order to describe this, one may use the Stokes parameters. Considering light propagating along the $\hat{z}$ direction, the electric field can be described in terms of two orthogonal components,

$$\vec{E} = \begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} E_{x0} e^{i\delta_x} \\ E_{y0} e^{i\delta_y} \end{bmatrix}. \tag{2.13}$$

The Stokes parameters are then defined as,

$$\vec{S} = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} (E_{x0})^2 + (E_{y0})^2 \\ (E_{x0})^2 - (E_{y0})^2 \\ 2E_{x0}E_{y0}\cos(\delta) \\ 2E_{x0}E_{y0}\sin(\delta) \end{bmatrix}, \tag{2.14}$$

where $\delta = \delta_x - \delta_y$ is the delay between the polarisation components. Each of the Stokes' parameters has an intuitive meaning: S_0 describes the total signal, S_1 describes the dominance of the horizontal over the vertical polarisation, S_2 describes the dominance of the diagonal over the anti-diagonal polarisation, and S_3 describes the dominance of the right-handed over left-handed polarisation. Under the assumption of fully-polarised light, another convenient way of representing the polarisation parameters is the ellipse, as seen in Fig. 2.2. This is parameterised via three numbers: The orientation angle, Ψ, the ellipticity angle, χ, and the length (equivalent to S_0). These are related to the Stokes parameters by,

$$\Psi = \frac{1}{2}\tan^{-1}\left(\frac{S_2}{S_1}\right), \tag{2.15}$$

$$\chi = \frac{1}{2}\sin^{-1}\left(\frac{S_3}{S_0}\right). \tag{2.16}$$

In this ellipse framework, perfectly circularly polarised light is found when the polarisation components have equal magnitudes and have a phase delay of $\frac{\pi}{2}$ between them. One sees that both $\pm\chi$, and therefore $\pm S_3$, will result in the same shape of the polarisation ellipse. Therefore, the sign of the ellipticity angle is also important for informing on the handedness of the light.

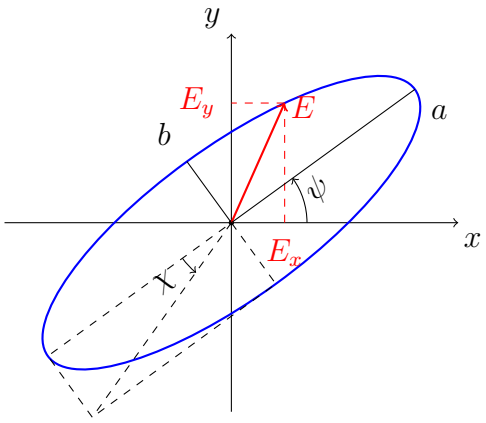


Fig. 2.2 Polarisation ellipse depicting the electric field vector $\vec{E}$ with components E_x and E_y along the x and y axes, respectively. The semi-major axis a and semi-minor axis b of the ellipse represent the maximum and minimum amplitudes of the electric field. The angle ψ indicates the orientation of the ellipse, while the angle χ represents the ellipticity angle, describing the shape of the polarisation ellipse.

2.2 Numerical Propagation of Light

This section details the relevant theory of light propagation, and the numerical methods used to simulate this, including discussions of resolution requirements. The Rayleigh–Sommerfeld (RS) diffraction integral provides an exact solution of the scalar Helmholtz equation in the half-space, $z > 0$, that lies infront of the initial plane, $z = 0$:

$$E(x, y, z) = \iint_{-\infty}^{\infty} E(x', y', 0) \frac{z e^{ikR}}{2\pi R^2} \left(\frac{1}{R} - ik \right) dx' dy', \tag{2.17}$$

where $R = \sqrt{\rho^2 + z^2}$, and $\rho = \sqrt{(x - x')^2 + (y - y')^2}$. Its validity is subject to some conditions; the complex field (or its normal derivative) must be specified across the entire initial plane (in the case of a laser one often approximates the field to be zero outside the known beam width), and it can only be used in regimes where the scalar treatment of light is valid. Within these conditions, it can be used for propagation at any distance [73], and a derivation is given in Chapter 3 of [74]. Eq. (2.17) is normally solved numerically, as the scenarios where analytical solutions exist are very limited. It can be seen as the convolution between the input field, $E(x, y, z = 0)$ and a kernel $G(x, y, z) = \frac{z e^{ikR}}{2\pi R^2} \left(\frac{1}{R} - ik \right)$. The convolution theorem then means that it is solved efficiently in Fourier space (although

it should be noted that this restricts the output field to be on the same domain as the input field).

There has been detailed analyses on the resolution criteria for this method, including in this Fourier-transform based method [75], and more generally [76]. In both cases, one must consider the necessary resolutions required for both the input field and the kernel, and choose the higher resolution (upsampling the other). Firstly, to avoid aliasing in the input field, it must be sampled at a rate at least twice its greatest spatial frequency, in accordance with the Nyquist sampling theorem. In practice, the input field may have hard edges (such as a top-hat beam) – to resolve these edges it is intuitive that the sampling rate would have to be infinite; however typically this can be neglected once a resolution is achieved that results in only a small amount of energy being lost. One must also consider the necessary resolution of the propagation kernel, G. It is apparent that the limiting factor is the fast oscillations of the exponential term, e^{ikR}; thus, one can differentiate with respect to a spatial dimension to find the maximum spatial frequency,

$$f_x = \frac{1}{2\pi}\frac{\partial\left(kR\right)}{\partial x} = \frac{1}{2\pi}\frac{kx}{R}, \tag{2.18}$$

and then use the Nyquist criterion once more to find the necessary sampling resolution,

$$\Delta x < \frac{1}{2f_{max}} = \frac{\lambda R}{2x}. \tag{2.19}$$

This criterion is plotted in Fig. 2.3. This highlights a difficulty for numerically simulating the focusing of ultra-intense laser pulses, which typically are very wide in the near field. Unless they are being propagated many tens of meters, the number of spatial samples required to solve the RS integral is very large (for example, propagating an 800 nm beam with width $x = 30$ cm a distance of $z = 30$ m requires 22500 spatial samples according to Eq. (2.19)). By rearranging to find the propagation distance, z, as a function of the

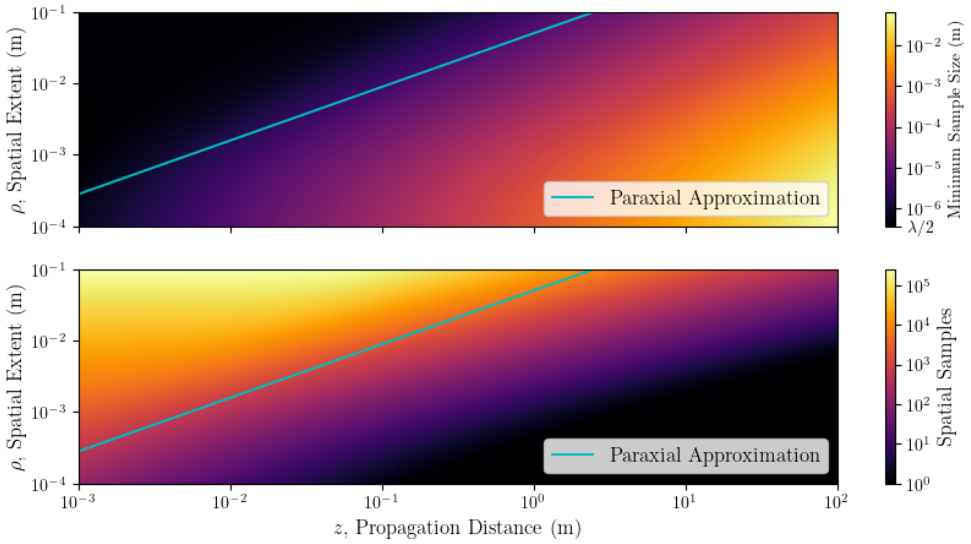


Fig. 2.3 Visualisation of the resolution criterion for numerically implementing the Rayleigh Sommerfeld Integral. Top shows the minimum sample size required according to Eq. (2.19). Bottom shows the consequential number of spatial samples. Also shown is the criterion for the paraxial approximation, $k\frac{\rho^4}{8z^3} = 2\pi$; thus this applies to everything significantly under the blue line.

resolution,

$$z^2 > x^2 \left(\frac{2}{\lambda} \Delta x - 1 \right), \tag{2.20}$$

one finds that as long as the sampling rate is below $\lambda/2$, any propagation distance is permitted.

Subsequently, approximations to this scalar diffraction theory are discussed that make numerical calculations more trivial, and can be applied in the paraxial regime (long propagation distance relative to transverse spatial extent). It's important to be mindful that there are real world applications that don't fit into this regime, and the RS integral is the more generally applicable method. However, to highlight the limitation of the scalar nature of the RS integral, consider the tight focusing of a laser pulse (perhaps for ion acceleration purposes), which will result in a longitudinal component of the field, E_z. Gauss's law (Eq. (2.3)) will then cause the coupling of the field components and the scalar treatment will no longer be applicable; thus the RS integral can no longer be used. In this case one must use vector diffraction theory [77, 78].

2.2.1 Paraxial Approximation

The paraxial approximation (PA) is very commonly used, and describes a regime where the angles between the optical axis and rays are small, and can make the numerical propagation of light more trivial. Mathematically, this can be expressed in two ways. Firstly, the paraxial approximation can be applied to the Rayleigh-Sommerfeld integral. First expressing $R = \sqrt{\rho^2 + z^2} = z\sqrt{1 + \frac{\rho^2}{z^2}}$, one knows that the angles of rays are small, $z \gg \rho$. Thus, one can approximate $R = z$ everywhere outside the exponential. Inside the exponential, one must be more careful due to its fast varying cyclical nature. Here one can find the Taylor expansion of R, and drop terms larger than second order in ρ/z under the condition, $k\frac{\rho^4}{8z^3} \ll 2\pi$. Substituting these approximations, one arrives at the paraxial form of the Rayleigh-Sommerfeld integral, the Fresnel diffraction integral,

$$E(x,y,z) = \frac{e^{ikz}}{i\lambda z}\iint_{-\infty}^{\infty} E(x',y',0)e^{\frac{ik}{2z}\left[(x-x')^2+(y-y')^2\right]}dx'dy', \tag{2.21}$$

$$= \frac{e^{ikz}e^{\frac{ik\left(x^2+y^2\right)}{2z}}}{i\lambda z}\mathcal{F}\left\{E(x',y',0)e^{\frac{ik\left(x'^2+y'^2\right)}{2z}}\right\}\left(\frac{x}{\lambda z},\frac{y}{\lambda z}\right). \tag{2.22}$$

The second line demonstrates a number of computational advantages of the Fresnel integral over the RS integral. Firstly, the input is directly related to the output of the Fourier transform. Thus, it is not required to also perform the FFT of a kernel, or the final IFFT. Another advantage is the rescaling of the spatial coordinate of the output field, which as shall be seen in subsequent sections, is crucial for computational efficiency in simulating the focusing of ultra-intense laser pulses.

One can also apply the paraxial approximation directly to the Helmholtz equation. The field varies slowly in the direction of propagation, or $|\frac{\partial^2 E}{\partial z^2}| \ll k^2|E|$. From this, the paraxial form of the Helmholtz equation is,

$$\nabla_\perp^2 E + 2ik\frac{\partial E}{\partial z} = 0. \tag{2.23}$$

These methods of applying the paraxial approximation are consistent, in that the Fresnel integral is an analytical solution of the paraxial Helmholtz equation. Another important advantage of the paraxial approximation to the Helmholtz equation is the existence of a number of analytical solutions – for example, the family of Gaussian beams. Not only can the propagation of these beams be calculated much faster, but also they give rise to intuitive parameters, such as the beam waist, w_0, and the Rayleigh range, z_R.

2.2.2 Angular Spectrum Method

One particularly intuitive method to think about light propagation is the angular spectrum method. Consider a complex field, $E(x, y, z = 0)$. The spatial Fourier transform of this field yields the frequency components. A further inverse Fourier transform returns it to real space;

$$\begin{aligned} E(x,y,0) &= \mathcal{F}^{-1}\left\{\mathcal{F}\left\{E(x,y,0)\right\}\left(\frac{k_x}{2\pi},\frac{k_y}{2\pi}\right)\right\} \\ &= \iint \mathcal{F}\left\{E(x,y,0)\right\}\left(\frac{k_x}{2\pi},\frac{k_y}{2\pi}\right) e^{i(k_x x+k_y y)} dk_x dk_y \\ &= \iint c(k_x,k_y) e^{i(k_x x+k_y y)} dk_x dk_y. \end{aligned}$$

The final line exhibits the fact that this equation can be interpreted as the integral over plane waves ($e^{i(k_x x+k_y y)}$), each with complex amplitudes $c(k_x, k_y)$. Given the wavelength of the light, λ, one can find the wavevector along z that means that $\lambda = \frac{2\pi}{|\vec{k}|}$ is satisfied. Defining $k = \frac{2\pi}{\lambda}$ and $k_\rho = \sqrt{k_x^2 + k_y^2}$,

$$k_z = \sqrt{k^2 - (k_\rho)^2}. \tag{2.24}$$

One then can propagate the field to a distance z via,

$$E(x,y,z) = \iint c(k_x,k_y) e^{i(k_x x+k_y y+k_z z)} dk_x dk_y. \tag{2.25}$$

Perhaps the key restriction for this method is the identical sampling spacing in both the input and output planes. However, this result proves useful in the following section.

2.2.3 Fourier Optics

The measurement of the pulse often occurs in the near field. In order to calculate the peak intensity of the pulse, it must be numerically propagated to the focus. The well known result of Fourier optics is that, under the paraxial approximation, the far field is found by the Fourier transform of the near field. For completeness, consider the near field, $U_\lambda(x, y, z = 0)$, which has a spatial Fourier transform of $U_\lambda(\nu_x, \nu_y, z = 0)$. This field travels a distance of F_l to a lens, where F_l is the lens' focal length. According to the angular spectrum method of propagation, the Fourier transform of the field in the lens plane is found as,

$$U_\lambda(\nu_x, \nu_y, z = F_l) = U_\lambda(\nu_x, \nu_y, 0) e^{i2\pi F_l \sqrt{\frac{1}{\lambda^2} - (\nu_x^2 + \nu_y^2)}} \tag{2.26}$$

$$\approx e^{i\frac{2\pi F_l}{\lambda}} U_\lambda(\nu_x, \nu_y, 0) e^{-i\pi\lambda F_l(\nu_x^2 + \nu_y^2)}, \tag{2.27}$$

where, the second line uses the paraxial approximation.

Now, the thin lens imparts a phase onto the field. Again, with the paraxial approximation, this is represented as a quadratic phase factor,

$$E_\lambda(x', y', F_l) = U_\lambda(x', y', F_l) e^{-i\frac{\pi}{\lambda F_l}(x'^2 + y'^2)}. \tag{2.28}$$

The Fresnel integral can then be used to propagate the pulse to the focal plane of the lens. Importantly, when inserting the lens' quadratic phase factor into the Fresnel integral, and propagating to the focal length of the lens, this factor cancels out with the factor inside the Fourier transform. Thus, propagating from the lens at $z = F_l$ to the focus at

$z = 2F_l$ is found by the simplified integral:

$$E_\lambda(x, y, z = 2F_l) = \frac{e^{ikF_l}}{i\lambda F_l} e^{i\frac{\pi}{\lambda F_l}(x^2+y^2)} \mathcal{F}\left\{U_\lambda(x', y', F_l)\right\}\left(\frac{x}{\lambda F_l}, \frac{y}{\lambda F_l}\right). \tag{2.29}$$

So finally, inserting Eq. (2.27) as the Fourier transform of the field at the lens plane, one finds that the field at the focus of the lens is,

$$E_\lambda(x, y, 2F_l) = \frac{e^{i2kF_l}}{i\lambda F_l} \mathcal{F}\{U_\lambda(x, y, 0)\}\left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right) \tag{2.30}$$

The prefactor of the Fourier transform is a global phase shift (it has no dependence on planar spatial position), and therefore will not affect the pattern. Thus, to find the field at focus, one first performs the spatial Fourier transform of the field in the near-field. This results in a field defined on a grid of its spatial frequencies, $\nu_x \sim \frac{1}{x}$. According to Eq. (2.29), this pattern is the spatial pattern in the far-field, defined at positions $(x, y) \to (\nu_x \lambda F_l, \nu_y \lambda F_l)$. The significance of this result, is that if one attempted to numerically simulate the focusing of a 30 cm ultra-intense laser pulse using the RS integral (or the Fresnel integral), the required resolutions may be prohibitively large. Instead, given that one is in the paraxial regime and the input field is sampled with sufficient resolution according to the Nyquist theorem, the field at focus can be calculated by a single Fourier transform.

Some details to note; each wavelength is now defined on a slightly different spatial grid, which can create a complication when one wishes to perform a further Fourier transform along the spectral dimension, in order to study the pulse in time. This issue can be solved by the interpolation of the spatial field for each wavelength onto a common grid. As with any Fourier based approach, one can apply zero-padding to the near-field in order to increase resolution in the far-field. This is suitable as one knows the spatial extent of the pulse in the near field, and can assume there is no field outside this region.

2.3 Ultra-Intense Laser Pulses

This section aims to detail the specific physics behind the properties and generation of ultra-intense laser pulses (ULPs), with specific application to the ATLAS-3000 facility (where the final experiments of this thesis were performed), although an exhaustive description of the laser system can be found elsewhere. ULPs achieve their high intensity by compressing energy both spatially and temporally. For the latter, an illuminating equation is that of the Fourier transform of a constant function, which yields the delta function,

$$\mathcal{F}\{1\} = \int_{-\infty}^{\infty} e^{i\omega t} d\omega = \delta(t) \tag{2.31}$$

Thus, in order to achieve a 'spike' in time, a natural strategy is to use the superposition of waves of different frequencies. As shall be described in proceeding sections, in practice it is not possible to use all frequencies, so the range of the integral is truncated to an interval $\in [\omega_0, \omega_0 + \Delta\omega]$. The specific shape of the spectrum will vary (typically some mixture of Gaussians), but for the case of a top-hat function, the Fourier transform yields the sinc function,

$$\int_{\omega_0}^{\omega_0+\Delta\omega} e^{i\omega t} d\omega \propto \mathrm{sinc}(t(\Delta\omega)) e^{-i\frac{2\omega_0+\Delta\omega}{2}t}. \tag{2.32}$$

Here, the ultra-intense laser pulse will have a sinc packet shape, along with a high frequency modulation. Both of these things depend on the spectral bandwidth, $\Delta\omega$, of the signal. Naturally, as the frequency range approaches all frequencies, one recovers the delta function. One further thing to note about the above equations, is the fact that the waves for all spectral components are in phase at $t = 0$, meaning that they all constructively interfere at this location (hence the delta function). This phase locking between frequency components is essential, as becomes apparent in the following section on the generation of ultra-intense laser pulses.

Considering the spatial profile of the pulse, Section 2.2.3 described how the pulse at focus is related to the near field by the Fourier transform. The task of achieving the

highest possible intensity then is seen trivially as making the power spectrum as confined as possible, typically by reducing the high spatial-frequency content of the pulse.

2.3.1 Ultra-Intense Laser Pulse Generation

Introduction to the Laser

The laser is an acronym for *Light Amplification by Stimulated Emission of Radiation.* Thus, it can be seen that the possibility of production of a laser beam began, like many other things, with Einstein and his 1917 prediction of stimulated emission [79]. When an electron is in an excited energy state, there is a probability of it spontaneously transitioning to a lower energy state and emitting a photon equal to the energy difference between the states; this is *spontaneous emission.* Due to the quantised energy levels of a medium, this will occur at specific frequencies of light [80]. Similarly, the presence of an electromagnetic wave with the same energy as an energy jump between levels causes the probability of an electron to transition to increase. Electrons can absorb the photon in order to transition to a higher energy state (*absorption*), or can be stimulated to emit a photon whilst dropping energy levels (*stimulated emission*). This additional photon is of equal frequency and is in phase with the original photon, enabling the coherent amplification of the light [81]. Interestingly, it was noted by Einstein that the transition probabilities of these two phenomena are the same. Denoting the occupancy of energy level i with n_i, while there are initially more electrons in the lower energy state ($n_1 > n_2$), the net effect will be absorption. However, when *population inversion* is achieved – more electrons are initially in the higher energy state ($n_2 > n_1$) – there will be a net effect of stimulated emission, electrons dropping energy level and emitting a photon of the same frequency as the original, yielding a net amplification of the electromagnetic radiation.

Laser Oscillator

It is Townes and Schawlow who are credited with the theoretical conception of the laser, when they proposed a design to achieve the deliberate amplification of light using these phenomena [82]. Their design, the laser oscillator – which is still used to this day – begins with a Fabry-Pérot cavity, consisting of two reflective components that face each other. Inside the cavity is a gain medium, which contains atoms with electrons in excited energy states. When the laser beam passes by these atoms, stimulated emission will cause the relaxing of the electrons, and the subsequent amplification of the laser beam.

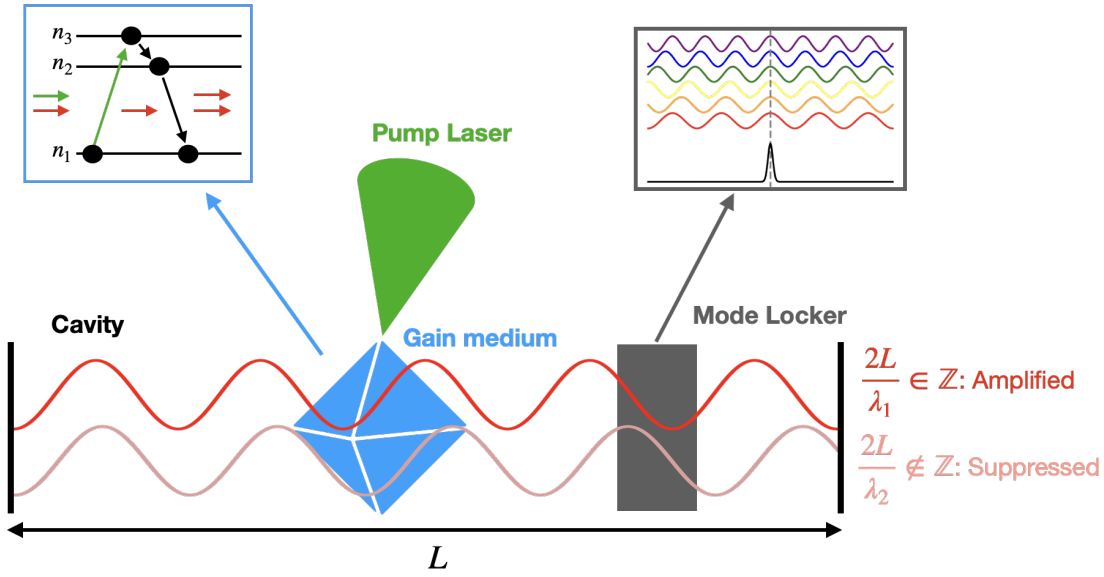


Fig. 2.4 A diagram showing the relevant components for a laser oscillator to create ultra-short pulses. A cavity contains the gain medium (which is being pumped by an external source), which must contain more than two energy levels. Mode locking is used to create ultra-short pulses.

From the above discussion, it is evident that a population inversion is necessary for the amplification of light by stimulated emission, on a timescale that is smaller than the associated spontaneous emission. Naturally, when the system is in thermal equilibrium, the system will obey Boltzmann statistics, and the lower energy level will be more populated. For this reason, the gain medium (containing the electrons) must be supplied external energy, which is referred to as pumping. In ultra-intense lasers this can be done using another laser ('optical pumping'). A difficulty arises in this two level system, when one

considers that the pump laser light will not only cause the desired absorption, but also stimulated emission. There will only be a net absorption of pump light while the occupancy of the lower energy level is less than the higher energy level, and thus, after perfect pumping one can expect $n_2 = n_1$; therefore population inversion is impossible. The most simple way to solve this issue is to find a material with a third energy level, n_3, that has a higher energy, and a short lifetime. The pump laser is then of a higher frequency, and excites the electrons to this third level, where they decay quickly to the second energy level, achieving population inversion relative to level one, $n_2 > n_1$. Finally, they can achieve stimulated emission and amplify the laser light. More energy efficient systems use a fourth level, with the laser amplification occuring between the middle two of these. The lower lasing level decays quickly to the lowest energy state, which means that the lower lasing level remains relatively empty and population inversion is achieved more efficiently.

The nature of stimulated emission means that the output of a laser is coherent spatially and, for each monochromatic component, temporally. As the light propagates in the oscillator, waves with wavelengths that fit an integer number of waves within the oscillator length will constructively interfere and form standing waves, while other modes cancel themselves out over time. The condition for a mode, with wavelength λ, to be amplified is that $\frac{2L}{\lambda} \in \mathbb{Z}$, for a cavity of length L. Due to the relative small size of wavelengths compared to the cavity size, there are a large number of modes that can be amplified (although this also depends on the energy level structure of the gain medium). The design of the laser cavity (and choice of gain medium) are tuned in order to yield a desired longitudinal mode profile (spectrum). A very commonly used gain medium, and the one used at the ATLAS-3000 facility, is a Titanium Sapphire (Ti:Sa) crystal; one reason as to why is the fact that it has the widest gain bandwidth of 650-1100 nm [83, 84], which is crucial for generating short pulses. There are many thousand of these modes (for Ti:Sa) propagating independently of each other and therefore not in phase; thus, the intensity output of the laser will be fairly constant over time.

As previously discussed, in order to achieve a short pulse it is necessary to create a fixed phase relationship between each of the spectral components, which describes the principle of *mode locking* [85, 86]. There exist many methods to achieve this; passive techniques typically exploit the fact that within the cavity, the random relative phases between the chromatic components will create many pulses. Then, they selectively transmit shorter, more intense pulses, which correspond to places where more spectral components are in phase [87]. This is the principle of the saturable absorber. This will absorb light until a pulse with sufficient intensity saturates it, allowing the transmission of just the peak. The Kerr lens is based on a similar principle which is a nonlinear process that preferentially focuses higher intensity pulses. A sketch of a laser oscillator generating ultra-short pulses is shown in Fig. 2.4.

At the ATLAS-3000 facility, the oscillator generates laser pulses with very short duration (< 10 fs), a high repetition rate and low energy ($\sim$ nJ). In order to increase the energy of the pulse, it must be amplified. Initially, while the pulse contains a low amount of energy, it can be amplified by simply passing through a Ti:Sa gain medium in a multi-pass configuration (such as a bow-tie). However, once the pulse becomes too intense, it causes nonlinear changes to the refractive index of materials, resulting in self-focusing which damages optical components.

Chirped Pulse Amplification

The technique for the amplification of ultra-intense pulses is known as chirped pulse amplification (CPA) [88]. This method is largely to mitigate the fact that even before focusing, ultra-intense pulses contain enough intensity to destroy optical equipment. Therefore, before the amplification process one can spread the energy across time, using a *stretcher*. The pulse is then amplified in a normal way using a Ti:Sa crystal, before being recompressed in a *compressor*.

In order to stretch the pulse, one can impart dispersion, meaning that higher or lower frequencies end up in the front of the pulse (referred to as normal and anomalous dispersion

respectively); to achieve this, the stretcher is an optical system designed so that its path length depends on the frequencies. The stretched pulse then passes through the amplifier where it gains energy but may also obtain further dispersion from the gain medium. Finally, a compressor recompresses the pulse; this is an optical system that aims to completely cancel out the accumulated dispersion from the stretcher and the amplifier so that in total, each frequency component leaves the CPA system with the same relative phase, restoring the original short temporal duration. A simple way to adjust the dispersion is by the combination of two gratings along with simple optical components, as seen in Fig. 2.5, which demonstrates anomalous dispersion. A simple setup reminiscent of this is used in the compressor to remove the dispersion (the ATLAS-3000 laser uses two pairs of gratings in order to achieve compression of a pulse with greater dispersion).

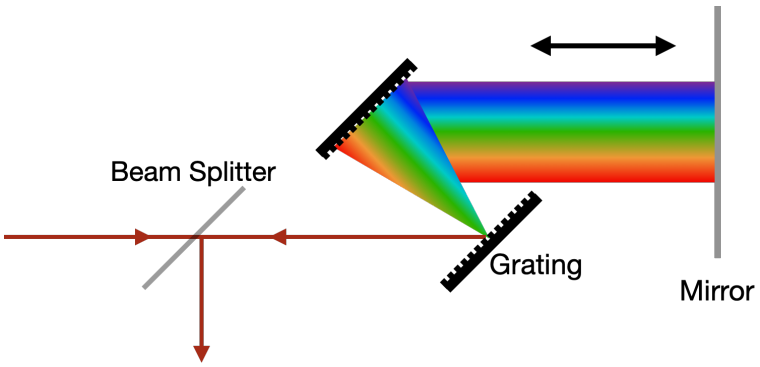


Fig. 2.5 The dispersing of a laser pulse using a pair of gratings. This setup imparts anomalous dispersion, where the higher frequencies (blue) travel a shorter distance and thus will be in the front of the resultant pulse.

A key challenge in the stretching and compressing is the precise control of the dispersion, including higher order terms (nonlinear relations between the frequency and its relative position in the pulse). If the dispersion is not perfectly cancelled out, spatiotemporal couplings will be present in the resulting pulse. The two grating system is very simple but does not provide the necessary control, especially over higher order dispersion terms [89]. The stretcher must provide a positive (normal) dispersion, and the most common method, and the one used at the ATLAS-3000 facility, is the Offner stretcher [90]. Here, the beam first hits a diffraction grating, propagates to a concave mirror, before being reflected by a

convex mirror and onto the concave mirror once more. This allows for a large dispersion, whilst maintaining control over undesirable higher order dispersion terms.

Once stretched, this chirped pulse undergoes amplification from a Ti:Sapphire crystal in multiple passes, before finally it is recompressed to a shorter duration. In ATLAS, the stretcher before the main amplification stage increases the pulse duration from ~ 20 fs to ~ 800 ps, which is more than 4 orders of magnitude. Beginning with an initial energy from the oscillator on the order of nano Joules, the final pulse has an energy of up to 70 Joules.

2.3.2 Beam Shaping

In all ultra-intense laser labs there are common devices in the beam-line that are used to try to control the properties of the pulse. For example, a deformable mirror is often used to control the spatial wavefront of the beam. This is a mirror consisting of an array of elements than can be shifted to better shape the wavefront. In order to control the spectral phase, an acousto-optic programmable dispersive filter is used (in its commercialised form called DAZZLER). This works by generating an acoustic wave in a birefringent medium, which creates a varying refractive index. This modulates the propagating light, imparting a desired phase delay between spectral components.

Of course, in order to use beam shaping techniques, one must have a measurement of the current pulse in order to know how to adjust the device, which is the basis of the proceeding section.

2.4 Ultra-Intense Laser Metrology

In order to assess whether ultra-intense laser systems are performing as desired, it is essential to take measurements of the laser pulses. Often it is the focused intensity which is of interest, and there exist proposed methods to measure this directly; for example via the ionisation of xenon gas [91]. However, these techniques do not provide insight into the underlying physical mechanisms that determine the intensity. By instead measuring

the structure of the laser pulse, valuable information can be obtained, as the processes involved in its generation leave characteristic traces in its structure. This not only helps in understanding the underlying physics but also provides feedback for beam shaping techniques. This section first introduces the problem of capturing the structure of an ultra-intense laser pulse, and proceeds to describe existing methods. In totality, there are a very large number of laser characterisation techniques and it is impossible to discuss them all in detail; one can refer to recent reviews for more details [92, 72]. This section will instead detail some of the most popular and capable methods, including information about their limitations.

2.4.1 The Measurement Problem

An ultra-intense laser pulse, without any assumptions on particular focusing conditions, can be described in general by a spatiotemporal vector with three polarisation components depending on coordinates (x, y, z, t),

$$\vec{E}(x,y,z,\lambda) = \begin{bmatrix} E_x(x,y,z,\lambda) \\ E_y(x,y,z,\lambda) \\ E_z(x,y,z,\lambda) \end{bmatrix} = \begin{bmatrix} \sqrt{I_x(x,y,z,\lambda)}e^{i\phi_x(x,y,z,\lambda)} \\ \sqrt{I_y(x,y,z,\lambda)}e^{i\phi_y(x,y,z,\lambda)} \\ \sqrt{I_z(x,y,z,\lambda)}e^{i\phi_z(x,y,z,\lambda)} \end{bmatrix}. \tag{2.33}$$

Its structure is essential to the applications of such pulses; in the simplest case in terms of focus-ability and therefore peak intensity. In all attempts at metrology, it is this object that one is trying to gain information about. However, this structure has high dimensionality, exacerbated by the fact that existing sensors only have access to the intensity rather than the field itself, meaning that the intensity and phase typically both have to be captured. Thus, the final data size is $2 \times 3 \times n_x \times n_y \times n_z \times n_\lambda$, where the factor of two comes from the need to measure intensity and phase, and the factor of three represents the polarisation components.

The challenge of capturing this information on sensors that are restricted to a maximum of two dimensions, in the case of the standard CMOS sensor, and are generally insensitive to the polarisation of light, is great. The task is simplified by collimating the beam, so that the propagation is assumed to be along $\hat{z}$. Then, a measurement of the field in a single plane at z_0 allows the subsequent calculation of the field at any plane $\tilde{z}$, by utilising a numerical propagator. Furthermore, as light is a transverse electromagnetic wave, the component of the electric field along $\hat{z}$ is 0, and thus one only needs to capture two components,

$$\vec{E}(x,y,\lambda) = \begin{bmatrix} E_x(x,y,\lambda) \\ E_y(x,y,\lambda) \end{bmatrix} = \begin{bmatrix} \sqrt{I_x(x,y,\lambda)}e^{i\phi_x(x,y,\lambda)} \\ \sqrt{I_y(x,y,\lambda)}e^{i\phi_y(x,y,\lambda)} \end{bmatrix}. \tag{2.34}$$

This object has a data size of $2 \times 2 \times n_x \times n_y \times n_\lambda$, which can contain couplings between four dimensions – two spatial (x, y), spectral/temporal (λ/t) and polarisation $(p \in [p_x, p_y])$. Once captured onto a grid of sufficient resolution, in order to fully characterise this object, one must be able to resolve the difference between neighbouring pixels in all four dimensions, i.e. $\Delta = E_x(x_i, y_i, \lambda_i) - E_y(x_j, y_j, \lambda_j)$.

In order to further simplify this problem, a number of common assumptions are often made. Firstly, that the field is scalar (linearly polarised); due to the properties of stimulated emission, this is normally valid for beams that have not had their polarisation structure deliberately manipulated. Another assumption is that of the separability of the field into spatial and temporal components, $E(x, y, \lambda) = A(x, y)B(\lambda)$, allowing the field to be resolved by two separate measurements in space and time/spectrum, thereby reducing the data size that needs to be captured to $2 \times (n_x \times n_y + n_\lambda)$. The following details measurement methods that are based on these assumptions, resolving either the spatial or the temporal profile of the field, and can be found at all ultra-intense laser facilities.

2.4.2 Spatially Resolved Measurements

It is possible to encode the wavefront of light into an intensity measurement, due to the fact that the wavefront plays a crucial role in shaping the intensity distribution of light as it propagates. Wavefront measurement techniques leverage this principle in different ways. Some techniques, such as the Gerchberg-Saxton (GS) algorithm [93] or transport of intensity equation (TIE) methods [94], rely on knowledge of the intensity at a reference plane and compare it with intensities measured at other planes to reconstruct the phase. This collection of methods based on propagation is not pursued in this thesis, due to the desire to create a compact device (prohibited by the measurements of intensity in multiple planes) and due to their relatively high noise sensitivity, which is not desirable with compressed sensing based method such as the one presented in Chapter 3.

Two other techniques of measuring the spatial field of a laser are described in detail, as they are used throughout this thesis. Both methods do not encode the wavefront directly, but instead capture the spatial gradient of the wavefront. Thus, a reconstruction process (termed *phase stitching*) is used to recover the wavefront, which is introduced in Section 2.4.2.

Shack Hartmann Sensor

The first method is the Shack Hartmann sensor [36]. Here the field is sampled across its spatial domain by an array of microlenses, creating characteristic spot patterns of foci. According to the local tilt at each location, the focal spot of each microlens is shifted. Consider a wavefront, defined at two points $x = -\Lambda/2$ and $x = \Lambda/2$, where Λ is the microlens pitch size. The gradient at $x = 0$ is then found by finite difference. This also can be seen as the angle of the slope of the wavefront, θ, found by trigonometry,

$$\frac{w(x = \Lambda/2) - w(x = -\Lambda/2)}{\Lambda} = w'(x = 0) = \tan(\theta). \tag{2.35}$$

When this slope passes through the microlens, it will cause a shift in spot at the detector, which is placed in the focal plane of the microlens. This shift is calculated again by trigonometry, using the focal length, f,

$$\Delta x = f \tan(\theta) = f w'(x = 0). \tag{2.36}$$

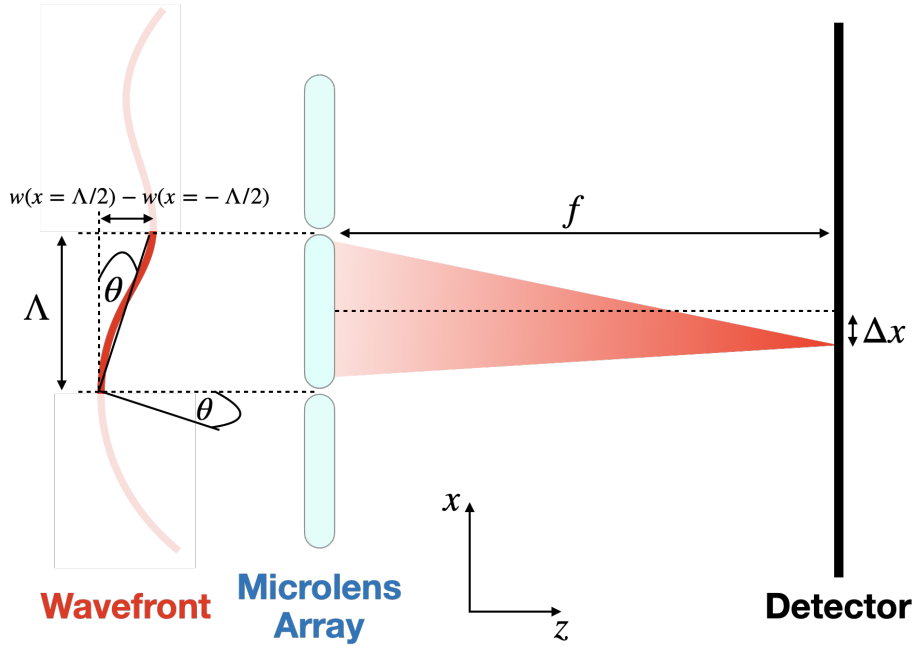


Fig. 2.6 A diagram showing geometry of the Shack Hartmann sensor. The wavefront is split into sections and each is locally focused to a foci at the detector. The shift of the spot encodes the derivative of the wavefront.

This geometry is seen in Fig. 2.6. Typically, the shift of the microlens foci is calculated using a simple centroiding algorithm (the change in centre of mass). One notes that in the spacing of two microlens, the wavefront may also have high frequency features, which may cause the foci to be aberrated. While this can cause an error in a standard centroiding algorithm, a deep learning approach can actually extract some additional information from the aberration [95]. From the spot displacements, the wavefront derivatives can be extracted, and stitched together, as is described subsequently.

Lateral Shearing Interferometer

The second method is based on interferometry, where classically, a pulse is allowed to interfere with a known reference pulse and the phase difference between them is then encoded into the resulting interferogram. Naturally, it is hard to obtain a perfectly-known reference pulse, so instead, the pulse can be interfered with a laterally shifted copy of itself; this is the case of the lateral shearing interferometer [96], which was first applied to the measurement of ultra-short laser pulses in the late 1990s [97]. The most common implementation is the quadriwave lateral shearing interferometer (QWLSI), where a grating is used to produce four copies of the beam, with the shearing directions aligned along two orthogonal axes. The resultant desired response from the QWLSI on an input field, E_{in}, can be described as the convolution, $E_{out} = E_{in} * S$, where S is,

$$S(x, y) = \delta(x - k, y) + \delta(x + k, y) + \delta(x, y - k) + \delta(x, y + k), \tag{2.37}$$

where k quantifies the level of shear. By considering the Fourier plane of a 4-f imaging system, this calculation can be considered by the convolution theorem. Thus, the desired transmission of the grating is given by the Fourier transform of S, and results in a bi-sinusoidal function. Fabricating this is hard to realise in practice, but an approximation of this ideal grating is fabricated by a three level system [98], which is the design used in the devices commercialised by the company Phasics. One further benefit of this four-copy design, is that it enables the extraction of two separate pairs of orthogonal gradients, meaning two distinct estimates for the wavefront are found, providing error estimation (in actual fact, this is also a property of the Shack Hartmann sensor [99]).

The grating causes a tilt to the wavefront of each copy. The azimuthal angle, ζ, of this tilt is related to the pitch of the grating, Λ, and the wavelength, λ, according to Bragg's law,

$$\zeta(\omega) = \arcsin\left(\frac{\lambda}{\Lambda}\right). \tag{2.38}$$

For each copy, this tilt is in a different direction, denoted by tilt unit vectors, $\vec{v} \in [\hat{x}, \hat{y}, -\hat{x}, -\hat{y}]$. The tilt adds a spatially varying phase for each copy m, relative to the original optical axis, as,

$$\Delta\phi_m(\vec{r}, \omega) = k\vec{r} \cdot \vec{v}_m \sin(\zeta(\omega)), \tag{2.39}$$

where $k = \frac{2\pi}{\lambda}$ is the wave vector of the spectral component of the pulse. Furthermore, as the copies propagate to the detector, they accumulate a lateral shear,

$$\vec{\Delta r}_m = \tan(\zeta(\omega))\vec{v}_m \Delta z. \tag{2.40}$$

The geometry described here can be seen in Fig. 2.7.

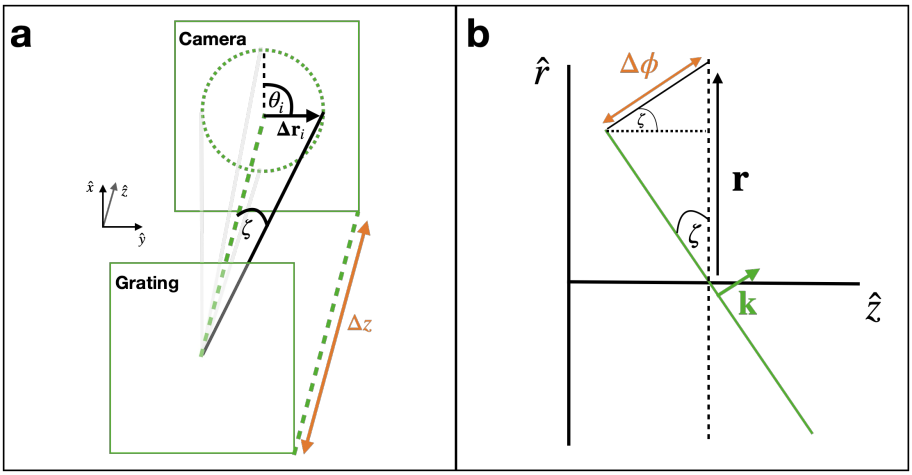


Fig. 2.7 Two diagrams showing the geometry of the quadriwave lateral shearing interferometer. Four copies of the beam are created, each with a tilt in a different direction. They then propagate where they accumulate lateral shear relative to each other. **a** shows the geometry of propagation. **b** shows the phase imparted by the tilt.

Thus, beginning with an input field, $E(\vec{r}) = \sqrt{I_0(\vec{r})}e^{i\phi(\vec{r})}$, the field of copy m, is expressed as,

$$E_m(\vec{r}, z = \Delta z) = \frac{1}{4}\sqrt{I_0(\vec{r} - \vec{\Delta r}_m, 0)}e^{i\left(\phi(\vec{r}-\vec{\Delta r}_m,0)+k\vec{v}_m\cdot\vec{r}\sin(\zeta(\omega))\right)}. \tag{2.41}$$

The corresponding intensity of the interferogram of these copies is,

$$I_{total}(\vec{r},\Delta z) = \left|\sum_{m=1}^{4} E_m(\vec{r},\Delta z)\right|^2 \tag{2.42}$$

$$= \frac{1}{16}\sum_{m=1}^{4} I_0(\vec{r}-\vec{\Delta r}_m,0) + \frac{1}{16}\sum_{m\neq n} E_m(\vec{r},\Delta z)E_n^*(\vec{r},\Delta z). \tag{2.43}$$

Analysing the cross terms,

$$E_m(\vec{r},\Delta z)E_n^*(\vec{r},\Delta z) = \sqrt{I_0(\vec{r}-\vec{\Delta r}_m,0)I_0(\vec{r}-\vec{\Delta r}_n,0)}e^{i\left(\phi(\vec{r}-\vec{\Delta r}_m,0)-\phi(\vec{r}-\vec{\Delta r}_n,0)+k(\vec{v}_m-\vec{v}_n)\cdot\vec{r}\sin(\zeta(\omega))\right)}, \tag{2.44}$$

one sees that encoded into the interferogram is the spatial gradient of the wavefront, between each set of beam copies. Furthermore, these terms are modulated by the difference in tilt vector, $(\vec{v}_m - \vec{v}_n)\cdot\vec{r}$, and are therefore separated in Fourier space. Accordingly, the gradient derivatives are extracted using Fourier filtering [100]; this method involves Fourier transforming the signal before shifting each modulated peak back to the centre (and therefore demodulating) and applying a low-pass filter. Finally, an inverse Fourier transform yields an object where the argument is proportional to the wavefront derivative. This method is further used in a later section (and depicted in Fig. 2.9). The wavefront is then obtained by the stitching of the gradients, as is described in the following section (or by other methods such as Fourier integration [101]).

Phase Stitching

Here is discussed the principle of phase stitching – the recovery of the phase from measurements of its derivatives; this is explained here in detail in order to give some background to a method developed as part of this thesis which is presented in Section 4.2. In both the previously discussed methods, the quantity captured in the raw signal is the spatial

gradient of the wavefront, which is, in vectorised form,

$$\vec{s} = \begin{bmatrix} s^x_{0,0} & s^y_{0,0} & s^x_{0,1} & s^y_{0,1} & \dots & s^x_{1,0} & s^y_{1,0} & s^x_{1,1} & s^y_{1,1} & \dots \end{bmatrix} \tag{2.45}$$

where $s^x_{i,j}$ describes the derivative of the wavefront at location (i, j) with respect to the x direction. This slope data is related to the vectorised wavefront, $\vec{w}$, via the derivative matrix, $\mathbf{D}$, forming a linear system, $\vec{s} = \mathbf{D}\vec{w}$.

In the simplified 1D case, this matrix has the form,

$$\mathbf{D} = \frac{1}{d} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \tag{2.46}$$

where d is the spacing by which the gradient is sampled. The variations in reconstruction come in the form with which one expresses the wavefront. Naturally, it can be expressed using any complete basis, so that, $\vec{w} = \mathbf{B}\vec{c} = \sum_i \vec{B}_i c_i$, where $\vec{c}$ is the vector of modal coefficients and $\mathbf{B}$ is the set of basis functions. Combining these equations, and defining a new matrix as the derivatives of a set of basis functions, $\mathbf{D_B}$, one arrives at,

$$\vec{s} = \mathbf{DB}\vec{c} = \mathbf{D_B}\vec{c}. \tag{2.47}$$

To then recover the wavefront, one notes that this system is overdetermined; if the wavefront has dimensions $(x \times y)$, the slopes contain $(2 \times x \times y)$ entries. Thus, the coefficient vector is found by minimising the squared error,

$$\vec{c} = \underset{\vec{c}}{\operatorname{argmin}} \left\{ |\mathbf{D_B}\vec{c} - \vec{s}|^2 \right\}, \tag{2.48}$$

$$= \left(\mathbf{D_B}^T \mathbf{D_B} \right)^{-1} \mathbf{D_B}^T \vec{s}. \tag{2.49}$$

The $\left(\mathbf{D_B}^T\mathbf{D_B}\right)^{-1}\mathbf{D_B}^T$ term is referred to the pseudoinverse of the derivative matrix $\mathbf{D_B}$, and can be easily calculated in python libraries such as pytorch (`pytorch.pinv`).

There exists two predominant reconstruction techniques, referred to as *Zonal* and *Modal*. Beginning with the former, the raw signal on the sensor is typically measured on a pixel grid arranged in Cartesian coordinates, often referred to as the *pixel* basis. Each pixel represents a mode in this basis, with its intensity value serving as the coefficient,

$$w(x,y) = \sum_{i=0}^{n}\sum_{j=0}^{m} c_{ij}\delta(x-i)\delta(y-j), \tag{2.50}$$

where, $w(x,y)$ is the wavefront, and c_{ij} are the coefficients.

Conceptually, it might seem straightforward to perform the reconstruction of the wavefront in this same basis, determining the phase value for each individual pixel. This approach, known as zonal reconstruction, is commonly used in various wavefront sensing methods, including lateral shearing interferometry (LSI) and Shack Hartmann (SH) sensors [102–106].

The modal reconstruction technique typically uses the Zernike basis [107]. This basis can be intuitively understood as a polar coordinate adaptation of polynomial regression, which is optimised for representing wavefront aberrations in optical systems. This makes the Zernike modes' applicability hierarchical, and by truncating the coefficients one can reliably calculate standard aberrations, improving the efficiency and reducing the computational cost of reconstruction. There are many examples of the use of this basis in wavefront reconstruction [108, 109, 39, 110–112].

$$w(R,\theta) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} c_n^m Z_n^m(R,\theta), \tag{2.51}$$

One notes that in the above formalism, the zonal reconstruction technique is just a special case of the modal in which the basis, $\mathbf{B}$, is simply the pixel (Cartesian) basis; the identity matrix. However, there are other special types of functions that cannot efficiently

be expressed by either of these bases. An example of which is the optical vortex, which provides the motivation for the method developed in Section 4.2.

2.4.3 Temporally/Spectrally Resolved Measurements

Temporally resolved measurements are important due to the fact that they offer a measurement of the pulse duration, which is fundamental to the calculation of peak intensity (although when used in isolation, this type of measurement provides, in general, an underestimate of the pulse duration). These techniques measure the spectral phase – the relative phases between spectral components, and typically use some form of interferometry. Principles such as those in lateral shearing interferometry cannot be applied, as neighbouring wavelengths interfere at frequencies that are too high to resolve. Therefore, the pulse must be interfered with itself, or with some reference pulse.

Here will be detailed two methods, the most popular that are used in labs; self-referenced spectral interferometry (SRSI, or known in its commercialised form as the WIZZLER), and frequency resolved optical gating (FROG).

SRSI

The first technique described here, and the one that is used to measure the spectral phase in the experiments of Chapter 5, is self-referenced spectral interferometry [113]. This is a widely used method in its commercial form (the WIZZLER from FASTLITE). This method creates an approximately flat reference, using a similar methodology to the principle of spatial filtering with a pinhole, before capturing the interference between the reference and the desired pulse. The reference is generated from the original pulse (hence 'self-referenced') by a nonlinear process. The cross-polarised wave generation (XPW) mechanism creates a signal that is approximately the original signal cubed,

$$E_{xpw}(t) \propto |E(t)|^2 E(t). \tag{2.52}$$

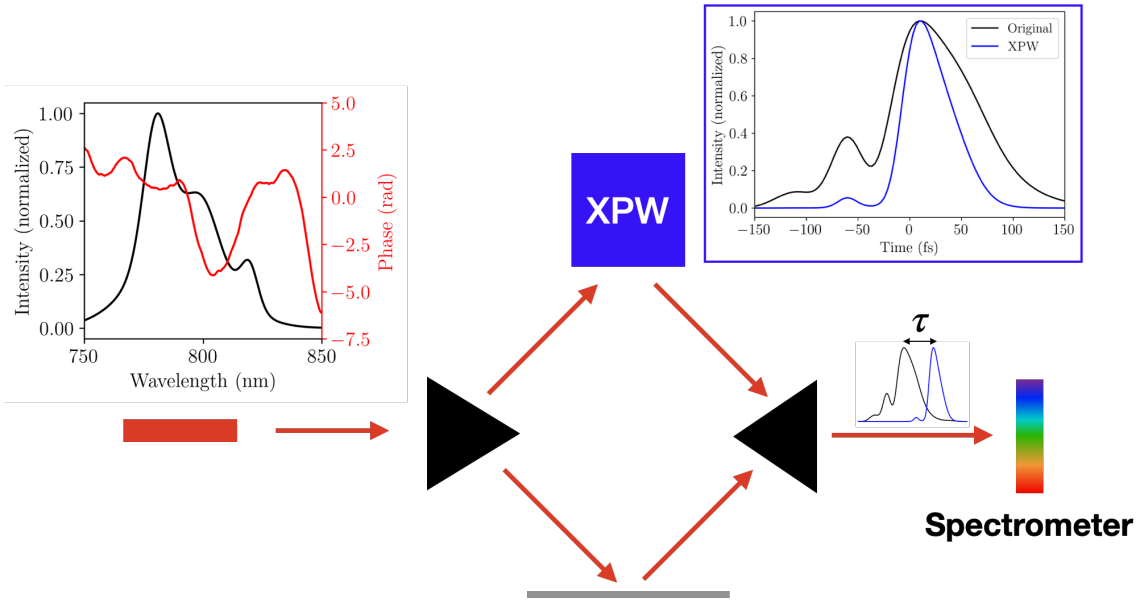


Fig. 2.8 A diagram showing the principle of self referenced spectral interferometry, reproduced from [113]. The unknown pulse is split into two copies. One copy experiences cross-polarised wave generation, which has the effect of spectrally filtering the phase, yielding a reference with a relatively flat wavefront. This is then recombined with the unknown pulse and the interferogram is captured on a sensor.

Relatively, this means that the section of the signal with the highest magnitude passes through a kind of pinhole. Thus, effectively only a small region of the phase is sampled, and expanded over the spectrum, and so it is flat compared to the original pulse. A relative delay is also inserted between the signal and the reference, in order to create a modulation which allows for the isolation of the phase shift via Fourier filtering. The signal is captured on a spectrometer and requires no scanning. This is seen in Fig. 2.8.

Along with the delay, the interferogram at the detector has the form,

$$I(\omega) = \left|E_{xpw}(\omega) + E(\omega)e^{i\omega\tau}\right|^2 \tag{2.53}$$

$$= |E(\omega)|^2 + |E_{xpw}(\omega)|^2 + 2\,|E(\omega)|\,|E_{xpw}(\omega)|\cos\left(\phi(\omega) - \phi_{xpw}(\omega) + \omega\tau\right). \tag{2.54}$$

The delay, τ, is what separates the interference from the DC term in Fourier space, and allows reconstruction to be possible. The Fourier filtering process is shown in Fig. 2.9.

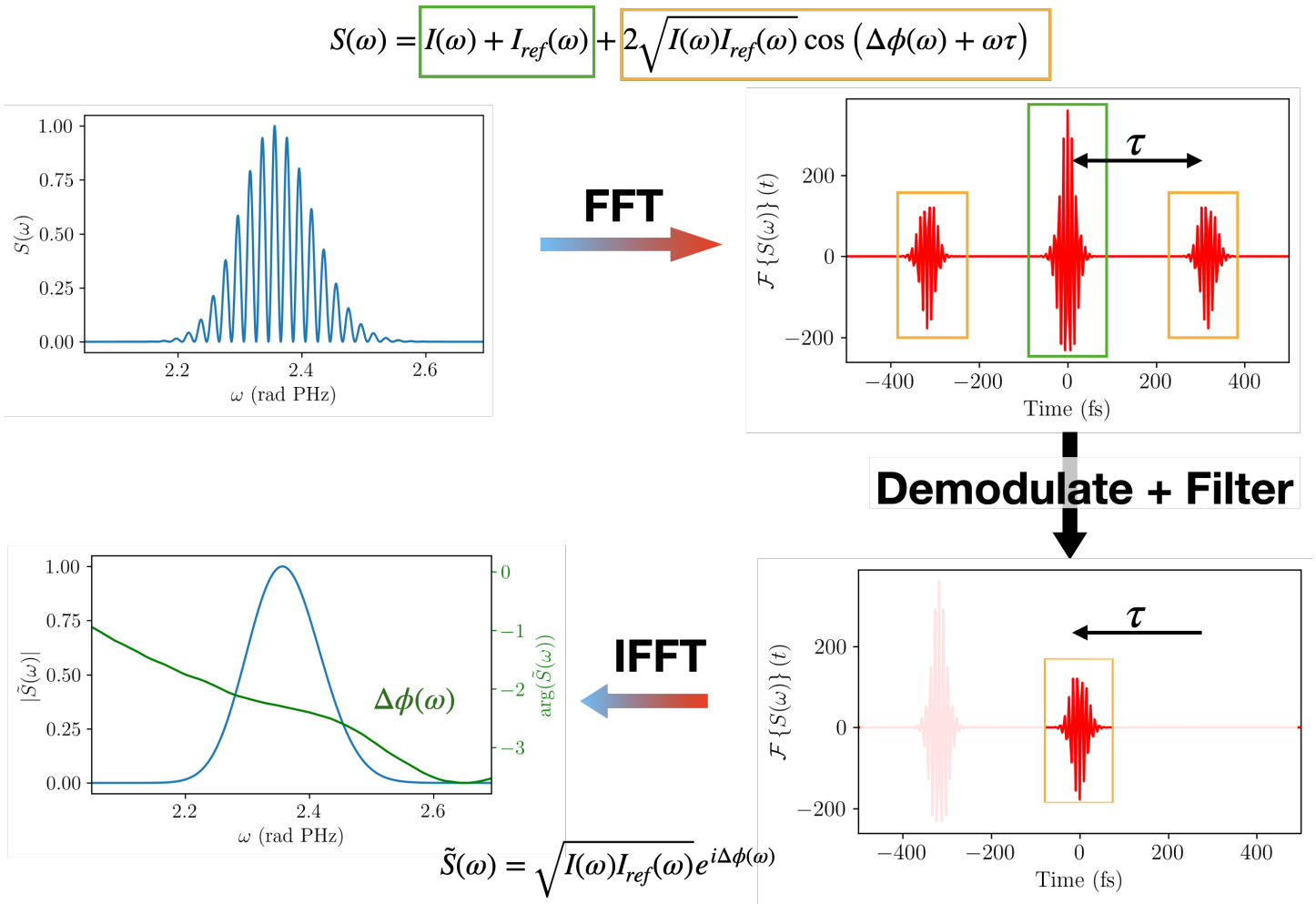


Fig. 2.9 Demonstrating the principle of Fourier filtering. During interferometry, the signal is captured with a modulation term (in this case $\omega\tau$). This allows for the separation of the interference term from the incoherent term in Fourier space. By demodulating (rolling the array), and applying a low pass filter, the term is isolated, before an inverse Fourier transform and the taking of the argument yields the phase difference.

FROG

Autocorrelation is a traditional technique in ultrafast optics used to estimate the duration of an optical pulse. It involves correlating a pulse with itself by splitting it into two identical copies, delaying one with respect to the other, and then measuring the intensity of their interaction as a function of the delay. The simplest form, intensity autocorrelation, is performed by combining the two pulses in a nonlinear medium, such as a second harmonic generation (SHG) crystal, and recording the generated intensity as a function of delay.

While autocorrelation provides an estimate of pulse duration, it does not directly yield information about the spectral phase. In the case of intensity autocorrelation, the measurement is symmetric with respect to delay and requires an assumption about the pulse shape to extract temporal information. This limitation led to the development of more advanced techniques, such as Frequency-Resolved Optical Gating (FROG), which capture both temporal and spectral information of the pulse.

The measurement of the spectrogram, $S(\omega, \tau)$, of a field, $E(t)$, is sufficient to completely determine the field. The spectrogram can be captured by scanning the field with a delay function, $g(t-\tau)$,

$$S(\tau, \lambda) = \left| \int E(t) g(t-\tau) e^{i\omega t} dt \right|^2 . \tag{2.55}$$

Due to the problem that it is not possible to generate a gating function shorter than the duration of the pulse, FROG uses the pulse itself. In the case of the second harmonic generation version of FROG (SHG-FROG), $g(t-\tau) = E(t-\tau)$, which means the above equation has the form of a measurement of the spectrally resolved autocorrelation of the pulse [114]. Experimentally, this autocorrelation is isolated by combining the two pulses in a second harmonic generation crystal.

The resulting spectrogram is symmetric in delay, τ, which as mentioned, means that if the frequency component is not resolved, it is only possible to estimate the pulse duration by making an assumption of the pulse shape, and it is impossible to find the spectral phase. Conversely, resolving the frequency content allows one to reconstruct the whole

pulse. Intuitively, one sees how each delay, τ, corresponds to a different phase shift for each wavelength. Thus, by capturing multiple delays, it is clear that it is possible to determine the relative phase shifts between wavelengths in the initial pulse.

It is important to note that this version of FROG – based on second harmonic generation – inherently possesses a time-reversal ambiguity; the retrieved temporal electric field $E(t)$ and its time-reversed counterpart $E(-t)$ will produce identical FROG traces. This means the absolute time direction of the pulse cannot be determined from the measurement alone, although the pulse shape and phase remain accurately reconstructed. If knowledge of the time direction is essential, one possible solution is to use two FROG diagnostics, placing a known amount of dispersive material in front of one of them. The induced asymmetry (chirp direction) then allows the correct time axis to be inferred by comparing the two reconstructions.

A measurement must resolve in frequency and in time (delays), meaning that this is traditionally a multi-shot technique which requires a scan [85]. However, by using larger detectors, such as a 2D line spectrometer, these two quantities can be measured with fewer shots (or even a single shot, which is the case in the commercialised Grenouille diagnostic).

With spectral phase measurements, one must decide which section of the pulse to measure; i.e. measuring at one spatial location, or integrated over space. A separate spectrally resolved measurement in each location of space will give the pulse duration for each place in space. However, this is not sufficient to obtain the peak intensity of a focused beam, as the relative phase changes between spatial locations are not captured. Subsequently, there will be discussion of spatiotemporally resolved measurements. The majority of these methods cannot resolve the spectral phase, and are therefore complemented by a measurement using one of the devices in this section.

2.4.4 Spatiotemporal Measurements

Spatiotemporal Couplings

The assumption of the separability of the field into spatial and temporal profiles leads to a vastly simpler measurement problem. However, it is now well known that it is not an assumption that holds for ultra-intense lasers, due to the presence of spatiotemporal couplings (STCs) [92]. Here, common chromatic optical elements such as prisms and gratings can impart different spatial phase to each spectral component. The most simple STC is that of pulse front tilt, which can be seen as a spectrally varying spatial tilt in the wavefront of the pulse. To see how this differs to the separable case, first consider introducing a constant tilt across all spectral components, so the field is transformed from $E(x, y, \omega) \rightarrow E(x, y, \omega) \cdot e^{iax}$, where a is some tilt coefficient. By Fourier transforming into the time domain, one sees that this tilt does not affect the intensity through time,

$$\mathcal{F}\left\{E(x, y, \omega) \cdot e^{iax}\right\} = \mathcal{F}\left\{E(x, y, \omega)\right\} \cdot e^{iax} = \tilde{E}(x, y, t) \cdot e^{iax}, \tag{2.56}$$

$$I(x, y, t) = |\tilde{E}(x, y, t) \cdot e^{iax}|^2 = |\tilde{E}(x, y, t)|^2. \tag{2.57}$$

Of course, if this pulse is focused spatially with a standard optic, then the tilt will have an effect.

Now instead consider the addition of a spatiotemporal coupling, a spectrally varying spatial tilt, $E(x, y, \omega) \rightarrow E(x, y, \omega) \cdot e^{i\omega ax}$. By the Fourier shift theorem, this results in a spatial tilt in the resultant temporal intensity profile,

$$\mathcal{F}\left\{E(x, y, \omega) \cdot e^{i\omega ax}\right\} = \int_{-\infty}^{\infty} E(x, y, \omega) e^{i\omega t + \omega ax} d\omega = E(x, y, \omega) e^{i\omega(t+ax)} d\omega \tag{2.58}$$

$$= \tilde{E}(x, y, t + ax) \tag{2.59}$$

$$I_{\text{pft}}(x, y, t) = |\tilde{E}(x, y, t + ax)|^2. \tag{2.60}$$

Thus, the pulse front – the surface of constant intensity – is tilted with respect to the direction of propagation. Upon focusing a pulse such as this, the pulse duration will be elongated and the peak intensity reduced. Pulse front tilt is added upon the propagation of a pulse through a tilted plane of a chromatic medium, such as a glass prism.

Another common STC is pulse front curvature, which instead represents a spectrally varying spatial curvature in the wavefront of the pulse. In the paraxial approximation, a lens can be represented by imparting a curvature onto the wavefront, so this effect can intuitively be caused by a chromatic lens. Again studying the field one sees that now the peak arrival time of the pulse varies quadratically with radial coordinate,

$$I_{\mathrm{pfc}}(r,t) = |\tilde{E}(x,y,t+br^2)|^2, \tag{2.61}$$

where b is some curvature coefficient. Again, if not properly compensated, this will have the effect of diminishing the peak intensity of the focused laser pulse.

One may consider applying the previously described temporal and spatial measurement devices to the pulses described above, for example with pulse front tilt. At each location in space, the temporal duration is the same as without the STC as the pulse has simply been shifted; thus if one were to apply a temporal measurement, even in multiple locations of space, they will not be able to detect this pulse front tilt, as these measurements are not sensitive to the arrival time of the pulse. Similarly, measuring in space would integrate spectrally, and one would simply see the spectrally averaged wavefront. Fig. 2.10 shows the generation of pulse front tilt and curvature, along with attempts to measure them using purely spatial and temporal measurement devices.

Upon attempting to measure the three-dimensional spatiotemporal field, one is confronted by the impediment of the lower dimensionality of sensors. This means that most methods resort to using multiple measurements; that is, measuring the field in two dimensions, whilst scanning through the third. These methods are referred to as *multi-shot methods*. For example, a lot of the spectrally resolved techniques described above can be

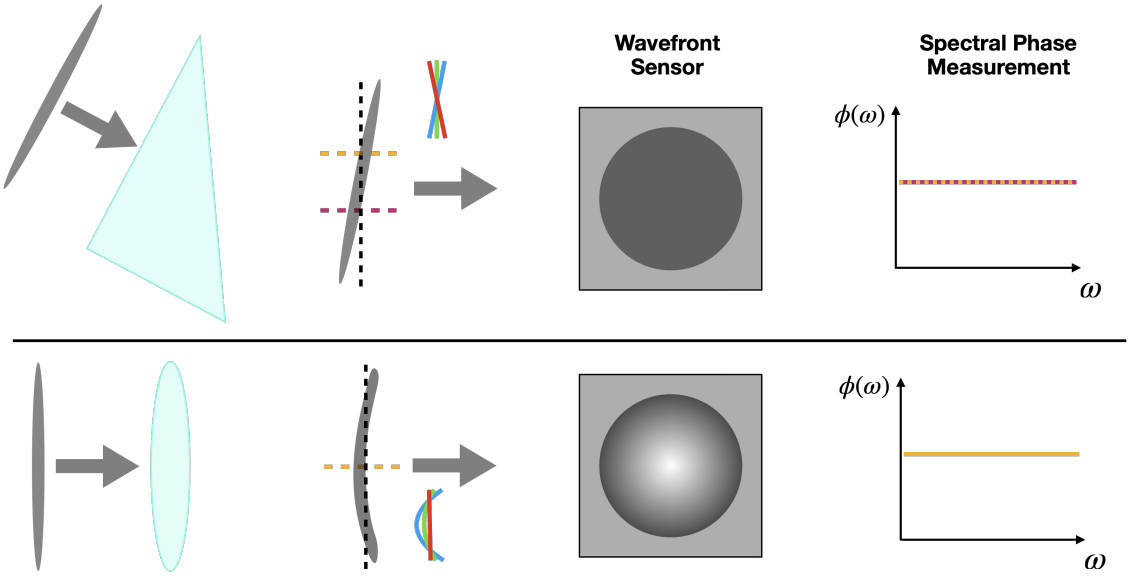


Fig. 2.10 The generation of some common spatiotemporal couplings, showing their invisibility with separate spatial and temporal measurements. The top row displays pulse front tilt, generated from propagation through a wedge. Also shown is the relative tilts of each spectral component. The wavefront sensor measures the spectrally integrated wavefront, and in this case measures a flat wavefront. The spectral phase measurement is blind to arrival time, and so for a transform limited pulse, would measure a flat wavefront even if multiple measurements were performed in different spatial locations. The bottom row shows pulse front curvature, generated from propagation through a chromatic lens. The generation part of this plot is inspired by a similar one in [92].

extended to one spatial dimension by using a camera rather than a spectrometer, but then to capture the full three-dimensional information, the remaining spatial dimension must be scanned through. Alternatively, one may spectrally scan with a spatially resolved method such as the Shack Hartmann.

A subtlety is made in the fact that these methods often resolve a series of 2D objects, which is not necessarily equivalent to the 3D object. For example, the Shack Hartmann sensor encodes the derivative of the phase, and is thus blind to the piston term (which has a derivative of 0). Therefore, one cannot simply combine measurements of the spatial phase for each wavelength, as the absolute difference between wavelengths has been lost. Instead, one must combine these individual measurements for each wavelength, $\phi_\lambda(x, y)$, with a single simultaneous measurement of the *spectral phase*, $\tilde{\phi}(\lambda)$, which is the relative

phase between wavelengths for a single spatial position (or averaged over space),

$$\phi(x, y, \lambda) = \phi_\lambda(x, y) + \tilde{\phi}(\lambda). \tag{2.62}$$

Spatiotemporal Strehl Ratio

The Strehl ratio is a common metric to assess the quality of optical image formation. It can be calculated by finding the peak intensity from an aberrated image of a point source, and dividing it by the maximum intensity from an image from an ideal optical system limited only by diffraction. Thus, it is a measure of the degradation of the peak intensity due to the aberrations in the system. The original definition of the Strehl ratio considers this process in space only, so it is defined as,

$$S = \frac{\max\left[I_{aberrated}(x, y)\right]}{\max\left[I_{ideal}(x, y)\right]}. \tag{2.63}$$

In the case of monochromatic light, it is the spatial wavefront (i.e., as measured by a Shack Hartmann sensor) that determines the Strehl ratio, and the ideal pulse will have a flat wavefront.

The fact that it quantifies the degradation in peak intensity due to aberrations makes it interesting for the assessment of ultra-intense laser pulses. However, the broad spectral bandwidth of these pulses means that a slightly different quantity must be evaluated instead – in particular, the pulse must be considered through time as well as space, giving rise to the spatiotemporal Strehl ratio. This ratio is calculated in the same way, except it also considers the focused pulse through time,

$$S = \frac{\max\left[I_{aberrated}(x, y, t)\right]}{\max\left[I_{ideal}(x, y, t)\right]}. \tag{2.64}$$

In the following methods, the pulse's spatiotemporal intensity and phase is measured in the near-field, so it must then be numerically propagated to the focus to calculate the spatiotemporal Strehl ratio.

A subtlety comes in what is defined as the aberrated pulse in this spatiotemporal setting. In this thesis, the *absolute ratio* is calculated by simply treating the 'aberrated pulse' as the measured spatiotemporal field and propagating to the focus. However, another interesting idea is to consider the spatiotemporal Strehl ratio that is purely due to spatiotemporal couplings [115] – the *STC-isolated ratio*. This represents the case where one has performed separate spatial and temporal measurements of the pulse, and has made perfect separate spatial and temporal corrections; therefore any further reduction in peak intensity is due to spatiotemporal couplings (which are not as easily detectable). In this setting, the 'aberrated pulse' is found by taking the measured spatiotemporal wavefront and making the spectrally-averaged wavefront flat (eg. by a deformable mirror), as well as setting the spatially-averaged spectral phase equal to 0 (which could be achieved by a DAZZLER), before numerically propagating to focus.

To understand the difference between these ratios clearly, consider the complete measured spatiotemporal phase, $\phi(x, y, \lambda)$. It can be expressed in the following way,

$$\phi(x, y, \lambda) = \phi_{space}(x, y) + \phi_{spectral}(\lambda) + \phi_{STC}(x, y, \lambda), \tag{2.65}$$

where ϕ_{space} is the spectrally-averaged spatial phase, $\phi_{spectral}$ is the spatially-averaged spectral phase, and ϕ_{STC} is the remaining phase (due to spatiotemporal couplings). A pulse that was both chirped and containing pulse front curvature could be expressed like this, with ϕ_{space} containing the average curvature, $\phi_{spectral}$ containing the chirp and ϕ_{STC} containing the spectral variations of the spatial curvature. The *absolute ratio* will use the total complete phase, ϕ, in its calculation of the spatiotemporal Strehl ratio, where as the *STC-isolated ratio* will set the phase of the near-field to ϕ_{STC} before numerically propagating to focus.

Multi-Shot Methods

Here will be detailed some of the most popular methods that are implemented in major laboratories. Each of these methods describes either a spectral measurement performed in multiple spatial locations, or a spatial measurement performed in multiple spectral locations. A technique, here referred to as TERMITES, that fits into the former category was developed simultaneously by two groups, and works on the basis of imaging Fourier transform spectroscopy (FTS) [40, 69]. In the most general case, this technique involves taking many images of the interference between two pulses, E_1 and E_2, whilst the delay between them, τ, is scanned. The captured signal will have the form,

$$I(\vec{r},\tau) = \int |E_1(\vec{r},t) + E_2(\vec{r},t-\tau)|^2 dt, \tag{2.66}$$

$$= I_1(\vec{r}) + I_2(\vec{r}) + \int \left[E_1(\vec{r},t)E_2^*(\vec{r},t-\tau)\right] dt + \int \left[E_1^*(\vec{r},t)E_2(\vec{r},t-\tau)\right] dt. \tag{2.67}$$

Considering the form of one of the coherent terms (terms three and four), one sees that they are modulated by a carrier frequency, $\int E_1^*(\vec{r},t)E_2(\vec{r},t-\tau)dt \propto e^{i\omega_0\tau}$. This means that by taking the Fourier transform of $I(\vec{r},\tau)$, these terms are separated from the incoherent terms in Fourier space. Applying that if $\mathcal{F}(f(t)) = \tilde{f}(\omega)$, then $\mathcal{F}(f^*(t)) = \tilde{f}^*(-\omega)$, one obtains,

$$\mathcal{F}_{\tau\to\omega}(I(\vec{r},\tau)) = I_1(\vec{r}) + I_2(\vec{r}) + \left[\tilde{E}_1 \cdot \tilde{E}_2^*\right](\vec{r},\omega) + \left[\tilde{E}_1^* \cdot \tilde{E}_2\right](\vec{r},-\omega) \tag{2.68}$$

Now considering the form of one of the coherent terms once more,

$$A(\vec{r},\omega) = \left[\tilde{E}_1 \cdot \tilde{E}_2^*\right](\vec{r},\omega) = \|\tilde{E}_1(\vec{r},\omega)\|\|\tilde{E}_2(\vec{r},\omega)\|e^{i[\phi_1(\vec{r},\omega)-\phi_2(\vec{r},\omega)]}, \tag{2.69}$$

one sees that the relative phase difference is encoded. In traditional FTS, it is typical to use the same beam for both, $E_1 = E_2$, which then results in A being equal to the intensity

(this is used for hyperspectral imaging). However, if one uses a known reference for E_2, the spatiotemporal wavefront of the unknown beam can also be found, $\phi_1 = \mathrm{ARG}(A) + \phi_2$.

In TERMITES, and as was the case in SRSI, the reference is generated from the original pulse. While in an ideal case, the beam is tightly spatially filtered to create a reference free of STCs [69], this is practically difficult for large-diameter ultra-intense beams. Thus, in the TERMITES implementation [40], a small section of the beam is taken and is diverged to properly overlap the other copy. This is similar to the principle of radial shearing interferometry [116], which is also related to LSI except instead of shearing laterally, the beam is enlarged radially. This is shown in Fig. 2.11. As the small section of the beam is not necessarily free of STCs, an iterative algorithm is used to reconstruct the field. Seeing as both copies are generated from the same beam, this method is insensitive to the spectral phase, and thus a further temporal measurement is still required. In the practical implementation, a Michelson interferometer is used with a delay stage to achieve the scan.

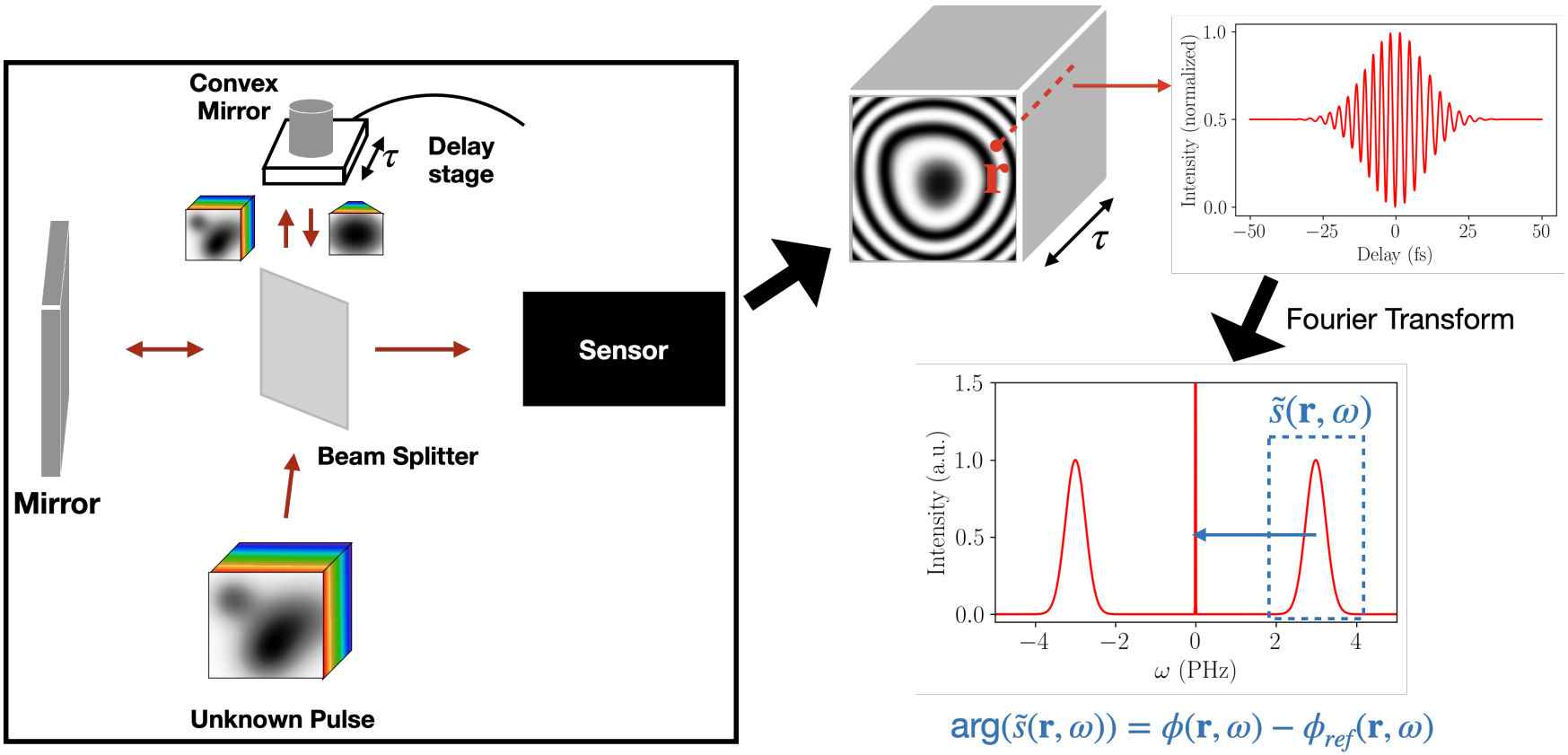


Fig. 2.11 Capturing the spatiotemporal structure of a pulse using Fourier transform spectroscopy, as in the TERMITES method [40]. A reference is generated by a convex mirror, which samples a small portion of the beam. This is interfered with the original pulse, whilst a stage scans through delays. For each spatial location of the resulting intensity cube, a Fourier transform is taken. Then Fourier filtering is used to isolate the modulated peak (corresponding to the interference). The argument yields the phase difference between the original and reference pulses.

Due to the use of the Fourier transform, the Nyquist criterion relates the range in one domain and the resolution in the other domain. Thus, in order to resolve the spectral bandwidth of the pulse (which is modulated by some frequency ω_0), the resolution of the sensor must be necessarily fine. Then, it is the range that is scanned through which determines the resolution by which the spectrum is captured.

INSIGHT [117] is related technique, developed in the same lab, but it operates in the far field, directly resolving the pulse at focus. This technique is essentially a spectrally resolved version of the Gerchberg-Saxton algorithm. Spatially resolved Fourier transform spectroscopy is again used in order to resolve the spectral information. However, this time the phase information is obtained differently to TERMITES (which used a reference generated from the unknown pulse). Here, the hyperspectral intensity is obtained in three separate planes, one of which is at focus, and the other two are at a small offset $\pm\Delta z$. A Gerchberg-Saxton algorithm is then used to recover the field for each spectral component. Both TERMITES and INSIGHT have been applied to measuring PW laser pulses at the BELLA laser facility [118], where it was found that the STC-isolated spatiotemporal Strehl ratio (intensity reduction due to spatiotemporal couplings) was 85% and 92% respectively. As the hyperspectral intensity is captured in three separate planes, INSIGHT requires a large number of laser shots or a setup containing multiple sensors.

Another technique that deserves mention is FALCON [119], developed by colleagues in Munich during the duration of this thesis. This is a method that begins to reduce the number of shots (and therefore scanning locations) that are required to resolve the field. Hardware-wise, it consists of a motorised wheel on which there are 9 narrowband pass filters. Therefore, the laser is measured nine times, each time with a different spectral band. This represents far fewer measurements than the previously mentioned techniques. The raw data from FALCON gives access to the wavefront derivatives for each of the nine spectral locations. It should also be said that this method is similar to the HAMSTER technique [120], where the spectral component is selected using an acouto-optic dispersive filter (AOPDF) instead. Key to the success of FALCON however, is its reconstruction

algorithm, which can be seen as the hyperspectral extension of the phase-stitching methods described in Section 2.4.2. It is discussed in detail in Section 2.4.6.

Single-Shot Methods

As was described in the introduction, there are many reasons to desire a single shot characterisation technique, where two of the most important are to detect shot-to-shot fluctuations, and to act as a diagnostic to enable spatiotemporal beam shaping. At the time of the start of this project, there existed no single shot technique to completely characterise ultra-intense (petawatt power) light. While the STRIPED-FISH method has existed for a long time [121], and is still being actively developed, current models are not able to resolve the spectral range required to measure ultrashort pulses and are limited to TW powers [122]. Simultaneously with the development of the content of this thesis, some other groups also attempted to work on this problem. For example, Kim *et al* developed a method to resolve the spatiospectral wavefront using a multispectral camera, but their work was limited to just 4 spectral channels [123]. Another limitation of this method, was that it used point samples spectrally rather than averaging, meaning it lacked robustness to noise. Another method, similar to the one developed in Chapter 3 of this thesis, is termed single-shot compressed optical field topography [124], and uses machine learning techniques to reconstruct the hyperspectral intensity in multiple planes before using a Gerchberg-Saxton algorithm to recover the phase. There is further discussion of this method also in Chapter 3. Neither of the aforementioned methods are sensitive to the polarisation structure of the pulse.

2.4.5 Vector Field Measurements

To obtain the full vector field, one must also consider the coupling in the polarisation dimension, which gives rise to the Stokes' parameters. Ultra-intense lasers are currently linearly polarised, but in recent times there has been an expanding literature on the effects

of vectorial light in laser-matter interactions. Many recently proposed acceleration schemes depend on the polarisation of the light, including the use of circular [125–127] and radial [65, 128–130] polarisations. In order for these effects to be realised experimentally in the future, it is necessary to also incorporate a sensitivity to this structure in a measurement device. Just as in the case of the spectrum, it is not sufficient to simply measure the individual intensities and phases for each of the two orthogonal polarisations independently, for the same reason that one requires the relative difference between them in order to resolve the coupling (i.e. the analogue of the spectral phase).

One method has been developed to capture the full vector field of pulses generated by a laser system [70]. This method used FTS to resolve spatiotemporally, by interfering the pulse with a known reference, while scanning through temporal delays. It extends the measurement to capture the polarisation structure by first also scanning through rotations of a linear polariser in order to capture two orthogonal polarisations, $\in [0°, 90°]$, and the interference between them by setting the angle of the polariser to 45°. This enabled the extraction of the Stokes parameters up to an ambiguity in handedness, and this was removed by deliberately manipulating the reference beam with a birefringent medium. Evidently, in this formulation the inclusion of the polarisation measurement increases the number of required measurements by a factor of $3\times$, making an already arduous process even more so.

Finally, there has also been an approach that captures the hyperspectral polarisation and intensity information in a single shot, but this does not include the phase [131]. There exists no single shot method to characterise the vector field of ultra-intense laser pulses.

2.4.6 Modal Spatio-Spectral Wavefront Retrieval

For some of the experiments conducted in the following chapters, it is useful to extract the coefficients of spatiotemporal couplings from the spatiospectral wavefront. Weiße *et al.* introduced a method for this purpose [132].

This method is the hyperspectral extension of the standard modal reconstruction method to wavefront reconstruction. In the monochromatic case, one can fit coefficients that multiply by the derivatives of modes of the Zernike basis, in order to minimise the squared error with the measured derivatives. In the hyperspectral case, additional modes are introduced that vary spectrally, forming the 'Zernike-Taylor' basis. A mode has three indices, corresponding to the Zernike mode, and the spectral dependence,

$$A_n^{m,j}(x,y,\lambda) = \left(\frac{(\lambda-\lambda_0)}{(\lambda_{\max}-\lambda_{\min})}\right)^j Z_n^m(x,y). \tag{2.70}$$

Most common STCs involve only a linear relation with spectra, for example, pulse front tilt (PFT) and pulse front curvature (PFC). For example, a mode to be fit which is related to PFT:

$$A_1^{1,1}(x,y,\lambda) = \frac{(\lambda-\lambda_0)}{(\lambda_{\max}-\lambda_{\min})} Z_1^1(x,y). \tag{2.71}$$

The spatiospectral wavefront is represented as the sum of these modes and coefficients,

$$\phi(x,y,\lambda) = \sum_i \sum_n \sum_{m=0}^{n} C_n^{m,i} A_n^{m,i}(x,y,\lambda). \tag{2.72}$$

The $Z_1^{\pm 1}$ Zernike modes describe the phase shift due to the tip or tilt respectively. In the implementation here, the modes are defined on a grid of extent: $x \in [-1,1]$ and $Z_1^1 = x$. In a monochromatic case, where one fits a coefficient, c, for Z_1^1, the edge-to-edge phase shift due to the tilt is $\Delta\phi = 2c$. To get the corresponding tilt angle, θ, one converts this phase shift into spatial units and uses trigonometry,

$$\theta = \arctan\left(\frac{\lambda_0}{2\pi}\frac{2c}{D}\right) \approx \frac{\lambda_0}{\pi}\frac{c}{D}, \tag{2.73}$$

where D is the diameter of the beam. Now one may wish to convert this formalism to the hyperspectral case, to find how this tilt angle varies with wavelength, causing

pulse front tilt. Using Eq. (2.71), one can find the phase shifts for each wavelength ($\Delta\phi(\lambda) = 2C_1^{1,1}\frac{(\lambda-\lambda_0)}{(\lambda_{\max}-\lambda_{\min})}$), and can then calculate the tilt as above. Then, defining pulse front tilt as $\alpha = \frac{\partial\theta}{\partial\lambda}$, one finds this quantity, given in units of rad nm^{-1},

$$\alpha = \frac{\lambda_0}{\pi D}\frac{C_1^{1,1}}{(\lambda_{\max}^{\mathrm{nm}} - \lambda_{\min}^{\mathrm{nm}})}. \tag{2.74}$$

Another common STC is pulse front curvature, which is related to the spectral variation of the defocus Zernike mode, Z_2^0. Its coefficient is perhaps not as intuitive as that of pulse front tilt; here it is simply defined in the same way, and measured in units of rad^2 nm^{-1} (as it represents a changing curvature per wavelength):

$$\gamma = \frac{\lambda_0}{\pi D}\frac{C_2^{0,1}}{(\lambda_{\max}^{\mathrm{nm}} - \lambda_{\min}^{\mathrm{nm}})}. \tag{2.75}$$

Other work may define this coefficient in different ways, but the important point here is the linear relationship between the defocus mode and the wavelength.

2.4.7 Intensity Reduction due to PFT

For some STCs, the reduction in peak intensity can be estimated using an analytical approach. To estimate the effect of pulse front tilt on the peak intensity, one begins by considering the flat-top spatial profile of the near field, and the Airy-disk like pattern in the far field. To maintain analytical tractability, this Airy disk is approximated by a Gaussian [133], using a width of,

$$w_{far} = 0.84\lambda F, \tag{2.76}$$

where $F = \frac{f}{D}$ is the F-number of the system, f is the focal length, and D is the diameter of the flat-top in the near field.

Also assuming a Gaussian distribution on the spectra, with a width of w_λ, the intensity in the far field can be found as,

$$I_{\text{far}}(x,\lambda) \propto \exp\left(-2\left[\frac{\lambda-\lambda_0}{w_\lambda}\right]^2\right)\exp\left(-2\left[\frac{x}{w_{far}}\right]^2\right). \tag{2.77}$$

A pulse front tilt describes a structure where each spectral component propagates at a tilt relative to the others. It can be represented by a coefficient α, describing the change of this tilt angle with wavelength (thus measured in units of µrad/nm). Due to this tilt, each spectral component focuses in a slightly different position in space, and thus introduces a spectrally varying centre of the spatial Gaussian. This spatial shift for each wavelength, $\Delta x(\lambda)$ is calculated by trigonometry with $\tan(\alpha\cdot(\lambda-\lambda_0)) = \frac{\Delta x}{f}$, and thus under the small angle approximation, $\Delta x = f\alpha(\lambda-\lambda_0)$. It follows that the equation for the intensity of the field at focus, with a pulse front tilt present, is,

$$I_{\text{far}}^{\text{PFT}}(x,\lambda) \propto \exp\left(-2\left[\frac{\lambda-\lambda_0}{w_\lambda}\right]^2\right)\exp\left(-2\left[\frac{x-f\alpha(\lambda-\lambda_0)}{w_{far}}\right]^2\right). \tag{2.78}$$

Setting $x=0$, one finds that the resulting spectral intensity also approximates a Gaussian, with a new effective width ω_e,

$$I_{\text{far}}^{\text{PFT}}(\lambda) \propto \exp\left(-2\left[\frac{\lambda-\lambda_0}{w_\lambda}\right]^2 - 2\left[\frac{-f\alpha(\lambda-\lambda_0)}{w_{far}}\right]^2\right) \tag{2.79}$$

$$\propto \exp\left(-2\left[\frac{\lambda-\lambda_0}{w_e}\right]^2\right), \tag{2.80}$$

where $w_e = \frac{w_\lambda w_{far}}{\sqrt{w_{far}^2+f^2\alpha^2 w_\lambda^2}}$. Under this approximation ($w_{far}\approx 0.84\lambda_0 F$), one assumes the time-bandwidth product for Gaussians still applies [134], $\Delta_\lambda\Delta_t \approx 0.44$, and writes the ratio of the pulse durations for the pulse with PFT, and for the pulse without, as

$$\chi_{\text{duration}} = \frac{w_\lambda}{w_e} = \sqrt{1+\left(\frac{f\alpha w_\lambda}{w_{far}}\right)^2} = \sqrt{1+\left(\frac{\alpha w_\lambda D}{0.84\lambda_0}\right)^2}. \tag{2.81}$$

The resulting peak intensity reduction is found as the inverse of the square of this ratio of pulse durations,

$$\chi_{\text{intensity}} = \frac{1}{1 + \left(\frac{\alpha w_\lambda D}{0.84\lambda_0}\right)^2}. \tag{2.82}$$

2.4.8 Nonlinear Effects

The extremity of ultra-intense laser pulses is such that nonlinear optics become relevant when passing through optics such as lenses, or even glass windows. A particularly relevant example of this is the optical Kerr effect, where ultra-intense light causes a variation in the index of refraction of a medium, that is proportional to the local intensity of light. As a result, the laser pulse accumulates a nonlinear phase shift as it propagates; this is the B-integral, which is defined as,

$$B = \frac{2\pi}{\lambda} \int n_2 I(z) dz, \tag{2.83}$$

where $I(z)$ is the intensity along the beam propagation direction, z, and n_2 is the nonlinear index so that $n_2 I$ describes the nonlinear change in refractive index.

This creates a complication when attempting to take a reference measurement. In metrology, one normally has a small optical setup to transport the beam to the measurement apparatus. As this optical equipment is not a part of the main experimental setup (that will be used for e.g. particle acceleration), it is desirable to remove any aberrations caused by it in the final measurement. An often used approach to this is to use a spatially filtered beam to first take a reference measurement, which then has a wavefront containing the aberrations of the optical system, due to things such as glass windows. Crucially, in the linear regime this wavefront will not depend on the beam itself. This can then be subtracted from a subsequent measurement to remove the effects of the beam-carrying optical setup.

The difficulty introduced by nonlinear effects is that the changes to the wavefront depend on the beam itself, so this spatially-filtered referencing technique will no longer work. In the experiments of Chapter 5, an assessment of the importance of nonlinear effects is undertaken by assessing the B-integral.

2.5 Inverse problems

Chapter 3 is largely about reconstructing high dimensional data, based on a single capture of a two-dimensional sensor. This problem is described as an inverse problem. A linear system is as follows,

$$\vec{n} = \mathbf{\Phi}\vec{m}, \tag{2.84}$$

where $\vec{m}$ and $\vec{n}$ describe the vectorised initial and final state respectively, and $\mathbf{\Phi}$ is the forward model (an $a \times b$ matrix). An inverse problem is then defined as task of recovering $\vec{m}$, given $\vec{n}$ and $\mathbf{\Phi}$. The level of difficulty of this problem is highly dependent on the determinacy of the forward matrix. In the case that $a \geq b$, the system is determined (or even over-determined), and therefore there exists only one possible initial state that satisfies the equation. Here, the system is solved by simple matrix inversion (or the pseudo inverse).

However, in the case that $b > a$, the system is underdetermined; accordingly, there are unlimited possibilities for the initial state that produces the final state. To see this clearly, consider that $\mathbf{\Phi}$ is simply a 1×2 matrix,

$$n = \begin{bmatrix} \Phi_0 & \Phi_1 \end{bmatrix} \begin{bmatrix} m_0 \\ m_1 \end{bmatrix}. \tag{2.85}$$

Clearly the coordinates of a line in two-dimensional space describe the possible solutions for $\vec{m}$. Therefore, one desires an approach that can select the correct solution out of the possible ones. This problem can be approached from two complementary directions.

Firstly, in the reconstruction process, one can use a regularisation term that biases the predicted solution to a restricted solution space. The other approach, termed *compressed sensing*, describes the deliberate designing of the forward matrix, so that information is captured in an optimal, efficient way.

2.5.1 Compressed Sensing

Compressed sensing is a crucial part of the method developed in Chapter 3. It is a technique designed to achieve the efficient capture of a sparse signal. This section serves as an introduction to the theory, including the fundamental conditions of sparsity and incoherence [135].

To demonstrate the intuitive way the principles of compressed sensing, consider a spectrum, $\vec{x}$, that one would like to capture, containing 300 spectral bands from 400-700 nm. However, the detector (a single camera pixel), is only able to measure the integrated spectra. This is, of course, analagous to the problem of capturing a hyperspectral image on a two-dimensional sensor. Before the camera captures the signal, one applies an operation to it; for example, the application of a filter which will only let through certain wavelengths. A single capture is then represented by the dot product of a measurement vector, $\vec{\Phi}_i$, with $\vec{x}$;

$$y = \vec{\Phi}_i \cdot \vec{x}. \tag{2.86}$$

For example, the application of a narrow bandpass filter yields the measurement of one band of the spectra;

$$y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot \begin{bmatrix} x_{400} \\ x_{401} \\ x_{402} \\ \vdots \\ x_{700} \end{bmatrix} = x_{401}. \tag{2.87}$$

Multiple measurements can be represented by using a matrix $\mathbf{\Phi}$. For example, two measurements, each with a different filter;

$$\begin{aligned}\vec{y} &= \mathbf{\Phi}\vec{x} \\ &= \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} x_{400} \\ x_{401} \\ x_{402} \\ \vdots \\ x_{700} \end{bmatrix} = \begin{bmatrix} x_{401} \\ x_{402} \end{bmatrix}.\end{aligned}$$

In the case that one has no prior knowledge about $\vec{x}$, setting the measurement matrix to the identity basis, and thus taking 300 measurements, is maximally efficient. However, as shall be seen, the prior condition of sparsity allows for the capture of $\vec{x}$ with far fewer measurements.

Sparsity

The first condition of compressed sensing is that the signal must be sparse in some domain. This is represented with a basis-transform matrix, $\mathbf{\Psi}$ and the signal's sparse representation in that basis, $\vec{f}$; $\vec{x} = \mathbf{\Psi}\vec{f}$. The measurement can also be written in terms of the sparse representation:

$$\vec{y} = (\mathbf{\Phi}\mathbf{\Psi})\vec{f}, \tag{2.88}$$

where $(\mathbf{\Phi}\mathbf{\Psi})$ is referred to as the effective sampling matrix. The condition of sparsity is intuitive; if it is possible to capture the signal in fewer measurements than n, one expects there to be redundancy.

Consider that one is given the information that when $\vec{x}$ is transformed from some specific basis, $\mathbf{\Psi}$, it is expressed by $\vec{f}$ that contains a single non-zero element. This prior

knowledge of sparsity enables one to form a sampling matrix that captures the information of $\vec{x}$ in just 2 measurements. For example, using the following form of $(\mathbf{\Phi\Psi})$,

$$\vec{y} = (\mathbf{\Phi\Psi})\vec{f} \tag{2.89}$$

$$= \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 3 & \dots & 300 \end{bmatrix} \begin{bmatrix} f_{400} \\ f_{401} \\ f_{402} \\ \vdots \\ f_{700} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \end{bmatrix}, \tag{2.90}$$

one is able to obtain the magnitude of the non-zero component by y_0, and its wavelength index by $\frac{y_1}{y_0}$.

This trivial example demonstrates that knowledge of sparsity leads to a much more efficient measurement process. In a more typical case however, the knowledge of sparsity does not extend to knowing exactly how many non-zero elements are in $\vec{f}$. The question of how to choose an optimal sampling matrix in this setting will be the next topic of discussion.

Incoherence

To understand what makes the ideal sensing matrix, consider the following points. In order to reliably recover the sparse signal, the effective sampling matrix must sample the full domain of $\vec{f}$, for the simple reason that if the non-zero elements are outside this region, they are invisible. Furthermore, it is desirable for each measurement to sample as much of $\vec{f}$ as is possible, as this provides more terms in the equation system. For an example of an especially bad sensing matrix, if one uses rows of sparse basis for the sampling matrix, $(\mathbf{\Phi\Psi})$ contains zeros except for 1s down the diagonal, thereby each measurement samples a single element of the sparse vector. Thus, the rows of $\mathbf{\Phi}$ must be *uncorrelated* with the

rows of $\mathbf{\Psi}$. As coherence is defined as the maximum correlation between rows of $\mathbf{\Phi}$ and $\mathbf{\Psi}$, this condition is equivalent to maximising incoherence:

$$C(\Psi, \Phi) = \sqrt{n} \max_{1 \leq i,j \leq n} |\Phi_i \cdot \Psi_j| \tag{2.91}$$

One further condition here, is that one requires each measurement to be uncorrelated with respect to all other measurements, otherwise this leads to redundancy; therefore, the following quantity also must be minimised,

$$\tilde{C}(\Phi) = \max_{i \neq j} |\Phi_i \cdot \Phi_j| \,. \tag{2.92}$$

In a more general case, one does not know specifically what basis the signal is most sparse in, although it is assumed that it is sparse in *some* basis (a fair assumption for most signals). Thus, one cannot formulate the minimisation equation in Eq. (2.93), and one cannot handcraft a sensing matrix to maximise incoherence. In this case, one employs *random* sampling, which approximately satisfies the incoherence condition, and then solves for the signal, $\vec{x}$, using iterative methods or machine learning approaches.

L1 Regularisation

In traditional compressed sensing, that is where one knows the form of $\mathbf{\Psi}$, the optimum solution is described by the minimisation of the L0 norm, which counts the number of non-zero elements:

$$\min_{\tilde{f}} |\tilde{\vec{f}}|_0 \;\; \text{subject to} \;\; y = (\mathbf{\Phi\Psi})\tilde{\vec{f}} \tag{2.93}$$

This is a non-convex function without smooth gradients, and is thus very difficult to optimise. However, minimising the L1 norm leads to the same result in many situations (See Appendix A), so the equation that is actually solved is,

$$\min_{\tilde{f}} |\tilde{\vec{f}}|_1 \;\; \text{subject to} \;\; y = (\mathbf{\Phi\Psi})\tilde{\vec{f}}, \tag{2.94}$$

where $|a|_1 = \sum_i |a_i|$. There are many methods designed to solve Eq. (2.94); for example the family of iterative shrinkage-thresholding algorithms (ISTA). Appendix A also describes a specific method for this problem named Lasso, and its implementation in the `sci-kit learn` library.

2.6 Machine Learning

Machine learning forms a pillar that is used throughout this thesis. Commonly split into the three categories of supervised, unsupervised and reinforcement learning, the methods used in this thesis typically belong to the first category. Supervised learning describes the learning of a function mapping inputs x to outputs y, given a dataset of input-output pairs. Practically, the function is learned by the tuning of parameters, θ, in order to minimise a loss function, $L(\theta)$. While supervised learning is vastly dominated by neural networks, any parameterised differentiable function can be used.

One must decide on a choice for the loss function, and an algorithm for how to use this loss to adjust the network parameters. For tasks involving regression, the most commonly used loss function is the mean squared error (MSE), which is defined for a prediction, $\hat{y}$, and a ground truth, y, as,

$$\text{MSE}(y, \hat{y}(\theta)) = \frac{1}{N} \sum_{n}^{N} |y_n - \hat{y}_n(\theta)|^2. \tag{2.95}$$

An intuitive way to adjust the parameters is stochastic gradient descent (SGD), where the parameters are adjusted in the opposing direction of the gradient with respect to the loss function, scaled by a learning rate, η. The update rule to a vector of parameters, θ_t, is then,

$$\text{SGD:} \quad \theta_{t+1} = \theta_t - \eta \nabla L(\theta_t). \tag{2.96}$$

The stochasticity comes from the fact that the gradient is calculated on a subset of the data, rather than the full dataset. One natural question arises on how to efficiently calculate the gradients of the loss with respect to the parameters.

In machine learning this is done in the nature of dynamic programming in a method called *backpropagation*, which involves successive applications of the chain rule [136]. For a simple explanation, consider a system which has two trainable matrices, $\mathbf{A}$ and $\mathbf{B}$, which are successively applied to an input, x, along with a nonlinear function f, to give a prediction $\hat{y}$,

$$\hat{y} = f(\mathbf{B} \cdot f(\mathbf{A}x)). \tag{2.97}$$

For clarity this process is divided into intermediate steps, so that $a_0 = \mathbf{A}x$, $z_0 = f(a_0)$, $a_1 = \mathbf{B}z_0$ and $\hat{y} = f(a_1)$. A loss function, L, is used to evaluate the performance compared to some ground truth, y. The task is to find the derivatives of the loss function with respect to the matrices of trainable parameters, $\mathbf{A}$ and $\mathbf{B}$. Initially, one uses the chain rule to evaluate the gradient of the loss function with respect to the parameters in the matrix, $\mathbf{B}$,

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{B}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \mathbf{B}} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a_1} \cdot \frac{\partial a_1}{\partial \mathbf{B}} \\ &= \frac{\partial L}{\partial \hat{y}} \cdot f'(a_1) \cdot z_0\end{aligned}$$

SGD is then used to adjust these parameters. In order to calculate the gradient of the loss with respect to the parameters in matrix $\mathbf{A}$, one simply applies the chain rule once more,

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{A}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \mathbf{A}} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a_1} \cdot \frac{\partial a_1}{\partial \mathbf{A}} \\ &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_0} \cdot \frac{\partial z_0}{\partial a_0} \cdot \frac{\partial a_0}{\partial \mathbf{A}} \\ &= \frac{\partial L}{\partial \hat{y}} \cdot f'(a_1) \cdot \mathbf{B} \cdot f'(a_0) \cdot x.\end{aligned}$$

One sees that the derivative with respect to the earlier parameter **A** also depends on the value, $f'(a_1)$, which was already calculated for $\frac{\partial L}{\partial \mathbf{B}}$. This demonstrates the computational efficiency of backpropagation; updating the final parameters first before propagating backwards through the network. It is clear that in order for backpropagation to be applicable, it is essential that all operations are differentiable. The above rules generalise to many layers (N), where layer n has parameters $W^{(n)}$, as,

$$\frac{\partial L}{\partial W^{(n)}} = \left(\prod_{k=n}^{N} J^n \right) \frac{\partial L}{\partial \hat{y}}, \tag{2.98}$$

where $J^n = f'(a_{n-1})W^{(n)}$.

Optimisation

For a convex problem, gradient descent converges to the optimum (with a suitable learning rate). However, machine learning is typically employed in non-convex settings. In this case, the loss is often thought of as a landscape, with many peaks and troughs. Troughs which are not the absolute minimum of the loss are then referred to as local minima, and SGD may get stuck in these (as $\nabla L(\theta_t) = 0$). In order to navigate this loss landscape, one can employ more sophisticated approaches to the optimisation of parameters. Cited nearly 200,000 times, perhaps the most commonly used optimiser is referred to as Adaptive Moment Estimation (Adam) [137], and this is what was used for the majority of the work in this thesis.

To explain the principles behind Adam, two further concepts must be explained. Firstly an intuitive simple addition, momentum. Here, one keeps an exponential moving average of the gradients (over successive time steps), and uses this to update the parameters, thus giving the optimiser the ability to see part of the loss landscape past local minima, and to pass through plateaus. A second addition is known as adaptive gradient updates (AdaGrad) [138]. This method is designed to solve the problem that some parameters in the model will receive much larger gradients than others. Thus, with a constant learning rate the

training of these parameters will be much faster than others. Consider a two-dimensional loss landscape, where the loss varies slowly with respect to the first dimension, but quickly with respect to the second. Using a constant learning rate will either result in moving through the first dimension too slowly, or through the second dimension too quickly. This problem is solved by weighting the parameter updates with a measure of this variance, effectively introducing individual learning rates for each parameter, based on the historical gradients, allowing parameters with smaller gradients to be trained much faster. The Adam optimiser essentially combines these two techniques, and the total learning algorithm is as follows. Note that there is one additional step of un-biasing the updates, due to the initial values of the moving averages being zero, but that is not conceptually important here.

Algorithm 1 Adam Optimisation Algorithm

Input: Learning rate η, exponential decay rates β_1, β_2, small constant ϵ, initial parameters θ_0
Initialise $m_0 = 0$, $v_0 = 0$, $t = 0$
while not converged **do**
    $t \leftarrow t + 1$
    Compute gradient $g_t = \nabla_\theta f_t(\theta_{t-1})$
    $m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t$ ▷ Update biased first moment estimate
    $v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$ ▷ Update biased second moment estimate
    $\theta_t \leftarrow \theta_{t-1} - \eta \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$ ▷ Update parameters
end while

The hyperparameters β_1 and β_2 are the exponential decay rates for the moving averages of the gradient and the squared gradient respectively. The ϵ term is a small number to prevent division by zero.

Regularisation

Another commonly used technique is adding regularisation to the loss function, which can allow one to inject their prior knowledge about the solution space. In the previously discussed example on compressed sensing, the measurement, y, was related to a sparse vector, f, via $\vec{y} = (\mathbf{\Phi\Psi})\vec{f}$. As this is an underdetermined equation, one could use the L1

norm to enforce sparsity on the prediction for $\hat{f}$,

$$L(\theta) = \text{MSE}(y, (\mathbf{\Phi\Psi})\hat{f}) + \beta|\hat{f}|_1, \tag{2.99}$$

where β is a constant. There exist many other regularisation techniques. One that is used in this thesis is total variation (TV), which assesses the changes of a function across a dimension. Considering a parameter $C_{i,j}$, the total variation along the first index is expressed as,

$$\text{TV}(C) = \sum_j \sum_i \|C_{i,j} - C_{i-1,j}\|. \tag{2.100}$$

Training

In supervised learning, the fundamental objective is to make accurate predictions on unseen data; a property referred to as generalisation. When training a parameterised model in a supervised way, one typically separates their available data into training, validation and test sets. The training data is what is used to optimise the networks weights. Then the validation data is used during training, to assess how the network is performing on unseen data (how well it is generalising), and one may adjust hyperparameters (for example, the learning rate or the network architecture) in order to achieve a lower validation loss. Finally, it is the error on the test data that gives the assessment of the model's performance.

Two undesirable results are that of underfitting and overfitting. The former describes the case where the model is not capable of learning the patterns in training data, perhaps instead predicting the mean value for each example. This is typically a sign that the model's capacity must be increased. Overfitting describes the problem where the model achieves a low error on the training data, whilst performing much worse on the validation data. This is a sign that the model has begun to learn the specific noise in the training set rather dataset-wide patterns, and can be mitigated by increasing the training data size, or limiting the capacity of the model.

2.6.1 Neural Networks

A specific case of supervised learning deserves its own section; the neural network is the most famous technique of artificial intelligence. Inspired by the structure of the brain [139], they are composed of connected layers of neurons. The input is passed to the first layer, where at each neuron, a weighted sum of the inputs is calculated, a bias added, before being passed through an activation function, σ. This activation function is typically nonlinear, and gives the network the capability to represent many complex, non-convex functions. The output of the neuron is then passed to the next layer and this sequential process is repeated until the final layer, where the prediction is made. The optimisable parameters of the network are the weights and biases of each layer. The connectivity of each layer with the previous is partly what determines the architecture of the network. Traditionally used activation functions included the sigmoid and hyperbolic tan (tanh) functions. However, these compress the input values into small ranges which can cause the gradients to become small in backpropagation, leading to slow learning. For this reason, most modern methods use an activation function based on the rectified linear unit (ReLU) [140], which is as follows,

$$\mathrm{ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0. \end{cases} \tag{2.101}$$

The derivative of this function has a simple, easy-to-calculate form, and ReLU does not restrict the activations to a small range.

Returning to the network architecture, in the most elementary case, every neuron from layer $m-1$ is connected to every neuron in layer m. This is referred to as a fully connected layer (FCL), and is shown in Fig. 2.12(a). Denoting the input to layer m as $z^{(m-1)}$, the value at neuron j in the layer is calculated as,

$$z_j^{(m)} = \sigma\left(\sum_i z_i^{(m-1)} W_{i,j}^{(m)} + b_j^{(m)}\right) \tag{2.102}$$

where $W_{i,j}^{(m)}$ describes weights connecting neuron i in layer $m-1$ to neuron j in layer m, and $b_j^{(m)}$ are the biases. This method is extremely expressive, which in some situations is beneficial, but can also cause the layer to overfit to noise or to the specific training set, losing generalisation. Also, due to the large number of parameters, they are often prohibitively computationally expensive. This is apparent in the case of image analysis, where a flattened megapixel image has 1000×1000 elements. Using a fully connected layer with the same number of outputs (in case one wished to use a network to edit the image) creates 10^{12} trainable weights, which if stored in 8 bit format, equates to 1 terabyte of storage.

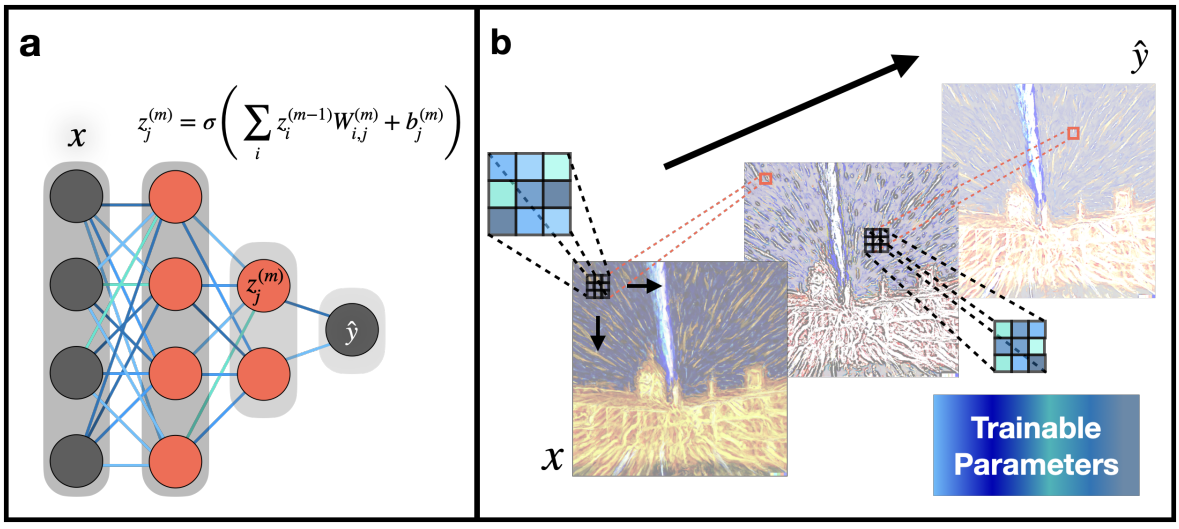


Fig. 2.12 Showing two of the most typical types of neural network, where blue colours represent trainable parameters (weights). **a**, the fully connected network, where every element of the input is connected to every element of the output by a weight. A bias is also added at every neuron (in orange). **b**, the convolutional network, which is used on image data. Here, trainable kernels are passed over the input, yielding the output of that layer. In some cases this continues until the final prediction which can also be an image, or the output of a layer can be flattened and fed into a fully connected network.

To reduce this expressiveness, and to improve efficiency, one can take advantage of the fact that often it is the local region of the input that is important. By allowing the neurons to only be connected to a local region of the previous layer, the number of parameters is reduced, which is the principle of the locally connected layer (LCL). Despite the fact that this module bares the most similarity to the structure of the brain [141], it has found

very limited applications [142], due to the fact it is nearly always outperformed by a simpler variant, the convolutional layer. Here, the locality is maintained while all the weights are shared across the layer, which is equivalent to passing a trainable convolutional kernel across the input. By sequentially stacking convolutional layers, convolutional neural networks (CNNs) extract highly abstract features, whilst also widening their receptive field to detect features that are broader than the individual convolution kernels; an example is seen in Fig. 2.12(b). Each convolutional layer can be defined by a few hyperparameters; the kernel size, the stride (how many pixels the kernel moves at a time), and the number of channels, which describes how many kernels are passed over the input and determines the dimensionality of the output of the layer. By also incorporating 'max pooling' layers – splitting the input into 2×2 cells and taking the maximum value of each, the spatial extent of the feature map becomes smaller, allowing non-local features to be detected. In computer vision, these two layers are traditionally the building block of nearly all models.

Many of the influential architectures of the last decade involve slight adjustments to these simple ideas. The first 'very deep' CNN was developed in Oxford in 2014, consisting of $\sim$19 layers and achieved state-of-the-art results [143]. Proceeding models, the Inception [144] and Xception [145] networks, revolved around the idea of using multiple filter sizes simultaneously in a single layer, allowing the model to capture different spatial features effectively. In the early days it seemed that deeper architectures equated to superior, due to increased abstraction capabilities. However, the community found that deeper models became increasingly difficult to train and gave diminishing performance improvements, due to the phenomena of 'vanishing gradient'. The successive applications of the chain rule in backpropagation means that the gradients become smaller and smaller, and the updates to the earliest layers become so small that these layers will not train in a practical amount of time. A revolutionary technique designed to alleviate this problem was the Residual Network (ResNet), in another paper with more than 200,000 citations [146]. The deceptively simple idea of residual (skip) connections is to allow the network to bypass certain layers, so that during backpropagation gradients propagate to the earlier

layers much more easily. This is seen in Fig. 2.13. To see the computational benefit of ResNet, one returns to the equations of backpropagation, and considers that with the skip connection, the result at a neuron is,

$$z^m = \sigma\left(z^{(m-1)}W^{(m)} + b^{(m)}\right) + z^{(m-1)}. \tag{2.103}$$

When one considers the application of the chain rule, one sees that this skip connection allows the gradient to flow to the previous layer $(m-1)$,

$$\frac{\partial L}{\partial z^{m-1}} = \frac{\partial L}{\partial z^m}\frac{\partial z^m}{\partial z^{m-1}} = \frac{\partial L}{\partial z^m}\left(\frac{\partial f}{\partial z^{(m-1)}} + 1\right). \tag{2.104}$$

This unlocked the potential of deeper networks (> 50 layers), achieving significant performance gains. While there has been a lot of research in creating new architectures since then, seemingly achieving further increases in performance, it was shown that the standard ResNet obtains the same accuracies on benchmarks when trained with more modern techniques [147].

A final convolutional architecture to mention is one used throughout this thesis – the U-Net [148], which is shown in Fig. 2.13. This network is most commonly used in image segmentation, and has an encoder-decoder architecture, where the input is spatially downsampled (with successive convolutions and max pool layers) before being upsampled again, creating the 'U' shape. Furthermore, residual connections are added between each stage of the up and down sampling. The U-Net can extract non-local patterns (due to the downsampling), and can synthesis large effective kernels.

Away from convolutions, another technique that should be mentioned is *attention*, as it bares similarly with the method developed in Section 4.1. This is the fundamental building block of the transformer architecture, which has revolutionised the field of natural language processing, and given rise to the large language models such as GPT4. The attention mechanism allows the network to focus on different parts of the input, and is particularly

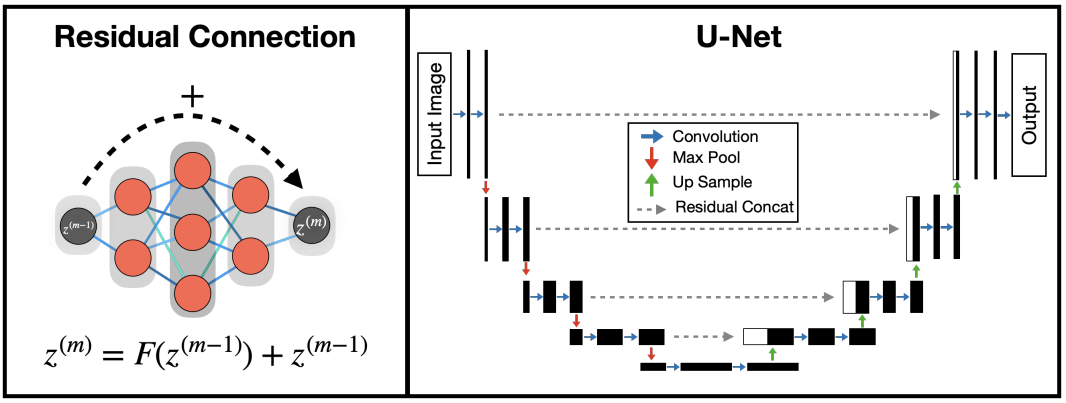


Fig. 2.13 On the left is an example of the residual connection in the ResNet. The input to some block of a neural network is added to the output, resulting in the alleviation of the vanishing gradient problem. On the right is the U-Net. The thickness of each black block represents the number of channels after convolution, and the height represents the spatial extent. Here, the residual connections are concatenated to the part of the upsampling section that has the same spatial extent.

useful in tasks where the input is sequential, while it has also received significant recent interest in image analysis tasks [149–152]. Attention acts on an input vector $\vec{v}$ to give an output vector $\vec{z}$, and its general form is given by,

$$\vec{z} = f(g(\vec{v}), \vec{v}), \tag{2.105}$$

where g is a function to generate the attention, and f applies the attention to $\vec{v}$. Common forms of f include element-wise multiplication, weighted sum, or concatenation, while g can be a simple linear transformation, a neural network, or even a more complex function. A popular evolution of the method, self-attention, refers to the case where the generator function, g, depends on pairwise function of elements of $\vec{v}$. When used as spatial attention, this addresses a limitation of the receptive fields of convolutions, and has thus received use in computer vision in order to capture non-local features [153].

Other types of parameterised operation can be trained using backpropagation and an optimiser such as Adam. A relevant example of this is the Spatial Transformer Network [154]. Here, the incoming signal is resampled onto a new grid. The method of resampling

is parameterised; for example it may be an affine transform with a trainable affine matrix, or a neural network.

Chapter 3

Hyperspectral Compressive Wavefront Sensing

This chapter presents a novel approach to the single-shot characterisation of ultra-short laser pulses, utilising the theory of compressed sensing. The technique is first implemented in a simulated environment, demonstrating the plausibility of the concept. Subsequently, an experimental implementation is discussed, including some unforeseen problems that ultimately led to the abandoning of this approach in favour of another new method that I developed, which is described in Chapter 5.

Gaining an insight into the structure of ultra-intense laser pulses is universally essential for applications. Furthermore, it is not sufficient to separate the measurement problem into the spatial and temporal domains due to the potential presence of spatiotemporal couplings (STCs) [155]. In order to resolve STCs, a three-dimensional measurement is required in space and time, and due to the fact that the electric field is not directly accessible, both the intensity and wavefront must be captured. As the duration of these pulses are on the scale of femtoseconds, a measurement in the time domain is not possible; instead, one measures the Fourier conjugate (wavelength). Thus, the dimensionality of the signal to be captured is, $2 \times n_x \times n_y \times n_\lambda$.

Existing detectors are, at most, two-dimensional; thus, capturing three-dimensional data on such a detector requires compromises. Existing methods usually resort to taking multiple measurements on the detector, scanning through the other dimension(s) [156–158, 119, 40]. While there exist some methods that are single-shot [121], these currently lack resolution, spectral range and are cumbersome to implement. There is, however, another compromise that can be made; the assumption of a prior on the signal to be captured. The theory of Compressed Sensing describes the efficient capturing of a signal, under the prior assumption that it is sparse under some basis transformation. This broad field includes the more specific case of Snapshot Compressive Imaging (SCI), which describes attempts to capture three-dimensional information on a two-dimensional sensor.

It was hypothesised that the combination of SCI and a standard wavefront sensor would enable the hyperspectral wavefront to be captured in a single laser pulse. The optical setup for this technique splits naturally into two parts. Firstly, the use of a conventional wavefront sensor in order to encode the wavefront information into the signal. Here, the quadriwave lateral shearing interferometer (QWLSI) is applied, due to its high spatial resolution. The result of the pulse passing through the QWLSI is a hypercube, containing interferograms for each spectral channel. Secondly, there is the capture of this hypercube via the principles of SCI. The theory of each of these parts is described in the subsequent sections, as well as the details of my simulation-based implementation.

3.1 Wavefront Sensing

A lateral shearing interferometer (LSI) measures the spatially varying phase of a light beam, and is described in detail in Section 2.4.2. Briefly, the LSI works by creating multiple copies of the laser pulse and shearing them laterally relative to each other before their interference pattern is captured on a sensor. Due to the shear, information about the spatial gradient of the wavefront is encoded in the interferogram. The most popular implementation is the aforementioned quadriwave lateral shearing interferometer (QWLSI).

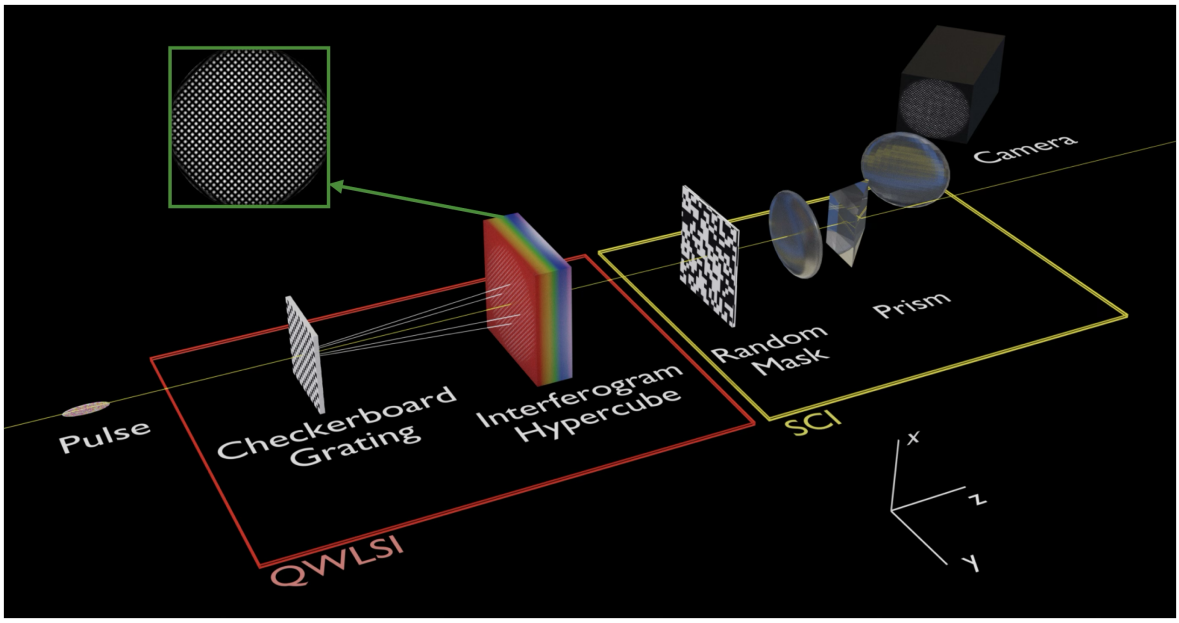


Fig. 3.1 Schematic of the experimental setup that was simulated. From left to right, the pulse first travels through a quadriwave lateral shearing interferometer, yielding a hypercube of interferograms, a slice of which is shown in the green box. The hypercube is then passed through a CASSI setup. This consists of a random mask and a relay system encompassing a prism, before the coded shot is captured on the camera. This diagram is not to scale.

The property of offering two orthogonal sets of derivatives, and therefore two individual estimates for the wavefront, is highly desirable for a sensor based on compressed sensing, because inevitable noise in the measurement corrupts the wavefront with reconstruction artifacts. It was for this reason that methods such as a two-plane Gerchberg-Saxton algorithm were deemed unsuitable. In contrast, redundancy of phase information in a QWLSI provides direct validation and thus makes the wavefront retrieval much more resilient to noise. A sketch illustrating the concept of a QWLSI is shown in the red box of Fig. 3.1. It should be noted that the method presented here is, in principle, applicable to any kind of wavefront retrieval technique, including the Shack Hartmann sensor.

3.1.1 QWLSI Simulation

A physical implementation of QWLSI usually consists of a phase grating with 'pixels' of alternating phase arranged in a chequerboard pattern [98], which leads to dominant diffraction in 2×2 copies of the beam. Instead of simulating this process, considered here

is an idealised setup where the four copies are generated analytically. One begins with creating the pulse of interest by defining a spatial-spectral intensity, I_0 and phase, ϕ_0,

$$E_0(\vec{r}, z = 0, \omega) = \sqrt{I_0(\vec{r}, z = 0, \omega)} e^{i\phi_0(\vec{r}, z=0, \omega)}. \tag{3.1}$$

The pulse is copied 4 times, and each copy's field is propagated to the detector plane according to the equations in Section 2.4.2, restated here for clarity. Note that for brevity, when the z index is not stated, $z = 0$ and the ω index is suppressed from the electric field.

Firstly, the pulse accumulates a spatially varying phase due to the tilt applied from the grating. This phase has the form,

$$\Delta\phi_j(\vec{r}, \omega) = k\vec{r} \cdot \vec{v}_j \sin(\zeta(\omega)), \tag{3.2}$$

where $k = \frac{2\pi}{\lambda}$ is the wave vector of the pulse, $\vec{v}_j$ describes the tilt unit vectors $\in [\hat{x}, \hat{y}, -\hat{x}, -\hat{y}]$, and ζ is the angle of the tilt (in Fig. 3.1, one identifies the azimuthal angle, ζ, as that between the white copy lines, and the central yellow line). This tilt is crucial in reconstruction as it provides a high frequency modulation that separates the gradients in Fourier space.

Furthermore, considering the j^{th} copy has travelled a distance Δz, one finds its planar displacement is,

$$\vec{\Delta r}_j = \Delta z \tan(\zeta(\omega))\vec{v}_j.$$

The resulting electric field of the copy is,

$$E_j(\vec{r}, \Delta z) = \frac{1}{4}\sqrt{I_0(\vec{r} - \vec{\Delta r}_j)} e^{i\left[\phi_0(\vec{r} - \vec{\Delta r}_j) + \Delta\phi_j(\vec{r}, \omega)\right]}. \tag{3.3}$$

Combining these two effects and summing over copies, one obtains the final changes to the field of,

$$E(\vec{r}, \Delta z) = \frac{1}{4} \sum_{j=1}^{4} E_j(\vec{r}, \Delta z). \tag{3.4}$$

After some propagation distance, Δz, one has a hypercube of interferograms. An example of a single frequency channel slice is shown in the green box of Fig. 3.1. In other applications one collapses the cube onto a sensor at this point (integrating spectrally); however this eliminates the chance of retrieving the spectrally-resolved phase. Instead, as discussed in Section 3.2, here snapshot compressive imaging is used to aid in the capturing of the cube.

3.1.2 Wavefront Reconstruction

Once the interferogram is captured, one must extract the wavefront. As mentioned in Section 2.4.2, current reconstruction methods usually involve multiple steps, i.e. extracting the gradients and stitching them together. This is a time-consuming process, especially in a hyperspectral setting where the reconstruction has to be done for every channel. To address this problem, presented here is a deep learning approach to wavefront reconstruction for LSI. While similar work has been done in the context of Shack Hartmann sensors [159, 160], this is the first application of deep learning to LSI reconstruction, to the best of my knowledge. The network that was used will be discussed in Section 3.3.

3.2 Snapshot Compressive Imaging

Snapshot compressive imaging (SCI) describes the application of compressed sensing theory to the task of capturing a three-dimensional datacube on a two-dimensional sensor. To determine if it is applicable to the problem of ultra-intense laser characterisation, one must consider the conditions of compressed sensing, as laid out in Section 2.5.1. Firstly, the underlying desired signal must be sparse in some basis. This is assumed to be true based

on the fact that it is well known that the intensity of a laser pulse is represented by a small number of Gaussians, whilst the Zernike basis provides an efficient way of representing the wavefront. The second condition concerns the incoherence between the measurement basis and the sparse basis. As previously mentioned, in the situation where one does not know the sparse basis, this condition is satisfied via the use of random sampling.

However, the task of randomly sampling a three-dimensional cube onto a two-dimensional sensor is not trivial. To do so, nearly all research in this area uses hardware based on the coded aperture snapshot compressive imaging (CASSI) system [161, 162]; this is a method that has also received recent interest in ultrafast optics [163, 124, 164].

CASSI

The hypercube is first passed through a coded aperture – this is a binary random mask with each pixel transmitting either 100% or 0% of the light. This mask is then imaged onto the sensor by a 4f optical system, which is a common relay imaging configuration consisting of two lenses of focal length, f_l. The object (i.e., the coded aperture) is placed a distance of f_l in front of the first lens, the second lens is placed a distance of $2f_l$ behind the first lens, and the image forms a distance of f_l after the second lens. A benefit of the 4f system is that it gives access to the Fourier plane which is located midway between the two lenses (a distance of f_l from each). By placing a dispersive medium (e.g. a prism or grating) in this plane, different wavelengths are spatially shifted by varying amounts, effectively encoding spectral information as a function of position. The resultant spectrally integrated image on the sensor – known as the *coded shot* – is composed of the contributions from all spectrally dispersed and masked channels. The desired effect of this optical system is that when the hypercube reaches the detector plane, each spectral channel is encoded with a different coded aperture, thereby approximating random sampling across the whole cube. It is then possible for a reconstruction algorithm to retrieve the cube. A diagram of a CASSI system is shown in the yellow box in Fig. 3.1, with an example of the coded shot shown on the camera. The setup is then simulated by multiplying the cube by the mask,

then shifting the channels of the cube according to the amount of (angular) dispersion imparted onto them, and finally summing over the spectral axis.

3.2.1 Reconstruction

Mathematically, the CASSI system discussed above is summarised into a matrix $\mathbf{\Phi}$, which operates on $\vec{m}$, a vectorised representation of the hypercube, to give $\vec{n}$, a vectorised coded shot,

$$\vec{n} = \mathbf{\Phi}\vec{m}. \tag{3.5}$$

Due to the shape of $\mathbf{\Phi}$, the task of recovering $\vec{m}$ is an under-determined problem; there are many possible versions of $\vec{m}$ that can satisfy this equation. In order to reconstruct $\vec{m}$, one can solve,

$$\tilde{\vec{m}} = \text{argmin}_{\vec{m}} \left[\underbrace{||\vec{n} - \mathbf{\Phi}\vec{m}||^2}_{\text{data term}} + \underbrace{\eta\mathcal{R}(\vec{m}, \psi)}_{\text{regulariser}} \right]. \tag{3.6}$$

The first term on the right-hand side is labelled the data term, and enforces that the hypercube must match the coded shot when captured. As mentioned, this alone is an under-determined system, so a regularisation term, parameterised by ψ, is added which restricts the solution space and selects the correct hypercube.

Most methods that have been developed to solve this non-convex equation fall into two classes: iterative algorithms or end-to-end neural networks. The former offers good generalisation but lacks abstraction capability and is slow, whilst deep networks are fast and have been shown to learn almost any function, but are prone to overfitting [165]. A middle ground that offers state-of-the-art performance is deep unrolling.

Deep Unrolling

While one could attempt to train a neural network to solve Eq. (3.6) directly, if it were possible to split the equation, the data term can actually be solved analytically. This is

desirable as it alleviates the abstraction needed to be done by the network resulting in greater generalisation and better parameter efficiency [166]. To perform such a separation, half quadratic splitting is employed. First an auxiliary variable $\vec{p}$ is substituted into the regularisation term, with Eq. (3.7) being equivalent to Eq. (3.6). Then, the constraint is relaxed and replaced by a quadratic loss term,

$$\vec{\hat{m}}, \vec{\hat{p}} = \text{argmin}_{m,p} \left[|\vec{n} - \mathbf{\Phi}\vec{m}|^2 + \eta R(\vec{p}) \right] \text{ s.t. } \vec{m} = \vec{p}, \tag{3.7}$$

$$\approx \text{argmin}_{m,p} \left[|\vec{n} - \mathbf{\Phi}\vec{m}|^2 + \eta R(\vec{p}) + \beta |\vec{m} - \vec{p}|^2 \right]. \tag{3.8}$$

Here, β is a variable that controls the strength of the constraint. High values of β will strongly enforce $\vec{m} = \vec{p}$ and better approximate the subject-to statement.

The benefit of this problem formulation is that it is then possible to split Eq. (3.8) into two minimisation sub-problems in $\vec{m}$ and $\vec{p}$, and effectively separate the data term from the regularisation term. When minimised iteratively, the following sub-problems can approximate Eq. (3.8):

$$\vec{\hat{p}}^{k+1} = \text{argmin}_{p} \left[\beta |\vec{p} - \vec{m}^k|^2 + \eta R(\vec{p}) \right] \sim \mathcal{S}(\vec{m}^k), \tag{3.9}$$

$$\vec{\hat{m}}^{k+1} = \text{argmin}_{m} \left[|\vec{n} - \mathbf{\Phi}\vec{m}|^2 + \beta |\vec{p}^{k+1} - \vec{m}|^2 \right], \tag{3.10}$$

where $\mathcal{S}$ on the right-hand side of Eq. (3.9) represents that a neural network will be used to solve the equation.

Eq. (3.10) is a convex equation with an analytical solution that can be found easily, by differentiating with respect to $\vec{m}$ and setting equal to 0,

$$\begin{aligned} 0 &= \frac{d}{d\vec{m}}\left[||\vec{n} - \mathbf{\Phi}\vec{m}||^2 + \eta|\vec{p^{k+1}} - \vec{m}|^2\right] \\ &= \frac{d}{d\vec{m}}\left[-2\vec{m}^T\mathbf{\Phi}^T\vec{n} + \vec{m}^T\mathbf{\Phi}^T\mathbf{\Phi}\vec{m} - 2\eta\vec{m}^T\vec{p}^{k+1} + \eta\vec{m}^2\right] \\ &= \left[-2\mathbf{\Phi}^T\vec{n} + 2\mathbf{\Phi}^T\mathbf{\Phi}\vec{m} - 2\eta\vec{p}^{k+1} + 2\eta\vec{m}\right] \\ \vec{\hat{m}}^{k+1} &= (\mathbf{\Phi}^T\mathbf{\Phi} + \eta\boldsymbol{I})^{-1}(\mathbf{\Phi}^T\vec{n} + \eta\vec{p}^{k+1}). \end{aligned}$$

This result can be further simplified by the matrix inverse lemma. However, upon implementation, it was found that this approach created numerical instabilities during training. Thus, a simple gradient based update was utilised instead. Defining L as the argument of the argmin in Eq. (3.10), the update was calculated as follows,

$$\vec{\hat{m}}^{k+1} = \vec{\hat{m}}^k - \epsilon\frac{dL}{d\vec{m}} = \vec{\hat{m}}^k - \epsilon\left[\mathbf{\Phi}^T(\mathbf{\Phi}\vec{\hat{m}}^k - \vec{n}) + \eta(\vec{\hat{m}}^k - \vec{p}^{k+1})\right] \tag{3.11}$$

where ϵ is a tunable parameter.

The intuition behind the two steps is as follows; Eq. (3.10) solves the least squares problem ("data problem") of ensuring that, when measured, the candidate solution, $\vec{\hat{m}}$, matches the observed measurement. Eq. (3.9) is the regularisation term tasked with selecting the *correct* solution, among the infinitely many that satisfy the data problem. In order to do this, a prior must be put on the signal to restrict the solution space. Whilst many classical such priors exist, for example the total variation prior which enforces smoothness, here a neural network was used in order to learn the more specific patterns that are known to exist in laser profiles.

The deep unrolling process is shown in Fig. 3.2(bi). Firstly $\vec{m}^{(0)}$ is initialised: $\vec{m}^{(0)} = \mathbf{\Phi}^T\vec{n}$. Then the two equations are solved for a fixed number of iterations, with the same architecture neural net being used to represent Eq. (3.9) in each iteration. Each iteration's network has its own set of weights, hence the unrolling of the algorithm (if all iteration's

networks shared weights, this is known in the literature as the 'plug-and-play' framework). The benefit of this independence of each iteration allows the reconstruction to gradually refine the cube. The architecture of the network will be discussed in Section 3.3.

3.2.2 Details of Implementation

There are a number of important factors in how to implement the simulation, in order to reduce the memory usage and to vastly speed up training.

Firstly, it is important to note that the calculations of $\mathbf{\Phi}\vec{m}$ and $\mathbf{\Phi}^T\vec{n}$ never actually need to be performed in this form. This is important to avoid memory problems; if $\vec{m}$ has $500^2 \times 31$ elements, then $\mathbf{\Phi}$ has $500^4 \times 31$ elements, which is a terabyte of information. Instead of forming the matrices, one creates functions that replicate the result of these matrix multiplications. For example, the multiplication of the cube with $\mathbf{\Phi}$, results in the dispersion, masking and spectral integration of the cube (yielding the coded shot). These operations are implemented via efficient functions from the `pytorch` library (such as `torch.roll` for dispersion), which crucially are differentiable, allowing the gradient to propagate backwards and the network to train.

Multiplication of a coded shot with $\mathbf{\Phi}^T$ is simulated by first tiling the coded shot along the spectral dimension, before multiplying with the mask. In fact, this reframing of the matrix multiplication into functions offers some more intuition about the reconstruction algorithm. In Eq. (3.11), there is a term in the update step of $\mathbf{\Phi}^T(\mathbf{\Phi}\vec{\hat{m}}^k - \vec{n})$. The $(\mathbf{\Phi}\vec{\hat{m}}^k - \vec{n})$ simply calculates loss between the measured predicted cube, and the true coded shot. $\mathbf{\Phi}^T$ then projects this back into the cube form, by tiling this error along the spectral dimension. Thus, the error is used to update all spectral channels, even if in the cube space, some of them already are perfect. This fact leads to the design choice of the final step of the unrolling being a regularisation step, rather than a data term.

One further simplification can be made, by pre-dispersing the mask, and reconstructing the *dispersed* form of the hypercube. This allows for a one-to-one relation between patches of the coded shot and the cube (as the projections are the same).

3.3 Network Architecture

As mentioned, neural networks were used in the reconstruction steps for both the QWLSI and the SCI (in the deep unrolling regulariser). This approach was chosen due to its proven state-of-the-art performance, as well as vastly superior speed to iterative algorithms. This section contains the architectures of the neural networks that were used. They will be discussed in the order they are used in the reconstruction process, which is displayed in the flow chart of Fig. 3.2a. Firstly the deep unrolling algorithm performs reconstruction of the interferogram hypercube from the coded shot, and secondly another network, Xception-LSI, reconstructs the spatial-spectral wavefront from the hypercube.

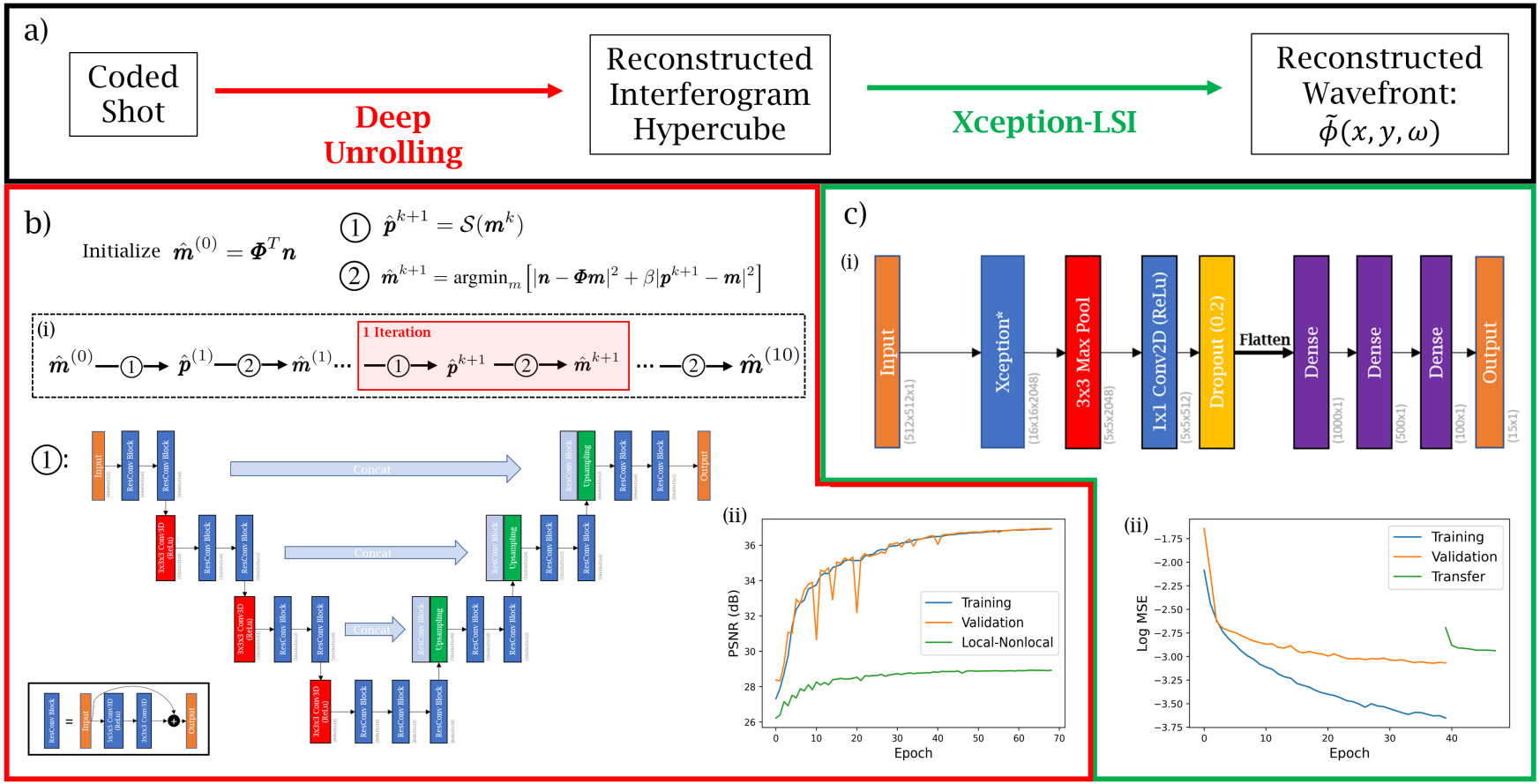


Fig. 3.2 A diagram showing the full reconstruction process of the wavefront from the coded shot. a): A flow chart of the reconstruction process. b): (i) The deep unrolling process, where sub-problems ① and ② are solved recursively for 10 iterations. Also shown is the neural network structure used to represent $\mathcal{S}(\vec{m}^k)$. b): (ii) The training curve for the deep unrolling algorithm. Plotted is the training and validation PSNR for the 3D ResU-Net prior that was used, as well as the validation score for a local-nonlocal prior. Here is demonstrated the superior power of 3D convolutions in this setting. c): (i) The network design for the Xception-LSI network. The Xception* block represents that the last two layers were stripped from the conventional Xception network. c): (ii) The training curve for Xception-LSI for training and validation sets, with the loss shown in log mean squared error. Also plotted is the validation loss when further training the model on the deep unrolling reconstruction of the data (Transfer).

3.3.1 Deep Unrolling Regulariser

As previously discussed, the neural network, $\mathcal{S}$, represents a regularisation term. This means one is able to exploit prior knowledge about the data in order to choose a suitable architecture. As has been described previously, STCs are described by a correlation between Zernike polynomial coefficients and wavelength. Accordingly, there are strong similarities in spot positions for neighbouring spectral channels. Due to this, I developed an architecture with 3D convolutions, which exploits these relations. Inspired by recent work in video snapshot compressive imaging [167], I chose a simplified ResU-Net architecture [168], with the standard 2D convolutions replaced with 3D ones. I used ten iterations for the unrolling process, as it has been found that adding more than this produces negligible performance gains [169]. A diagram of the network is displayed in Fig. 3.2b(i).

3.3.2 Xception-LSI

A wavefront retrieval network was developed that takes a single spectral channel QWLSI interferogram and predicts the spatial wavefront in terms of Zernike coefficients. The network is based on the Xception network [145], but as the original 71-layer network is designed for classification, some changes were made to adapt Xception to the application of wavefront recovery. Firstly the final 2 layers were removed. A max-pool layer and a convolutional layer were added to shrink the output in the spatial and spectral dimensions respectively before 3 fully connected layers with 1000, 500 and 100 nodes using the ReLU activation function [140]. The output layer consists of 15 nodes with linear activation, corresponding to the number of Zernike coefficients to predict. The network is given the name Xception-LSI, and it can be seen in Fig. 3.2c(i). To aid training and to reduce overfitting, dropout was applied to these fully connected layers; this is a technique which randomly sets the outputs of 50% of nodes in a layer to 0 in each training step. Once trained, the dropout effect is removed, and the resulting prediction utilises all the nodes. An intuition of dropout can be thought of as training an ensemble of smaller networks

(depending on the subset of nodes for each training step), and then during the testing phase, predictions are formed as approximately the average of the predictions of the ensemble [170].

3.4 Simulation Results

3.4.1 Training data generation

To represent the initial pulse, a total of 300 cubes were generated with dimensions $(n_x \times n_y \times n_\omega) = (512 \times 512 \times 31)$. The data was randomly split at a ratio of $4:1:1$ into training, validation and test sets, respectively. The wavelength range considered was $750 - 850$ nm, representing a broadband Ti:Sapphire Laser, giving $\Delta\lambda \approx 3.23\,\text{nm}$. For each cube, the wavefront for each channel was first initialised to a randomly weighted sum of 15 Zernike basis functions, where the random numbers were generated from a uniform distribution, $\mathcal{U}[0, 0.05]$. Then, to simulate an STC, one Zernike function was chosen and was made to vary either linearly or quadratically with frequency. Indeed, common STCs such as pulse front tilt and pulse front curvature can be represented in this way [92]. The amplitude of this coefficient was also made to be higher, by multiplying by the maximum noise level (0.05), and a randomly generated signal-to-noise ratio that was between 5 and 20, SNR $= (\mathcal{U}[0, 15] + 5)$. Thus, the generation of the coefficients for the main Zernike function was achieved by the following equation,

$$C_{\text{main}}(\lambda) = 0.05 \cdot (\mathcal{U}[0, 15] + 5) \cdot \left[m \cdot \left(\frac{\lambda - \lambda_{min}}{\lambda_{max} - \lambda_{min}} \right)^n + (1 - m) \cdot \left(\frac{\lambda_{max} - \lambda}{\lambda_{max} - \lambda_{min}} \right)^n \right], \tag{3.12}$$

where $n \in \{1, 2\}$ is a random variable to determine the order of variation (linear or quadratic), and $m \in \{0, 1\}$ randomly selects the direction of variation with frequency. This choice of Zernike coefficients is arbitrary, but allows for a demonstration that the

method identifies all Zernike basis functions. The intensity of slices of the cube were set to an image taken of a real laser intensity.

Each cube was then processed according to Fig. 3.1. Firstly it was passed through the QWLSI simulation (see Section 3.1.1), yielding a hypercube of interferograms – these are the training labels for the deep unrolling algorithm. This hypercube was then passed through the SCI simulation, yielding a coded shot – the training data. The wavefront was reconstructed via the process in Fig. 3.2a. The interferogram hypercube was reconstructed via deep unrolling, before being passed into the Xception-LSI network to predict the spectral Zernike coefficients.

The pitch of the lateral shearing interferometer was set to $\Lambda = 80\,\mu\text{m}$, and the dispersion of the prism, measured at the camera plane, was set to 1 pixel per channel (each channel having a width of 3.23 nm). The propagation distance was chosen to be the Talbot self-imaging plane, $\Delta z = 2\Lambda^2/\lambda$; this is not necessary for a QWLSI, but it gave a suitable amount of shear qualitatively.

Before being passed through the deep unrolling network, the cubes and coded shots were split spatially into 64×64 oblique parallelepiped patches, allowing for a one-to-one reconstruction between the input and output [169]. The initial learning rate was set to 0.01 and decayed by 10% every 5 epochs. The total number of epochs was 70, and the batch size was 8. The network was trained on an Nvidia GeForce 3090 GPU, and took approximately ten hours to train.

The Xception-LSI network was fed individual channels of the ground truth interferogram hypercubes and predicted Zernike coefficient polynomials. Normal random noise ($\mathcal{N}(\mu = 0, \sigma = 0.01)$) was applied to the input, to make the model robust to noise produced by the SCI reconstruction. The initial learning rate was set to 10^{-5} and decayed by 10% every 5 epochs. The total number of epochs was 40, and the batch size was 16. Once trained on the ground truth hypercubes, the model was trained on interferogram hypercubes that had been reconstructed by deep unrolling, for a further 8 epochs. The aim of this

transfer learning was to allow the network to account for any systematic noise in the SCI reconstruction, resulting in a more accurate wavefront reconstruction.

3.4.2 Results and Discussion

The training curves for the SCI reconstruction and QWLSI reconstruction are shown in Fig. 3.2b(ii) and Fig. 3.2c(ii) respectively, and the full reconstruction process is seen in Fig. 3.3. Beginning with the SCI reconstruction, the training curve is shown using the peak signal-to-noise ratio (PSNR) metric, which is typically used in image analysis, and is defined as,

$$\mathrm{PSNR}(\vec{m}, \hat{\vec{m}}) = 10 \cdot \log_{10}\left(\frac{\max(\vec{m})^2}{\mathrm{MSE}(\vec{m}, \hat{\vec{m}})}\right). \tag{3.13}$$

Also plotted here is the validation loss when a "local-nonlocal" prior [171] was used, which was the state-of-the-art for natural images at the time of this work. This demonstrates the benefit of using the three-dimensional convolutional model in this setting; one sees that when both architectures were used with 10 iterations of unrolling, the 3D convolutional model achieved a far superior PSNR, of 36 dB compared to 29 dB. Furthermore, it contains $\sim$45% fewer parameters. The PSNR when performing the reconstruction using the standard U-Net [148] was 27 dB.

In order to reconstruct the hyperspectral wavefront for a full interferogram hypercube, each spectral channel is fed through the Xception-LSI network sequentially. After training, the final mean squared error on the ground truth test set was 6.80×10^{-4}. Fig. 3.2c(ii) displays the training curve with the training, validation and transfer loss curves. This transfer loss describes the effects of first training the network on 'clean' data, before further training the network on interferogram hypercubes that were reconstructed by the deep unrolling method, and proves to be extremely effective in reducing the error of the wavefront predictions when working with reconstructed interferogram hypercubes. The final mean squared error on the reconstructed test set was 9.18×10^{-4}.

An example of the full reconstruction process, from coded shot to spatial-spectral wavefront, is displayed in Fig. 3.3. It is qualitatively apparent that the deep unrolling network was able to accurately reconstruct the interferogram hypercube, and the Xception-LSI network was able to reconstruct the wavefront. Fig. 3.4 shows further results, where it is clear that the technique was able to identify the dominant Zernike polynomial, and its relation with wavelength, in all the examples.

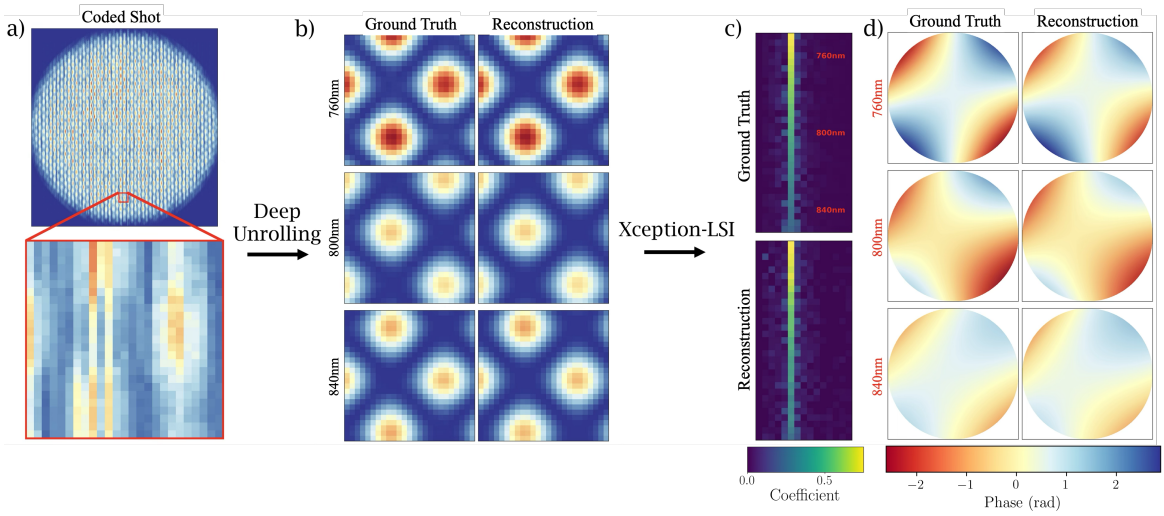


Fig. 3.3 Example results of the reconstruction process. a): An example of the coded shot, along with a zoomed section. b): Deep unrolling reconstruction of the interferogram hypercube in the same zoomed section at different wavelength slices. c): The Xception-LSI reconstruction of the spatio-spectral wavefront displayed in terms of Zernike coefficients, where the x-axis of each plot is the Zernike function, the y-axis is the wavelength and the colour represents the value of the coefficient. d): The spatial wavefront resulting from a Zernike basis expansion of the coefficients in c) at the labelled spectral channels.

3.5 Multi View Deep Unrolling

This section describes further work, where the novel application of Deep Unrolling for systems with multiple views (simultaneous camera measurements) was developed. Naturally, one way to improve the reconstruction accuracy in SCI is to simultaneously capture more measurements, thereby reducing the possible solution space. These additional measurements refer to the capturing of more views, with potentially differing masking and

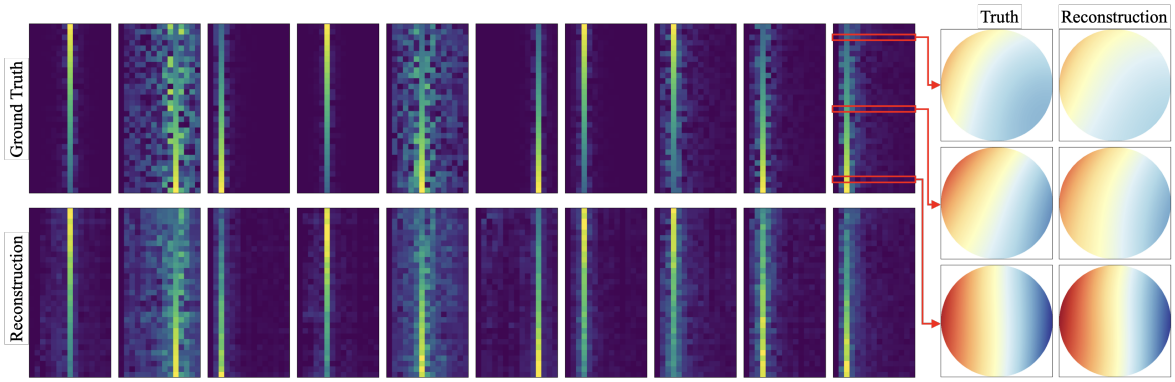


Fig. 3.4 Further results, the x-axis of each plot is the Zernike function, the y-axis is the wavelength and the colour represents the value of the coefficient. It can be seen that the main Zernike polynomial, and its relation with wavelength, is correctly identified in each case.

levels of dispersion. In this setting, the sensing process of A views can be described by the following tensor equation,

$$\begin{bmatrix} \vec{n}_0 \\ \vec{n}_1 \\ \vdots \\ \vec{n}_{A-1} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_0 \\ \mathbf{\Phi}_1 \\ \vdots \\ \mathbf{\Phi}_{A-1} \end{bmatrix} \vec{m} \qquad (3.14)$$

$$\vec{n} \quad = \quad \mathbf{\Phi} \quad \vec{m}.$$

The first case of this was introduced in 2016 [172], involving two cameras. In this design, the first camera captures the masked, dispersed and integrated hypercube, as in the single view case (*coded shot*). The additional camera captures the integrated hypercube without any coding or dispersion, yielding the *preview*. In this case ($A = 2$), and referring to Eq. (3.14), $\mathbf{\Phi}_0$ and $\vec{n}_0$ describe the CASSI process and the coded shot respectively, and $\mathbf{\Phi}_1$ transforms $\vec{m}$ into the vectorised preview, $\vec{n}_1$. This work used the CAVE natural hyperspectral image dataset [173], and solved the inverse problem in Eq. (3.6), using the TwIST algorithm [174] with total variation for the regularisation. The addition of the second view resulted in a significant increase in reconstruction accuracy, going from a PSNR of 24 dB with a single view, to 32 dB with the two views. Other work in the literature demonstrates the further benefit of utilising a RGB camera, as this further

conditions the spectral information [175]. This work led to the further increase of the PSNR, by around 1 dB.

The previously discussed work used relatively simple algorithms, and thus, computational limits were less of an issue. However, whilst the use of large deep neural networks can significantly improve reconstruction accuracy, they are also more memory intensive. For this reason, one often resorts to patching the input data. However, an issue occurs with the use of multiple views with differing levels of dispersion, as there is a mismatch in the projections of the cube.

Patching

Consider patching a hypercube into cube sections. As the hypercube is sensed into the coded shot, the dispersion will cause the patches to mix, meaning that a one-to-one relation between patches of the hypercube and sections of the coded shot is impossible. This problem can be overcome by the idea to reconstruct the *dispersed* form of the hypercube, as described in Section 3.2.2, allowing for a one-to-one projection between pixels of the coded shot and voxels of the hypercube. Now consider that another view is added with a different level of dispersion; e.g. the preview, with no dispersion. A pixel in the preview can be projected onto a voxel of the normal hypercube, but not onto the dispersed hypercube. It is impossible to have a one-to-one relation between a patch of the hypercube and a patch of each of the views, as shown in Fig. 3.5.

A simple way to address this issue, is to only update the cube prediction in the region where the view's projections overlap. This can be implemented practically by creating A compatibility tensors, $\mathbf{\Omega}_a$, which are of the same size as the hypercube patches. The values of these tensors is one in the region of overlap between the view-projected-cube and the reconstruction-cube, and zero elsewhere. For example, if the view, v_a and the reconstructed hypercube possess the same level of dispersion, then $\mathbf{\Omega}_a$ will contain only 1s.

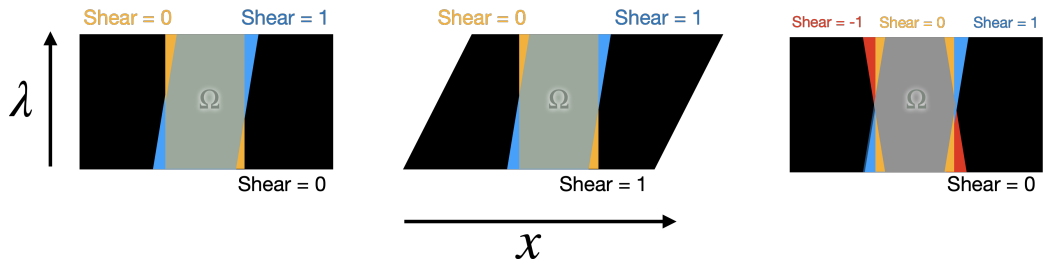


Fig. 3.5 The concept of overlapping projections, for three examples. The black background represents the reconstructed hypercube (the middle example is reconstructing the dispersed form of the hypercube). The other colors represent patches of different views, each with differing levels of dispersion. In the scheme presented here, the only section of the cube that is updated in the 'data term' step, is the overlapping region. The patches are chosen to align at the middle spectral channel, as this allows for maximum overlap.

The compatibility tensors are incorporated into the conjugate gradient update, shown in Eq. (3.15), along with trainable weighting tensors for each view, μ_A,

$$\vec{\hat{m}}^{k+1} = \vec{\hat{m}}^{k} - \epsilon \left[\sum_{a=0}^{A-1} \mathbf{\Omega}_a \mu_a \mathbf{\Phi}_a^T (\mathbf{\Phi}_a \vec{\hat{m}}^{k} - \vec{n}_a) + \eta (\vec{\hat{m}}^{k} - \vec{p}^{k+1}) \right]. \tag{3.15}$$

One problem with this approach is that in order to ensure that every section of the cube has been updated using information from every view, the cube must be patched more times (where some of these patches overlap). This makes the process of reconstructing a cube more time-consuming. However, the cube-reconstruction time of the deep unrolling algorithm introduced in Section 3.2.1 was on the order of 10^{-3} seconds, while the repetition rate of ultra intense lasers are typically on the order of 1 Hz. Thus, the additional time constraints are not prohibitive.

3.5.1 Results

Different combinations of views were tested. In order to be able to use the simple 'rolling' dispersion model in Section 3.2, the amount of dispersion in pixels per wavelength was limited to $[-1, 0, +1]$. The results are displayed in Table 3.1. In all cases, the same network was used for the regulariser as in the single view case and 10 steps of unrolling were used. In agreement with the literature, it was found that the addition of an unmasked

0 dispersion view to the usual +1 dispersion view caused a large gain in reconstruction accuracy, due to the preservation of sharp features [172]. In the presence of low-magnitude noise, the further addition of the −1 dispersion view caused only a small increase in reconstruction accuracy, but this became increasingly beneficial as the relative noise level increased.

Also tested was the case when the 0 dispersion view was masked in the same way as the ±1 dispersed orders, which yielded only a slight drop in performance from the unmasked case – this is significant for the experimental implementation of Section 3.6.

Table 3.1 Multi View Deep Unrolling Results. m and u denote masked and unmasked respectively.

Views	Test PSNR ($\mathcal{N}(\mu = 0, \sigma = 0.01)$	Test PSNR ($\mathcal{N}(\mu = 0, \sigma = 0.05)$
[+1], 1	36.2	30.7
[+1, -1], 0	37.1	33.4
[+1, 0u], 0	42.1	38.7
[+1, 0u, -1], 0	43.2	41.5
[+1, 0m, -1], 0	42.0	40.3

3.6 Experimental Implementation of Diagnostic

An experimental prototype was constructed of the instrument, shown in Fig. 3.6; this corresponds to the SCI section that is illustrated in Fig. 3.1. To maintain the compactness of the CASSI optical setup, while obtaining multiple views, a transmissive diffraction grating (110 lines/mm), was used instead of a prism, and three diffraction orders were captured on the sensor. The fundamental order contained no dispersion, whilst the ±1 orders obtained dispersion in opposing directions, whilst all the views were masked by the same random mask (it was seen that this did not significantly reduce performance as in Table 3.1). This prototype of the dispersing relay-imaging system was set up with the use of f/2.8 MVL50TM23 camera lenses and the Lucid Vision Triton 5 megapixel camera which has a camera pixel size of $x_c = 3.45$µm. The random mask was created by

JD PhotoData, with a pixel size of $x_m = 2x_c$. This setup allowed all three views to be captured on a single camera, which removed the need for aligning multiple cameras, and was a physically compact setup.

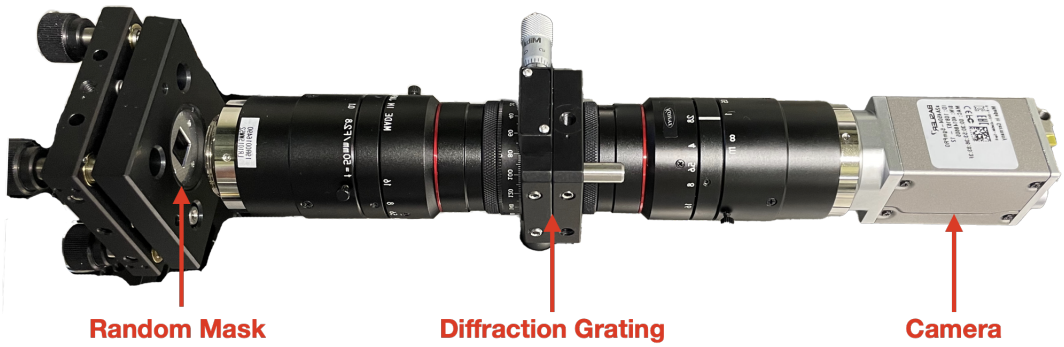


Fig. 3.6 The experimental prototype of the CASSI section of the device. The random mask is imaged onto the camera, with a diffraction grating placed in the Fourier plane, creating the three diffraction orders that are captured on the camera.

Any experimental implementation introduces a number of challenges relative to the simplified world of simulations. One of the key challenges here is the imaging of the mask by the device. In the simulated case described above, perfect imaging was assumed, and the dispersion was treated as a simple 1 pixel-per-nm. Fig. 3.7 shows an example of the signal captured on the prototype, where a random mask was in the object plane and was illuminated by an ultra-short (broadbrand) laser pulse from a Rainbow Femtosecond oscillator. With perfect imaging, the fundamental order should contain the image of the mask with each pixel having sharp defined edges. This is clearly not the case, with the pixels obtaining distortion (this will also be the case for the ± 1 orders). Furthermore, studying the streaks in the the $+1$ order, the dispersion does not appear to be perfectly aligned with the horizontal direction. Therefore, any models that have been trained on the simulated data, and then are naively applied to the experimental data, will not perform reliably. In order to train models on data that is representative of the experimental setup, a more accurate forward model is required [176]. While the forward model was previously assumed to be linear, here it is generalised to a function, Φ, which describes the process of

transforming the known profile of the mask, x, into the image captured on the sensor, y; $\Phi(x) = y$.

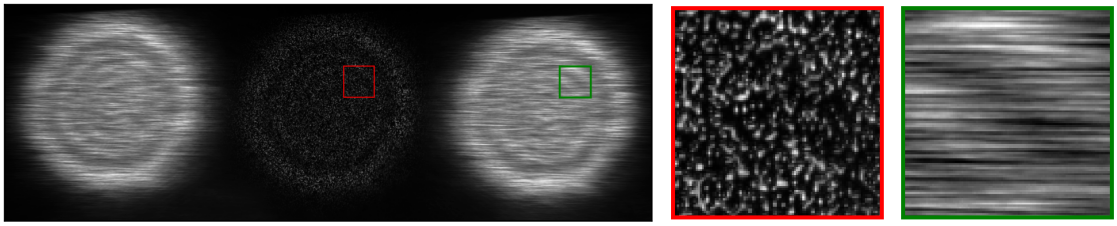


Fig. 3.7 An example of a random mask imaged on the prototype diagonostic. The red and green boxes on the right show zooms of the corresponding square in the measurement on the left. While in the case of perfect imaging one would expect the fundamental to contain a sharp image of the mask, in reality there is clearly distortion due to the imaging system. Furthermore, studying the streaks in the green box (the +1 order), it appears that the dispersion is not perfectly aligned with the horizontal direction.

Initially, efforts were made to create a more accurate forward model, that could be used to generate training data that was closer to the experimental data. Calibration measurements were made for the dispersion curve of the grating, the spatially and spectrally varying point spread function and the spatio-spectral intensity response of the system. In order to do this, a number of masks were created (in addition to the random masks to be used in the CASSI implementation). Firstly, a mask with a single hole was used to calibrate the dispersion curve of the device. This measurement was performed with a monochromator, progressing through $\Delta\lambda = 1$nm steps, where the location of a spot in the dispersed view was recorded. Following this, a mask containing sparse holes, regularly spaced, was used to measure the spatiospectral PSF. Again using the monochromator, the mask was imaged for each wavelength. Then with prior knowledge of the mask's pixel size, the PSF was calculated. All of these experiments were performed on an ultrafast rainbow oscillator.

This forward model could now accurately disperse the spectral components and apply the point spread function via convolution. However, I found that this forward model was not still not sufficient to translate the known mask profile to the measured intensity profile on the camera. To describe the observation that revealed this, consider the fact that the

process of convolution can only spread energy, it cannot focus it. Therefore, if one takes a mask pixel in the camera image, that was uniformly illuminated 2×2 camera pixels originally, no camera pixel should have a value that more than a quarter of the whole mask pixel. However, upon observation, there were certain pixels that contained large fractions of total energy of a whole mask pixel. This demonstrates conclusively that a purely convolutional model does not have the capability to describe the physics of the camera, and a natural candidate for the source of this effect was diffraction.

A study was carried out where a number of different pixel sizes were simulated, and small misalignments from the focal plane were considered. The Rayleigh-Sommerfeld integral was used to simulate the resulting changes in intensity following small displacements of the imaging plane. It should be noted that the paraxial approximation may not have been appropriate here, due to the strong significance of the sharp aperture which diffracts light at sharp angles. Using camera pixels of size, $x_c = 3.45$ um, a number of mask pixels sizes were investigated, $x_m \in [2x_c, 4x_c, 8x_c, 16x_c]$, where $2x_c$ was what was originally planned for the device. Displacements were considered in the range $\Delta z \in [10^{-6}\text{ m}, 10^{-5}\text{ m}, 10^{-4}\text{ m}]$; for context 10^{-4} m is the width of a human hair, and is at the limit of what is detectable to the eye from a typical viewing range of 30cm. The Rayleigh-Sommerfeld integral was simulated with $\lambda = 800$ nm light (corresponding to a typical central frequency of a Ti:Sapphire pulse) and with a spatial sampling size of $\ll \lambda/2$, to ensure the kernel was appropriately sampled. As a sharp-edged aperture was being considered, theoretically the input field required sampling with infinite resolution. The error according to this was mitigated by using a sampling resolution, s, that resulted in a pattern that was qualitativly the same as when a resolution of $s/2$ was used. The results are displayed in Fig. 3.8.

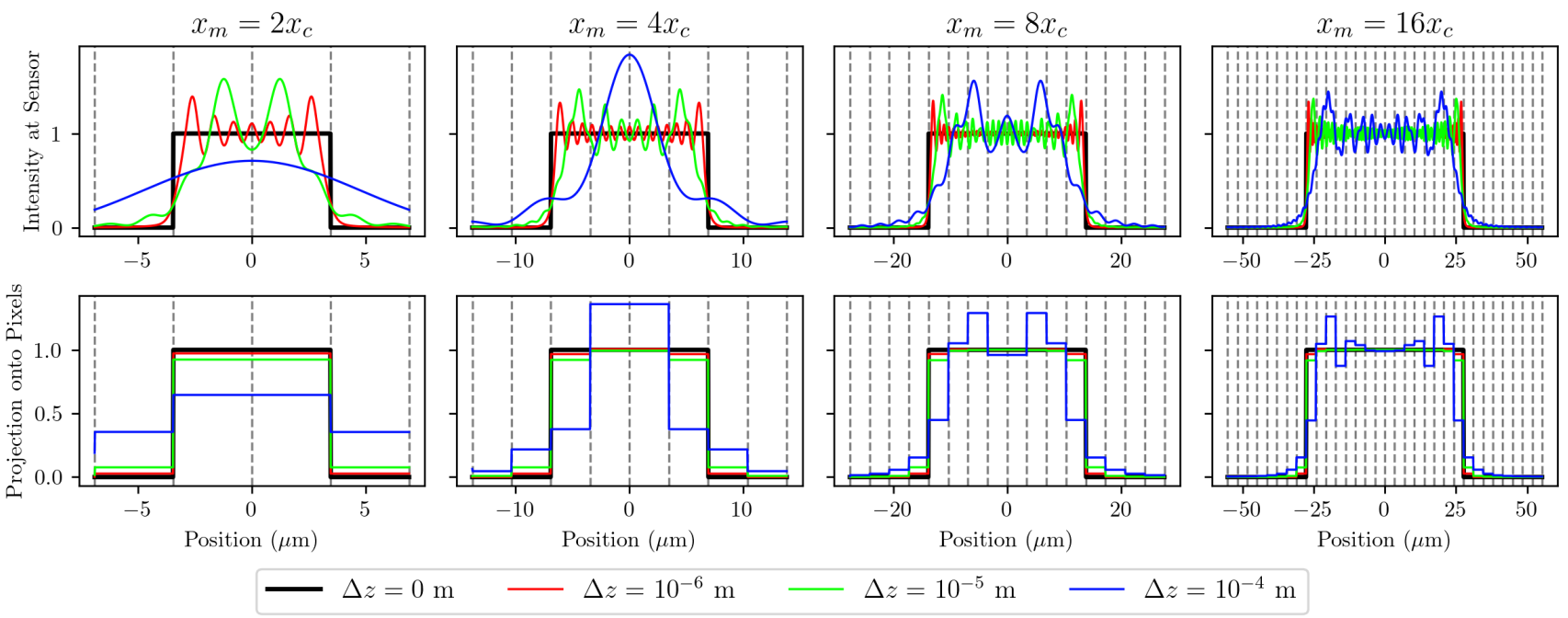


Fig. 3.8 Displaying the resultant intensity distribution at the camera, with small displacements from the image plane, for different mask pixel sizes. With displacements of 10^{-4} m, there is clear distortions in the projections onto the camera pixels. While this problem is mitigated for larger mask sizes, this leads to a diminished spatial resolution of the sensor. The camera pixels are shown with vertical dashed lines, and their size was $x_c = 3.45$ um, while the wavelength of light was set to $\lambda = 800$ nm.

One sees that the intensity pattern is highly distorted for displacements of 100 µm. This problem is compounded heavily when one considers that the mask pixels will not perfectly line up with the camera pixels. Fig. 3.9 demonstrates this effect for the case of $x_m = 2x_c$, showing that the intensity distribution is also heavily distorted for even a shift of 10 µm.

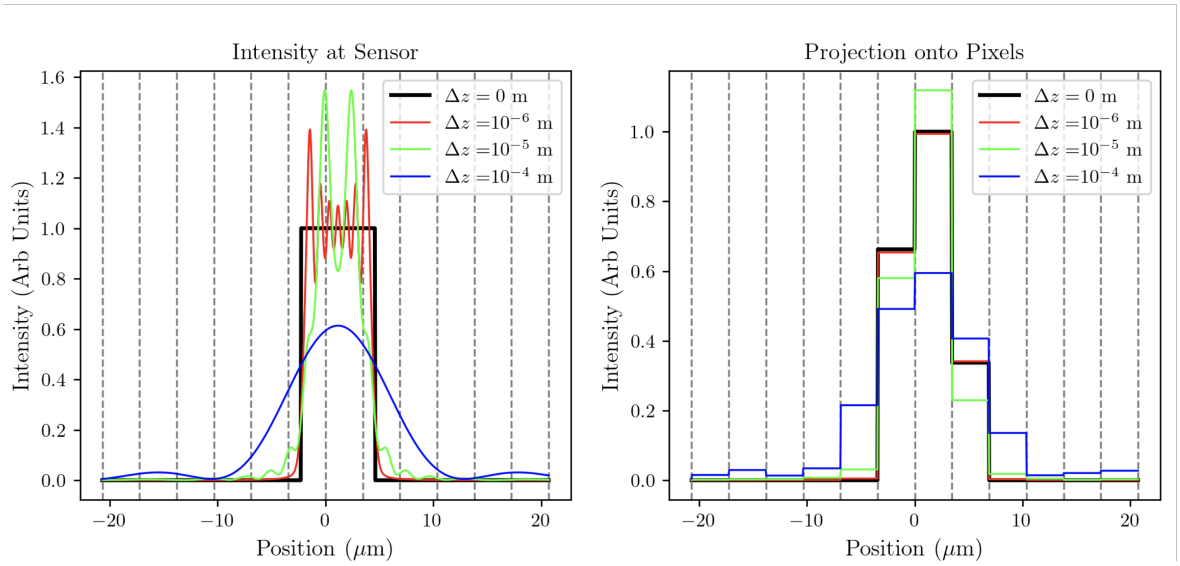


Fig. 3.9 Showing the effects when the mask pixel is not perfectly aligned with the camera pixels, for the case of each mask pixel is twice the size as a camera pixel. The camera pixels are shown with vertical dashed lines, and their size was $x_c = 3.45$ um, while the wavelength of light was set to $\lambda = 800$ nm.

The above figures have displayed the significant effects of diffraction on the intensity pattern recorded by the camera. In particular this problem is larger for smaller mask pixels. While it may be possible to perform a calibration for a completely static system, the extreme sensitivity of < 10µm lead to an impracticality of using the originally planned mask pixels of $x_m = 2x_c$. This problem becomes less pronounced with larger mask pixels. However, as the reconstructed hyperspectral intensity is also defined on the mask pixel grid, one loses spatial resolution. If using $x_m = 16x_c = 55$µm, then this yields a reconstruction with a spatial domain of 125×125 cells. Methods that have implemented CASSI with coherent light since this work was done have used these large pixel sizes, for example 80 µm in [177] and 50 µm in [178] and [124]. Considering that using a wave sensing technique

such as the QWLSI or the Shack Hartmann sensor yields phase information defined on a smaller size than the raw sensor image (eg. the number of microlens in the SH sensor), the spatial resolution of this sensor will quickly become small enough that compressed sensing becomes unnecessary, as one could simply separate spectral components over the sensor such as in STRIPED-FISH. While a GS algorithm can be used to maintain the spatial resolution [124], this algorithm is not robust to noise, which is likely in a SCI reconstruction.

3.7 Summary

This chapter has presented the earliest work of the project; the initial design of a device capable of capturing the hyperspectral wavefront of an ultrashort laser pulse in a single shot, by combining a phase-encoding wavefront sensor with spectrum-encoding Snapshot Compressive Imaging. This involved a simulation-based implementation, where a deep unrolling algorithm was used in the SCI reconstruction, which proved crucial in order to achieve sufficient accuracy.

When attempting to create this device experimentally, I found that there existed large differences from the simulation based environment – in particular, due to diffraction, which caused an extreme sensitivity to small displacements in the optical elements. To mitigate this effect the whole setup must be enlarged, including using larger mask pixels and magnifying the beam. However, at this point the sensor may be big enough to simply separate the spectral components spatially (such as what is done in STRIPED-FISH). Naturally, the significance of diffraction effects depends on the relation between the size of the mask pixels and the wavelength of light, so the work presented in this chapter may still offer a way to achieve higher spatial resolutions when working with light of a shorter wavelength. However, for the purposes of characterising an ultra-intense Ti:Sapphire laser pulse, a different approach is developed in Chapter 5 which begins with the questions of the necessary resolutions needed to capture the physically-significant information.

Chapter 4

Image Deconvolution and Over-complete Dictionaries

As detailed in the previous chapter, while a simulation-based proof of concept may appear successful, extending it to an experimental setting introduces additional challenges that must be addressed. A particularly relevant issue was the blurring of the signal due to the spatially varying point spread function of the imaging optical system, which was not an effect included in the simple simulation-based forward model, and therefore simply applying pre-trained neural networks would not yield accurate results.

When searching the literature for a solution to efficiently model such a point spread function, or to remove its effect in 'image denoising', it was found that the state-of-the-art solutions revolved around convolutional neural networks. However, as shall be discussed in detail, it was recognised that convolution's property of translational equivariance suggests that it is fundamentally inefficient for describing spatially varying point spread functions. This led to the creation of a new method, described in the first half of this chapter.

The second half of this chapter addresses another problem that was encountered, and the developed solution is utilised in Chapter 5. When performing wavefront reconstruction from the measurements of the wavefront gradients (as is provided by the Shack Hartmann sensor or lateral shearing interferometry), one must choose a suitable basis to perform the

reconstruction in, a point elucidated in Section 2.4.2. As the standard Cartesian basis can be inefficient, it is typical to use the Zernike basis, as was done in the previous chapter. However, there are some increasingly popular specialised wavefronts, such as the optical vortex, which cannot be represented by a truncated form of the Zernike basis. Presented here is a more general technique to find efficient representations of wavefronts, and it is used in the measurement of a spectrally varying optical vortex in the following chapter.

While the described methods are fairly distinct, they were both developed to aid the experimental realisation of the laser characterisation diagnostic, and have been combined into this chapter as they are both contributions to the growing interdisciplinary merger between ultrafast laser science and machine learning [179–181, 160, 110]. These are fields that individually have featured in three of the last seven Nobel Prizes in Physics, and in the context of laser-plasma science, this merger is detailed thoroughly in recent reviews [67] and the 2025 LPA Special Workshop on Intelligent Systems, hosted in Oxford's Department of Physics. In order to demonstrate the presented methods' superiority over existing ones (as part of publications), here they are sometimes implemented on datasets that are unrelated directly to ultra-intense laser physics. However, it is clearly described what was the intentioned use.

4.1 Image Denoising

A commonly addressed problem in computer vision is that of image denoising. When an optical system images a plane, it will cause a blur in the image, according to the point spread function (PSF) of the system. This was apparent when building the experimental diagnostic in the last chapter (Fig. 3.7). Before the discovery of the diffraction effects, it was planned to reduce the workload on the snapshot compressive imaging (SCI) reconstruction by using a modular approach: the SCI method would recover the blurred form of the hypercube, before the hypercube was denoised by an additional method. Upon identifying existing methods that are typically used to achieve this, it was noticed that these methods

often revolved around using convolutional neural networks (CNNs). As will be described, this makes them inefficient for this task, and hints at an opportunity to improve the state-of-the-art.

The PSF – a system's response to a point source – is a convolutional kernel, and by sliding it over the input signal, one obtains the convolved signal. While classical convolution requires a PSF that is spatially-invariant (i.e. no change as it is passed over the input), this is a property that is rarely found in reality, due to, for example, optical aberrations. When one considers a convolution with a spatially-variant PSF, the expressibility of the operation becomes even greater. However, with greater functionality, comes greater computational complexity. Discretised spatially-varying convolution is described by the matrix multiplication,

$$\vec{n}(\mathbf{i}) = \sum_{\mathbf{j}\in\Omega(\mathbf{i})} \mathbf{h}(\mathbf{i},\mathbf{j})\vec{m}(\mathbf{j}), \tag{4.1}$$

where $\vec{m}$ and $\vec{n}$ are the input and convolved signals, $\mathbf{h}$ is the point spread function, $\mathbf{i}$ is a position, and $\Omega(\mathbf{i})$ describes a relevant region around $\mathbf{i}$ (which may include the full domain of $\vec{n}$). Classic, spatially-invariant, convolution enforces that $h(\mathbf{i},\mathbf{j}) = h(\mathbf{i}-\mathbf{j})$.

When imaging an object with an optical system, a PSF is known to be dependent on the in-plane spatial coordinates, (x, y), as well as the angles of incidence, θ (or equivalently depth, z). Using a typical single two-dimensional (2D) sensor does not allow access to the latter variable, which leads to deconvolution becoming underdetermined. Here, a *static* spatially-varying convolution is defined as one where the PSF is consistent for every data sample and depends only on the in-plane coordinates, so that the other degree of freedom is removed. Such a condition is satisfied in many scenarios in optics; for example, when imaging at a fixed depth ($z(x, y) = C(x, y)$), such as in microscopy or in relay imaging, or at very far distances ($\theta \to 0$), such as in astronomy. The aforementioned definition also typically removes examples that include motion blur, which are not static between data samples; work into ML approaches for this regime has already been investigated

[182]. Finally, one notes that the (pseudo-)inversion of a static convolution gives rise to a static spatially-varying deconvolution, which in a noise-free case is totally determined. Specifically what is being considered here, is the problem of performing static spatially-varying convolution and deconvolution. This is most relevant in the case that one has a fixed optical system that they are trying to calibrate, which is a crucial part in the development of the laser characterisation diagnostic presented in this thesis.

Deep learning methods are the standard approach to solving these problems, outperforming traditional methods long ago. Due to the problem's convolutional nature, approaches centre around convolutional neural networks, which are composed of convolutional layers, and have the typically desirable characteristic of weight-sharing – significantly reducing the number of trainable parameters in the network and allowing for efficient position-independent extraction of local features from images. By sequentially stacking convolutional layers, CNNs extract highly abstract features, whilst also widening their receptive field. However, the property of weight-sharing also imposes a constraint on the model's ability to learn spatially-aware representations. This quality of 'translational equivariance' intuitively makes it impossible for a CNN to be sensitive to position at all.

However, standard CNNs inadvertently leverage spatial variance due to padding effects [183, 184]. It is common to pad the input to a convolutional layer with zeros before passing the kernel over it (termed *'same'* padding), as this maintains the spatial size of the image. For simplicity, considering that the input (an image) is uniform with pixel values of a, and the $n \times n$ convolutional kernel has elements with values of b, pixels in the centre of the output will have a value of $\sum_n \sum_n b \cdot a = n^2 ba$. However, when one considers the edges of the input, some terms in the sum will be equal to 0 due to the padding, causing the resulting values in the output to be lower. Post-convolution then, a systematic defect appears around the output's boundary, which propagates inwards with each subsequent convolutional layer. This effect is observable in Fig. 4.1a), where a (12×12) uniform input is convolved with a (3×3) uniform kernel using 'same' padding. The resultant feature map (output) displays the position-encoding effect propagating inward. A different

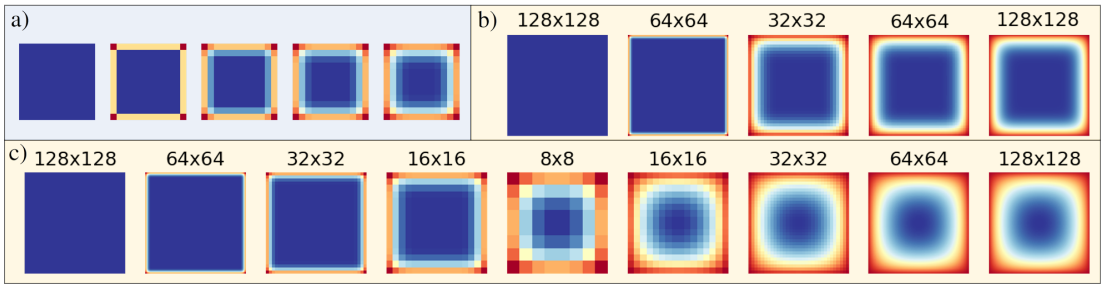


Fig. 4.1 The position encoding effects from *'same'* padding. a): Convolving a uniform input with a 3×3 uniform kernel 5 times. b&c): The same effect for 2 U-Net architectures, containing 2 and 4 steps of down-and-up sampling respectively. Before each dimension-changing operation, three 3×3 convolutions were applied, except for the middle layer in b), where 12 additional convolutions were applied so that each model had the same total number of convolutions.

kernel would disrupt symmetry in these maps. A network cannot detect spatial variance in a region until the defect reaches it; this is visible in the central region of Fig. 4.1a) that retains a uniform value. In a standard computer vision task with an input size of $\sim 500 \times 500$, a significant number of 3×3 convolutions are needed to reach the centre. Using encoder-decoder architectures like the U-Net [148] alleviates this issue. Convoluting at the downsampled layer in U-Net helps encode positional information in fewer steps, but at a downsampled resolution, as shown in Fig. 4.1b-c). Thus, for a U-Net to exploit positional information, it is necessary for the network to be deep and it may require a large number of channels to describe more complicated positional relationships. This unintended positional encoding of padding contributes to deep CNN's performance in spatial tasks, but it encourages exploration of more deliberate encoding methods for potentially higher accuracy and efficiency.

4.1.1 Related Work

Here are detailed the other existing methods to achieve this spatially variant deconvolution.

It is possible to adapt a classic convolution in order to give it some spatial variance. Perhaps one of the most direct methods in this class is the CoordConv layer, which concatenates normalised spatial coordinates to the input features before they are passed through a convolution, aiming to allow the network to learn spatially-aware representations

more explicitly [185], and it has been adopted for various applications [186–191]. In the case of *static* spatially-varying convolution, one desires that the coordinate information affects the feature map in the same way for every data sample. Throughout the training set, the values of data will vary, while the value of the coordinates will be static. CoordConv layers include the coordinate within a weighted sum with the data values which makes it impossible to find weights for a convolutional kernel that will allow the coordinates to effect each sample in the same way. This is a fundamental limitation of CoordConv.

Another method trying to address this issue is the pixel-adaptive convolution (PAC) [192]. In this method, the actual convolution function is multiplied by a pairwise function of pixel features (such as the coordinates). However, in PAC, the pairwise function has a fixed parametric form, such as a Gaussian, which limits this technique's generalisability, as a certain choice of function may not suit all problems. A spatially-varying optical PSF is nearly always smoothly changing, and can be represented as a superposition of a small number of kernels. For this reason, there is no need to introduce an expensive pairwise operation, if a more desirable light-weight method is able to simply interpolate between the kernels. The proposed method is based on this simplification, and enables easier training and faster inference. Furthermore, being a modification of the convolution itself, PAC cannot take advantage of hardware acceleration for standard convolutions.

For the sake of completeness one should also mention the arguably most general case of an adaptive convolution, the locally-connected network (LCN) [193]. Here, each position uses its own kernel to connect to a region in the feature map. Unfortunately, this flexibility comes at a price and LCNs require a much larger number of parameters than CNNs, which requires more memory, makes them prone to overfitting and harder to train. Furthermore, similarly to pixel-adaptive convolution, these networks cannot use the GPU-level optimisation of pure convolutions and thus, are significantly slower to execute than deep CNNs.

As with great proportions of ML, recently the computer vision community has also turned to the attention mechanism. One method utilised self attention for image recognition

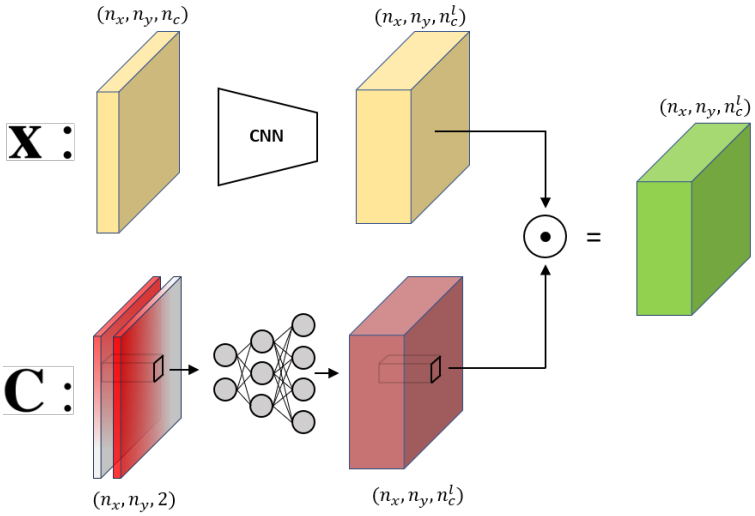


Fig. 4.2 CoordGate. The data, $\mathbf{x}$, and coordinates, $\mathbf{C}$, are fed through a CNN and a fully connected network respectively, before the Hadamard product is used between the resultant tensors.

[187], and also applied a CoordConv inspired technique of concatenating the coordinates to the feature map. Future methods will likely centre around the Vision Transformer (ViT) [152] which has achieved near state-of-the-art performance on classification tasks [194], although they still do not outperform CNNs significantly, and there are issues, such as defects that come from the patching of the image [195]. The ViT uses a method of adding the positional information, which shares the limitation of the CoordConv method in the context of performing a static deconvolution.

4.1.2 Proposed Method

As has been discussed above, current methods either rely on (indirect) modification of the convolution kernel, on concatenating a coordinate map that acts as additional weight, or on the use of an attention mechanism. My proposed method takes inspiration from these approaches, but condenses them into a lightweight module, which is ideally suited for applications such as static spatially-varying convolution in optical imaging. I developed the convolutional **CoordGate** module (see Fig. 4.2), where the input, $\mathbf{x} \in \mathcal{R}^{n_x \times n_y \times n_c}$, is first fed through a standard convolutional block, h, with the final layer containing n_c^l

output channels. As discussed previously, these channels correspond to globally applied convolutions. To synthesise a locally-varying convolution from these, the output channels are then multiplied with a gating mask of the same size, somewhat similar to an attention map. However, in contrast to attention that is based on the input signal, here the gating mask is created by taking a static coordinate map, $\mathbf{C} \in \mathcal{R}^{n_x \times n_y \times 2}$, as in CoordConv, and feeding it through a pixel-wise fully-connected encoding network, g, with the last layer containing n_c^l neurons. If one is using residual learning, a final 1×1 convolutional layer is used to yield an output with the original number of channels. Denoting an index of the two-dimensional arrays with the vector $\mathbf{i}$, and the channel slice of the resultant vector, $\mathbf{y}$, with a, CoordGate is described by,

$$\mathbf{y}_{\mathbf{i},a} = \left(h(\mathbf{x})_{\mathbf{i}} \cdot g\left(\mathbf{C}_{\mathbf{i}}\right)\right)_a . \tag{4.2}$$

The intuition behind the method is as follows. The convolutional network $h(\mathbf{x})$ learns a wide range of resultant kernels, and stores n_c^l different ones in the channels of its feature map. If the network includes downsampling and upsampling, such as the U-Net, then these kernels are very large and incorporate non-local features. In order to selectively attenuate filters, adopting a different resultant convolution for each pixel, the feature map is multiplied by the channel-wise attention generated from the coordinates. In other words, the feature channels form a basis whose amplitudes at each position are encoded in the gating map. Importantly, the encoding network $g(\mathbf{C}_{\mathbf{i}})$ is only dependent on the coordinates and once trained, one directly saves the parameters of the gating map. As a result, during inference the only computational overhead compared to a standard convolution is an element-wise multiplication.

The previously discussed zero-padding boundary effects can also be seen as a position-encoded multiplication operation. A 3×3 kernel being passed over an image will include 3 pixels of 0 when at the boundary (assuming it is not in a corner); accordingly, the corresponding pixel in the feature map will have an average relative magnitude of $\frac{2}{3}$ compared

to a pixel in the centre. This effect propagates inwards with successive convolutions and the proposed method involves a more deliberate and efficient use of this desirable property of multiplication.

Also note that one could, in principle, remove the position decoder network and make the gating map directly trainable. However, this not only results in a much larger number of parameters in the network that will complicate training, but more importantly it ignores the fact that many systems exhibiting a spatially-varying convolution vary smoothly over the input space and hence, can be parameterised.

4.1.3 Experiments

Here the proposed method is implemented to find solutions to a number of problems. Firstly the trivial example of performing 1D spatially-varying convolution is considered, in order to clearly demonstrate the technique's effectiveness compared to common alternatives. Secondly, the method is applied to the practical problem of image deblurring (2D spatially-varying deconvolution).

Learning 1D spatially-varying convolutions

For the 1D case, Eq. (4.1) can be rewritten as the following matrix multiplication,

$$\vec{n} = \mathbf{H}\vec{m}, \tag{4.3}$$

where $\vec{m}$ and $\vec{n}$ are the original and convolved data, and $\mathbf{H}$ is the convolution matrix. Note that in the case of a spatially invariant convolution, $\mathbf{H}$ has the form of a Toeplitz matrix, and can be represented exactly by a single-channel convolutional layer with a suitable kernel size.

Here, 10000 normalised uniform random samples of size (30) are generated for $\vec{m}$, before being multiplied with a custom convolution matrix, $\mathbf{H}$, to give $\vec{n}$. The task of the network is to predict $\vec{n}$ given $\vec{m}$, thus approximating $\mathbf{H}$. The proposed method is benchmarked

against a number of convolutional architectures, demonstrating their limitations. Each network is trained to convergence using the Adam optimiser [137] (with default parameters) to minimise the mean squared error (MSE).

A convolutional neural network is denoted by CNN(L, k, c), containing L convolutional layers with a kernel size of k with c channels, except from the last layer which has 1 channel. A convolutional model utilising CoordConv layers is denoted by cCNN(L, k, c) and the CoordGate architecture is denoted by CG(L, k, c, p), where L, k, c describe the parameters of the convolutional network, and p describes the number of fully connected layers to encode the coordinates, each with c nodes. The custom convolution matrix consists of a Gaussian for each pixel, with varying offset, $\delta(x')$, and standard deviation, $\sigma(x')$, as described by the following equation and shown in Fig. 4.3,

$$\mathbf{H}(x, x') = e^{\frac{-[x-\delta(x')]^2}{2\sigma(x')^2}}, \quad [\delta(x'), \sigma(x')] = \begin{cases} (1 - \frac{x}{15}, 0.5), & \text{if } x < 15 \\ (0, 0.5\left[2 - \frac{x}{15}\right] + 2\left[1 - \frac{x}{15}\right]), & \text{otherwise} \end{cases}. \quad (4.4)$$

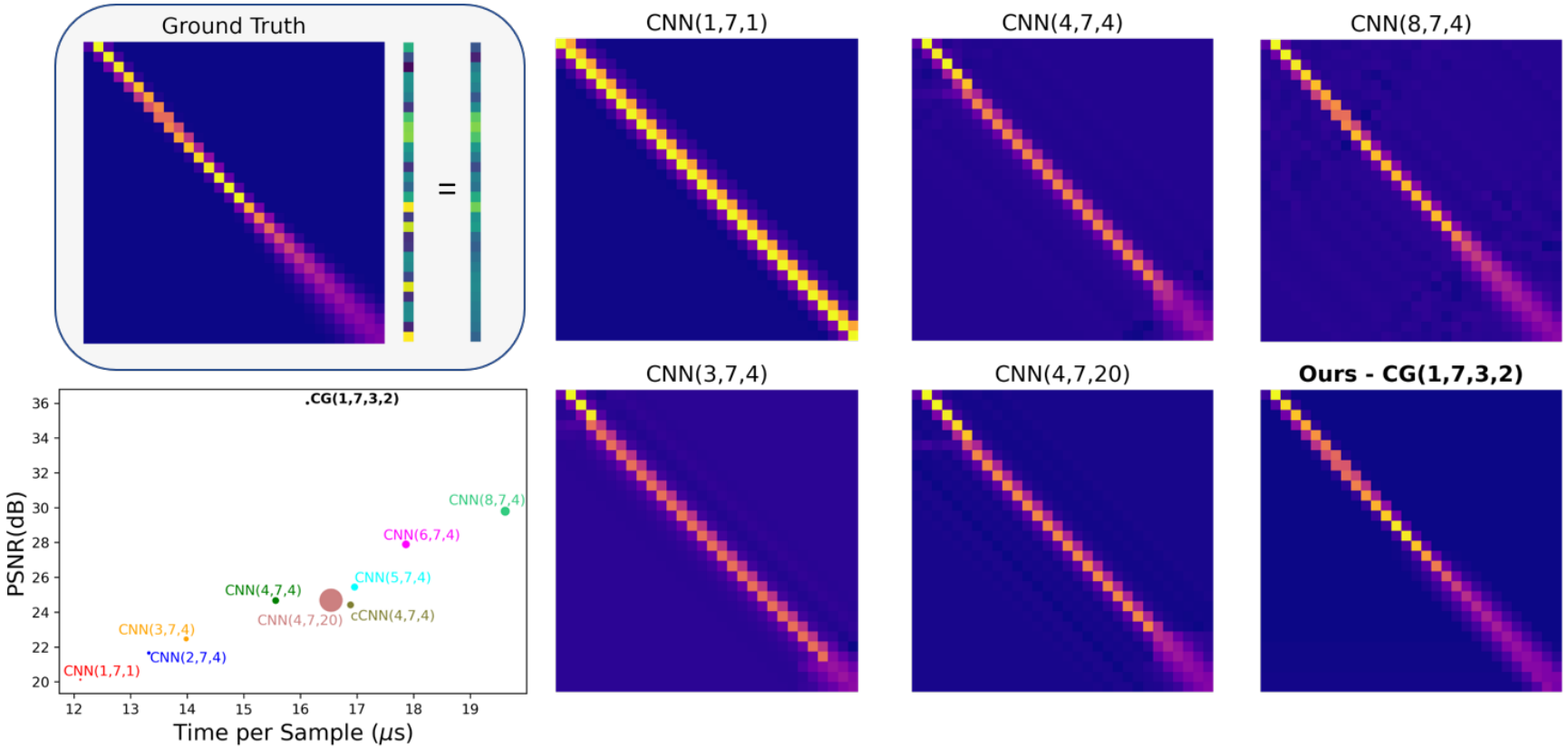


Fig. 4.3 Showing the approximations of the convolution matrix by different models. The CNN with a single layer, CNN(1,7,1), is forced to learn the mean kernel. Then with the addition of each successive layer, it can better approximate the spatially varying kernel near the edges, while learning the mean for the rest. On the other hand, CoordGate can learn the correct kernel with a single convolutional layer. Also shown is a plot of PSNR against inference time for each model, with the spot size being proportional to the number of parameters in the model.

Discussion. The graph in Fig. 4.3 displays the peak signal-to-noise ratio (PSNR), defined as $10 \times \log_{10}(\frac{1}{MSE})$, against the inference time for each model. Beginning with the CNNs, the simple case of a single convolutional layer is intuitive, with the network learning the mean kernel. One then sees an increase in performance with successive layers added. This is explained by the padding effects described earlier, and is seen between the CNN(3,7,4) and CNN(4,7,4) plots, where the boundary is moving down the learned convolution matrix, allowing the network to learn spatially variant information in that region, whilst it learns the mean for the rest. This propagation of spatial information is also seen for CNN(8,7,4). The comparison of CNN(4,7,4) and CNN(4,7,20) demonstrates that adding more channels does not help the model to learn spatially varying features. The CoordConv model possesses no benefit over a standard convolutional model in this task – in fact, studying the convolutional kernel, the model tries to discard the effect of coordinate layer (by learning weights equal to 0). On the contrary, the model utilising CoordGate learns the convolution matrix using one single convolutional layer, with significantly less inference time and a fraction of the number of parameters required by a convolutional model. The convolution matrix can be thought of as a superposition between 3 Gaussian kernels: ($\delta = 1$, $\sigma = 0.5$), ($\delta = 0$, $\sigma = 0.5$) and ($\delta = 0$, $\sigma = 2$), so it is intuitive that the CoordGate model required 3 channels in order to interpolate between them. Also tested was the case when instead of coordinates, the static variable was initialised as random numbers ($\mathcal{U}[-1, 1]$), which was found to not converge, highlighting the use of the coordinates.

Furthermore, the CoordGate method scales much better with increasing sample size, or with reduced kernel size, as it does not require subsequent convolutions to encode the position.

Image deblurring

Image deblurring is an example of a spatially-variant deconvolution. Of the different possible types of blur, many are dynamic – that is, they may vary from image to image – including things like motion blur. Here a static blur is considered, solely originating

from imperfections in the imaging system. In the context of this thesis, an effective image deblurrer could also be used to sharpen images from an imperfect relay imaging system, such as the work in Chapter 3 [110], but this is also a problem that has received interest in fields such as microscopy [196, 197] and astronomy [198, 199].

The U-Net is well suited to spatially-varying deconvolution due to its ability to synthesise resultant kernels with wide receptive fields, which helps with two problems; firstly, deconvolution typically requires a much wider receptive field than convolution, due to a more non-local resultant kernel (which theoretically spans the whole image due to mixing, but can be approximated with a smaller kernel). Secondly, as has already been discussed, in order to capture spatially varying features throughout the image, the zero-padding defect must be propagated to the centre. This second condition requires the U-Net to be very deep, as is seen in Fig. 4.1, which adds a large number of parameters (as typically the number of channels increases with depth). Here, it was hypothesised that adding CoordGate to the model will allow for a more efficient extraction of spatially-varying behaviour, and thus allow a much shallower U-Net to be able to achieve equal performance, with a fraction of the parameters.

Multiple U-Nets are used, and an architecture with depth d is denoted as U-Net(d). The form of these models is seen in Fig. 4.4a). In every step there are two 3×3 convolutions with the rectified linear unit activation function (ReLU) applied [140]. To test CoordGate's effectiveness in this scenario, it is added to both the most shallow and deep U-Nets at each down-or-up sampling point. For comparison with existing methods, CoordConv-UNet [189] is employed, along with a state-of-the-art method in this setting of image deblurring from a static point spread function, named MultiWienerNet [196]. This technique first requires the measurement of the PSF at a number of locations across the sensor (which is an inconvenience not required for CoordGate). Then Wiener deconvolution is performed each of these PSFs, resulting in a number of feature maps that are relatively unblurred in the positions of where each PSF was measured. Finally, a U-Net is used to combine and refine these maps to give the prediction. One notes that the MultiWienerNet was developed

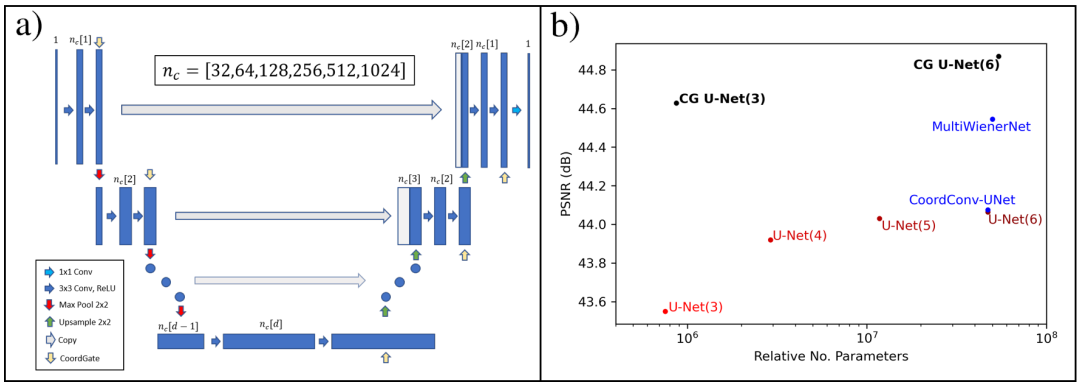


Fig. 4.4 a): The backbone U-Net architecture. A model with depth, d, has $n_c[d]$ channels in its deepest layer. The yellow CoordGate arrows are added for the CG U-Net(d) models. b): A plot of PSNR against the logarithm of the number of parameters, demonstrating the advantage of adding CoordGate to the U-Net architecture. Also included are the CoordConv-UNet and MultiWienerNet models.

in the setting where the PSF varies very quickly (to be used with compressed sensing [200]), whereas here a PSF more typical of a lens is considered. Both CoordConv-UNet and MultiWienerNet were implemented with the deepest U-Net architecture, U-Net(6).

Each model is evaluated using a database of microscopy images of live cells collected from multiple sources [201, 202]. A synthetic PSF was applied to each of the 20000 samples, with defocus increasing towards the edge of the image to try and simulate a realistic lens, which is seen in Fig. 4.6. The job of a model, is to recover the original unblurred image from its blurred counterpart. Each model was trained with the Adam optimiser for 600 epochs, when they showed convergence. The initial learning rate was 0.001, and was halved each time the validation loss didn't decrease for 20 epochs. The training curves are displayed in Fig. 4.5, and an example of a reconstruction is shown in Fig. 4.6.

Discussion. Fig. 4.4b) shows the validation PSNR of each trained model against the number of parameters. Comparing the different U-Net architectures, one sees that the PSNR increases with depth, as expected due to the increased capacity to represent spatially-varying features that has been described previously. The form of the PSF includes a region of focus in the centre, which explains why U-Net(6) performs only marginally better than U-Net(5), as the latter model can still represent this section well with the average kernel.

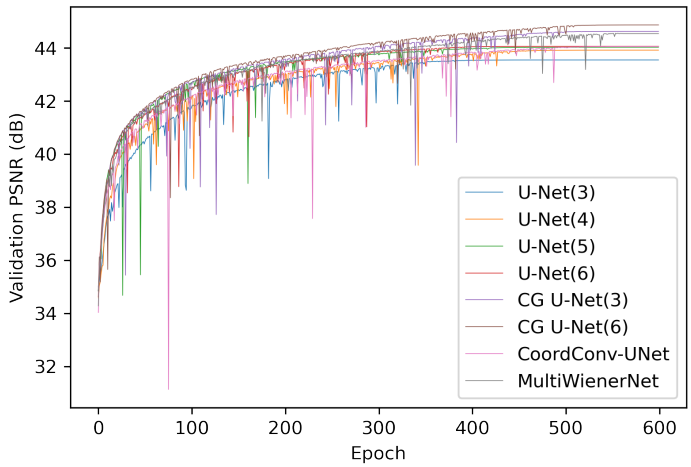


Fig. 4.5 Curves showing the PSNR evaluated on the validation dataset during training, for each model.

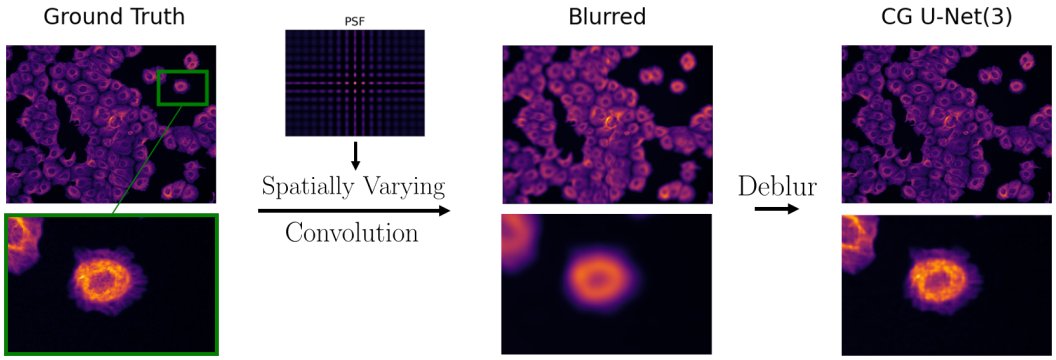


Fig. 4.6 An example showing the blurring and subsequent de-blurring process using the CoordGate technique. This example is generated from data in the test set.

The addition of CoordGate causes a large increase in performance and efficiency. In fact, upon the implementation of CoordGate, the most shallow model (CG U-Net(3)) is able to outperform the deepest normal model (U-Net(6)), despite the fact that it contains 60× fewer parameters. The deeper model has much more capacity to synthesise bigger kernels, which gives it a natural advantage, and thus the fact that it is outperformed shows that CoordGate is superior in learning the spatially-varying features. This claim is supported by studying the spatial loss averaged over all examples, which shows a significantly lower loss close to the centre of the image for CoordGate, as seen in Fig. 4.7. Finally, one notes that CG U-Net(6) performs better than CG U-Net(3), which is explainable by the fact that it

synthesises wider kernels, therefore better approximating the true global deconvolution kernels.

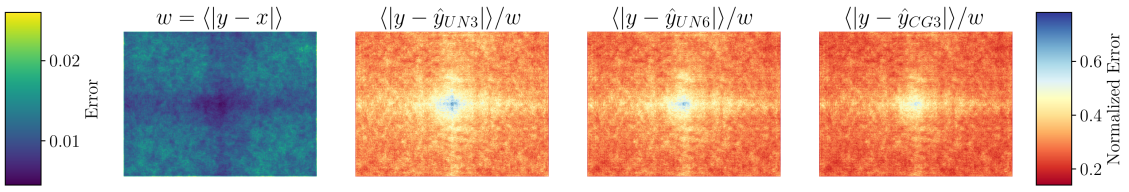


Fig. 4.7 Showing the absolute error in predictions, averaged across all samples in the test set. The plot on the left shows the absolute error between the blurred inputs, x, and the clean target images, y. The error predictably shows the pattern of the spatially varying PSF, and the magnitude is relatively small as many of the examples contain a lot of blank space (see Fig. 4.6.) This quantity is then used to normalise the average absolute error for the predictions from the various models, seen in the right three plots (in order to make the prediction error in the centre of the image more clear). One sees that even the shallow CG U-Net(3) model (CG3) performed better than the U-Net(3) and U-Net(6) models (UN3 and UN6 respectively), especially in the centre of the image.

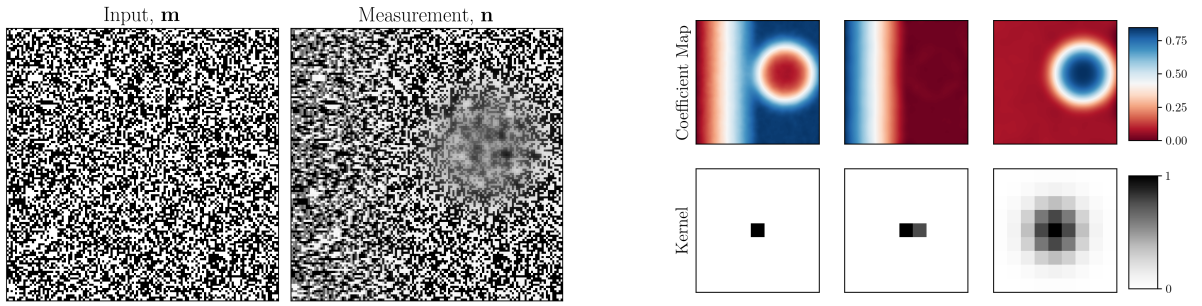
It was found that the CoordConv-UNet model performed nearly identically to the base U-Net(6) model, suggesting that the concatenation operation doesn't provide the coordinate information to the network in an optimum way. The incorporation of Wiener deconvolution in MultiWienerNet resulted in a significant performance increase over the base U-Net. However, it must be remembered that this model is given more information, in the form of some prior measurements of the PSF. Even so, it is outperformed by the shallower CoordGate model, despite containing significantly more parameters.

4.1.4 Extensions

While the objective of designing a novel efficient denoising technique for static spatially-varying convolution had been achieved, and was ready to be used in the processing of the images from the device, there were other points of interest that were investigated during this project.

Interpretable Forward Models

Firstly, as was hinted in Section 4.1.3, this method can be used to find physically-interpretable forward models from calibration data. This forward model, described by a spatially varying PSF, can then be used as part of the input for other image deblurring algorithms (such as MultiWienerNet described above), and can also potentially provide information on the causes of the blur. One way to measure this would be to image a sparse array of pinholes, which would give the PSF for each spot; one could then fit some spatially-parameterised function to make the changing blur interpretable (for example the change in Gaussian width with coordinate). CoordGate offers a potentially quicker method for this task, which does not require the sparse pinhole mask (instead one may use any pair of known object and measured image). For example, consider the simulated calibration measurement for a random mask in Fig. 4.8a, with a pair of a known input $\vec{m}$ and a measured output, $\vec{n}$, which has been blurred with multiple different kernels.



(a) Calibration images

(b) Learnt kernels and coefficient maps

Fig. 4.8 a) Calibration data, a pair of a known input and a measurement. b) The kernels and coefficient maps learnt by the CoordGate method.

By implementing CoordGate with a single convolutional layer (with multiple kernels), one extracts the effective point spread function at different positions, which remains interpretable, potentially unveiling the physical reasons for the blurring. Each convolutional kernel is enforced to sum to 1 and no activation function is used. The learned kernels along with their respective coefficient maps are shown in Fig. 4.8b. This information may then be used in a subsequent deconvolution process (such as in the MultiWienerNet).

Regularised Locally Connected Layers

If one were to treat the map as a trainable parameter directly (rather than generating it from coordinates), and the kernels within the convolutional layer form a complete base, this method becomes equally expressive as a locally-connected layer with shared bias term, as seen in Fig. 4.9.

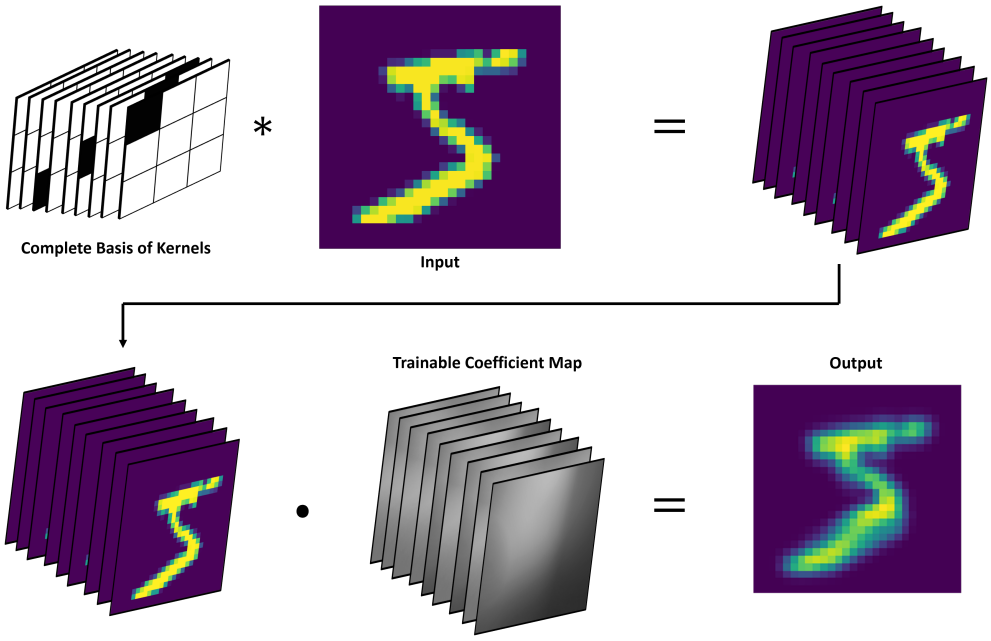


Fig. 4.9 By making the kernels within the convolutional layer of CoordGate a complete basis and the coefficient map fully trainable (rather than being generated by a neural network), the method becomes equally expressive as a locally-connected layer.

Even more interesting, is when one considers applying some form of spatial regularisation to the map. The regularisation could be, for example, total variation, which is then added to the loss function during training,

$$\mathrm{L}(y, \hat{y}, C) = \mathrm{MSE}(y, \hat{y}) + \beta \mathrm{TV}(C), \tag{4.5}$$

with β being a parameter to control the strength of the regularisation. Naturally, when $\beta = 0$ the layer acts as a locally connected layer. However, in the case that β is very large, it will result in the values in each channel slice of the map being the same, $C_{i,j,k} = c_k \, \forall \, i, j$. This means that in each spatial location, the same effective kernel is synthesised from the

complete basis of kernels; this is now a convolutional layer. Thus, the value of β allows one to continuously change from the convolutional to locally connected layer.

The locally connected layer is objectively more expressive than the convolutional layer (after all it can in-theory learn a standard convolutional kernel). However, in practice, these layers are nearly always outperformed by their simpler variation (convolutional), due to their large capacity leading to the overfitting of the training data. There has been recent work demonstrating that a middle ground between these approaches may lead to improved results [203]. Thus, the described method has the potential to be a technique to realise these performance gains.

4.1.5 Conclusion

The CoordGate module has been demonstrated to be efficient in the tasks of performing static spatially varying convolution and deconvolution. The technique works by multiplying the output of a CNN by a gating map, that is generated by a pixel-wise coordinate-encoding network, thereby selectively attenuating the resultant convolutional kernels for each pixel. CoordGate can be seen as lightweight in two ways. Firstly, the implementation itself adds minimal parameters to the existing model. Secondly, the addition of CoordGate allows a much simpler backbone network to achieve a superior performance than that of a more complex model, as has been proven in the experiments. CoordGate's utility was first verified on the simple case of 1D convolution, before being applied to a more challenging problem of removing a spatially-varying blur caused by a lens. In the latter, adding CoordGate modules to a shallow U-Net architecture enabled it to achieve a higher accuracy than a much deeper U-Net architecture, despite having almost two order of magnitude less parameters. For this problem, CoordGate also outperformed two more sophisticated methods MultiWienerNet and CoordConv-UNet in terms of outright accuracy and efficiency. More recently, similar mechanisms for positional encoding have also been applied to transformer architectures [204].

This was developed in an effort to denoise the measurements from the relay imaging system. However, for reasons detailed in the previous chapter, this approach was largely abandoned in favour of the method presented in the following chapter.

4.2 Wavefront Reconstruction Using an Over-complete Dictionary

The next method that was developed focused on the task of wavefront sensing. This section describes a technique that is to be used in the special circumstances of measuring wavefronts that are not efficiently expressed by commonly uses bases, and its development was largely motivated by a single shot measurement of a spatiotemporal optical vortex, that is presented in the following chapter.

Wavefront sensing serves as a pivotal element in numerous domains ranging from computer vision and phase contrast imaging to optical imaging and astrophysics. The wavefront plays a crucial role in shaping the intensity distribution of light as it propagates. This fundamental relationship between phase and intensity forms the basis for various wavefront measurement techniques. As light travels through space or interacts with optical elements, the phase of the wavefront undergoes changes, which manifest as variations in the observed intensity patterns. By carefully analysing these intensity variations, it becomes possible to infer the underlying phase information. For any wavefront measurement technique, the wavefront is extracted from the raw signal using a reconstruction process. This process varies depending on the method – for example, it might involve propagation calculations in the Gerchberg-Saxton algorithm [93], or the stitching of gradients in other techniques such as interferometry or wavefront slope sensors, which is described in detail in Section 2.4.2.

In this reconstruction step, an important consideration is the basis in which the wavefront is expressed. The pixel basis maintains a direct correspondence between spatial positions in the raw signal and the reconstructed wavefront, making it well-suited for capturing high-frequency features such as edges. This has made it prominent in applications like computer vision and phase contrast imaging. However, the pixel basis has significant limitations: its proneness to overfitting and a lack of efficiency. Given the high degrees

of freedom in the Cartesian basis – one value per pixel – there is an increased risk of modelling noise instead of the actual signal, especially in the presence of limited data or high noise levels. These limitations of the pixel basis motivate the exploration of alternative representations for wavefront reconstruction, that leverage prior knowledge about the optical system to mitigate overfitting and provide more robust reconstructions.

With prior knowledge about the system, one can choose a more suitable basis that avoids the problems of overfitting noise. The Zernike basis can be intuitively understood as a polar coordinate adaptation of polynomial regression, optimised for representing wavefront aberrations in circular optical systems. By design, it is directly related to common optical surface aberrations like coma and astigmatism, making it highly suitable to describing wavefronts that were generated with optical components. Defined on a unit disk, the Zernike polynomials provide an efficient way to handle rotationally symmetric systems, common in optical applications. The Zernike modes are also somewhat hierarchical and by truncating the coefficients one is able to reliably calculate standard aberrations. For example, a laser wavefront is typically described by a number of Zernike modes on the order of 10, which represents many fewer coefficients than required in the pixel basis. However, the efficiency of the Zernike basis is diminished when confronted with more complex or non-standard aberrations, for example those with sharp edges; Fig. 4.10a-b) displays the benefit of employing the Cartesian or Zernike basis in certain scenarios.

There exist certain wavefronts which are not efficiently captured by either the Zernike or Pixel basis, such as optical vortices [205], which possess an azimuthally varying phase that remains radially constant. The phase profile of an optical vortex can be written as $\phi(\theta) = \ell\theta$, where ℓ is the topological charge and θ is the azimuthal angle. Despite its very simple form, this phase profile cannot be efficiently captured by either Cartesian or Zernike polynomials, a fact demonstrated in Fig. 4.10c). In addition to optical vortices, myriad other complex wavefront morphologies, such as Bessel beams [206], Laguerre-Gaussian beams, and Airy beams [207], each carrying unique phase and amplitude profiles, are often encountered in optics. Each of these can be captured via concise mathematical

descriptions, but require complex expressions in other bases. Therefore, it is perhaps apparent that a more versatile and encompassing framework is required for wavefront reconstruction.

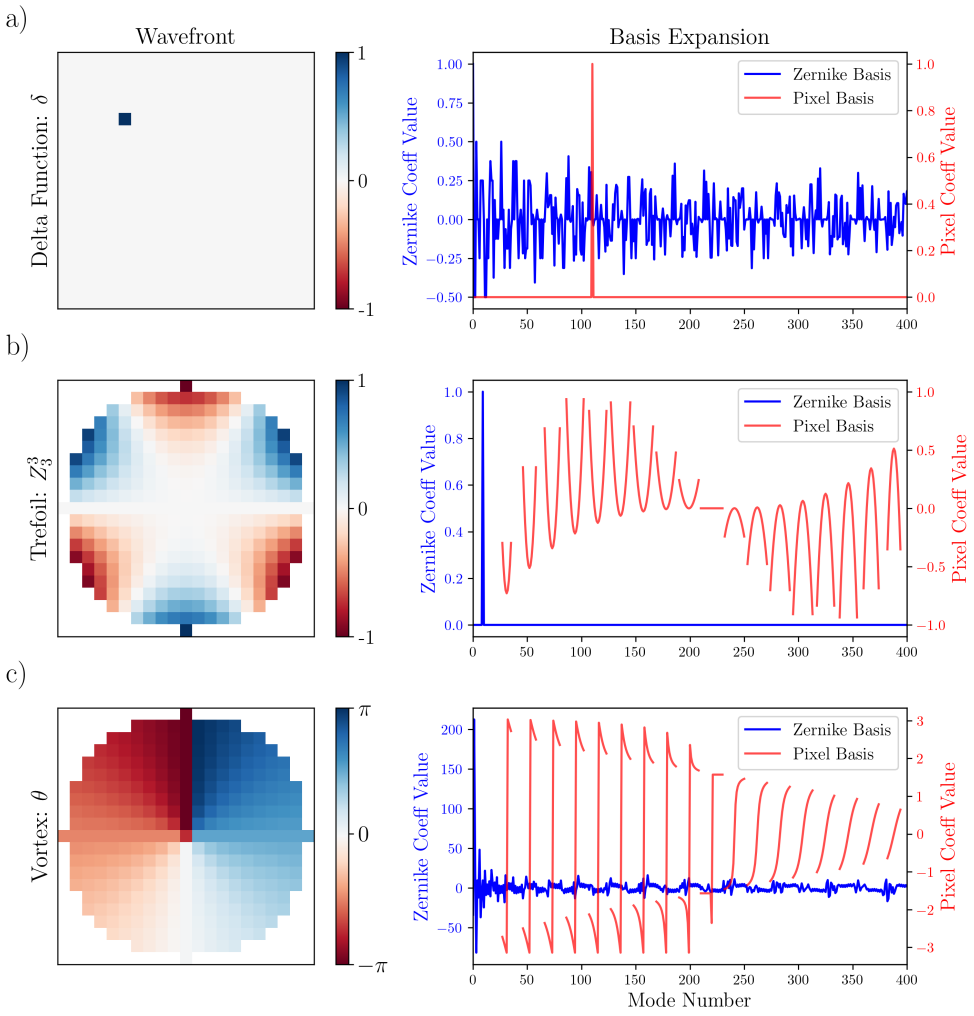


Fig. 4.10 The expansion of a number of wavefronts in terms of the Pixel and Zernike bases. In order to describe wavefront function, Ψ, the coefficient for mode n in basis Φ was calculated by $c_n = \langle \Psi | \Phi_n \rangle$. a) To express a single pixel requires many Zernike modes. b) A single Zernike mode requires many modes in the Pixel basis. c) There exist wavefronts, such as the optical vortex, which are not efficiently represented in either the Zernike or Pixel basis. The Zernike modes are ordered according to their Noll index [208].

Over-complete dictionaries offer a promising solution to this challenge. In contrast to traditional basis sets which are typically complete and orthogonal, over-complete dictionaries contain a surplus of basis functions. While traditional basis sets allow each signal to be represented in one unique way, over-complete dictionaries offer multiple

representations for the same signal, providing additional flexibility. By then using principles from compressed sensing, in particular the use of L1 regularisation [135], one finds the sparsest, most efficient solution, which is also often the most intuitive.

The following section will discuss the problems with existing approaches, before subsequent sections present the over-complete method and demonstrate it on experiments. Chapter 5 then utilises the method in a practical environment.

4.2.1 Limitations of Zonal and Model Bases

The following considers the reconstruction of the wavefront from measurements of its gradient, $\vec{s}$, which is the relevant reconstruction process for techniques such as the Shack Hartmann sensor and lateral shearing interferometry. This is related to the wavefront via the derivative matrix $\mathbf{D}$,

$$\vec{s} = \mathbf{DB}\vec{c} = \mathbf{D_B}\vec{c}, \tag{4.6}$$

where the wavefront, $\vec{W}$, is expressed in the basis $\mathbf{B}$,

$$\vec{W} = \mathbf{B}\vec{c}. \tag{4.7}$$

Discussed here are the limitations of the Pixel and Zernike bases. These are both complete bases, meaning that Eq. (4.7) is capable of expressing any wavefront. While each has its merits, they both possess inherent limitations that render them inadequate for capturing complex wavefronts. The constraints observed in these approaches call for the formulation of a comprehensive, versatile basis set that efficiently handles a broad array of wavefront morphologies.

Each basis set has its limitations. The Cartesian basis is often inefficient for low-frequency components due to its high dimensionality. Conversely, Zernike polynomials struggle to capture high-frequency details or non-standard aberrations. As a result, researchers are often forced to compromise between robustness and representational power.

It has already been seen that there are certain wavefronts which cannot be described efficiently by these bases. However, there are other overlooked issues of existing approaches to wavefront sensing when the Zernike basis is implemented; the following details one which has not been identified before to the best of the author's knowledge.

Misaligned Zernike modes

Consider a wavefront that can be represented by a single Zernike coefficient, e.g. the oblique trefoil mode, $Z_3^3(x, y)$, which is described in polar and Cartesian coordinates as,

$$Z_3^3 = R^3 \cos(3\theta) = x^3 - 3xy^2. \tag{4.8}$$

Now consider a small shift, $x \to x + \Delta$, so the Zernike mode is offset from the centre,

$$\begin{aligned} Z_3^3(x+\Delta, y) &= (x+\Delta)^3 - 3(x+\Delta)y^2 && (4.9)\\ &= x^3 - 3xy^2 + x^2(-3\Delta) + x(3\Delta^2) - \Delta^3 - 3\Delta y^2 && (4.10)\\ &= Z_3^3(x,y) + \underbrace{(x^2-y^2)(-3\Delta)}_{\text{n=2}} + \underbrace{x(3\Delta^2)}_{\text{n=1}} - \underbrace{\Delta^3}_{\text{n=0}}. && (4.11) \end{aligned}$$

This demonstrates that a simple coordinate shift of a Zernike coefficient in row N of the pyramid introduces additional terms from rows $[0, ..., N-1]$; a basis that was sparse will now be very dense. One also notes for a row, n, the shift factor is proportional to Δ^{N-n}. As $\Delta < 1$ (the Zernike basis is on a unit circle), the higher order modes are added with greater weighting. To generalise this observation, one employs a Taylor expansion of the shifted Zernike mode,

$$Z_N^M(x+\Delta, y) = Z_N^M(x,y) + \Delta\frac{\partial Z_N^M}{\partial x} + \frac{\Delta^2}{2!}\frac{\partial^2 Z_N^M}{\partial x^2} + \cdots. \tag{4.12}$$

The derivatives of the Zernike polynomials can be expressed as linear combinations of Zernike polynomials of lower order. Therefore, the first derivative is,

$$\frac{\partial Z_N^M}{\partial x} = \sum_{n=0}^{N-1} \sum_m a_{nm} Z_n^m(x, y), \tag{4.13}$$

and similarly, higher-order derivatives involve even lower-order polynomials. Substituting back into the Taylor expansion, one obtains,

$$\begin{aligned} Z_N^M(x + \Delta, y) = Z_N^M(x, y) + \Delta \sum_{n=0}^{N-1} \sum_m a_{nm} Z_n^m(x, y) \\ + \frac{\Delta^2}{2} \sum_{n=0}^{N-2} \sum_m b_{nm} Z_n^m(x, y) + \cdots . \end{aligned} \tag{4.14}$$

This series demonstrates that a small shift Δ introduces contributions from lower-order Zernike modes down to $n = 0$. The coefficients a_{nm} and b_{nm} depend on the specific mode and its derivatives. Even though Δ is small, the impact on the representation is significant because the higher-order terms decay slowly due to the factorial in the denominator being offset by the increasing number of contributing lower-order modes.

Thus, a wavefront that was sparsely represented by a single Zernike coefficient when centred becomes densely represented upon shifting. This density arises from the need to account for the introduced lower-order modes to accurately describe the shifted wavefront. The practical implication is that even minor misalignments in optical systems, or in the detection process, can lead to significant inefficiencies in wavefront representation and interpretation when using Zernike polynomials.

This effect underscores the limitations of the Zernike basis in situations where the wavefront is not perfectly centred or aligned, which is often possible in the case of an imperfectly-circular beam. It highlights the necessity for a more adaptable representation that maintains sparsity despite shifts or misalignments. It should be noted that this effect

also applies to attempts to fit many other bases. For example, attempting to fit a vortex to a de-centred vortex leads to errors.

4.2.2 Over-Complete Dictionaries

Over-complete dictionaries [209, 210] provide a solution to the problems outlined above. These dictionaries contain more basis functions than the dimensions in the space they represent, offering flexibility to adapt to a variety of signals while maintaining robustness. A wavefront, $W(x, y)$, can be expressed as a basis expansion,

$$W(x, y) = \sum_{i=1}^{N} c_i \phi_i(x, y), \tag{4.15}$$

where ϕ_i are the basis functions from the over-complete dictionary, and c_i are the coefficients. The overcompleteness of the dictionary presents a fundamental challenge: the existence of multiple valid solutions for a given wavefront. This non-uniqueness might seem to undermine the very purpose of this approach. However, it is precisely this redundancy that allows one to overcome the limitations of single-basis representations that have been previously discussed. The pursuit of sparsity in this context is not merely a mathematical convenience, but a reflection of the underlying physics of wavefronts. Throughout this section, it has demonstrated that different physical phenomena in optics – be it Zernike-type aberrations, optical vortices, or misaligned beams – naturally lend themselves to sparse representations in appropriate bases. This sparsity is a consequence of the wave equation and the boundary conditions typical in optical systems. The limitations of individual bases in efficiently representing diverse wavefront phenomena stem from the fundamental nature of optical wavefronts. These are in most cases solutions to the paraxial wave equation:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + 2ik\frac{\partial \psi}{\partial z} = 0 \tag{4.16}$$

where $\psi(x, y, z)$ is the complex amplitude of the field, k is the wave number, and z is the propagation direction. The solutions to this equation encompass a wide range of phenomena, including Gaussian beams, Hermite-Gaussian modes, Laguerre-Gaussian modes (which include optical vortices), and Bessel beams, among others.

The pursuit of sparsity in our over-complete dictionary is fundamentally a search for the most appropriate set of solutions to the paraxial wave equation that describe the observed wavefront. This approach can be formalised mathematically as an optimisation problem:

$$\min_{\vec{c}} |\vec{c}|_0 \quad \text{subject to} \quad \vec{W} = \mathbf{\Phi}\vec{c} \tag{4.17}$$

Here, $|\vec{c}|_0$ denotes the L0 norm, which quantifies the number of non-zero elements in the coefficient vector $\vec{c}$, and $\mathbf{\Phi}$ is the dictionary of modes. This formulation seeks to identify the minimal set of fundamental modes that accurately represent the observed wavefront.

L0 minimisation is non-convex – intuitively it is clear that it has discontinuities, making it hard to optimise. Candès et al. proved that in many situations minimising the more-tractable L1 norm can result in the same solution [211] (see Appendix A). Thus, the equation becomes;

$$\min_{\vec{c}} \|\vec{c}\|_1 \quad \text{subject to} \quad W = \mathbf{\Phi}\vec{c}. \tag{4.18}$$

Crucially, the paraxial wave equation is defined with respect to an optical axis, which in real optical systems may not perfectly align with the measurement apparatus. As discussed in Section 4.2.1, this misalignment can lead to apparent complexity in the wavefront when viewed from the perspective of the measurement system. By including the centre coordinates, (Δ_x, Δ_y), as variables in the optimisation, one may account for this potential misalignment:

$$\min_{\vec{c},\Delta_x,\Delta_y} |\vec{c}|_1 \quad \text{subject to} \quad W = \mathbf{\Phi}(x - \Delta_x, y - \Delta_y)\vec{c}. \tag{4.19}$$

This formulation allows us to simultaneously determine the most appropriate set of paraxial modes and the true optical axis of the system. It recognises that apparent complexity in the wavefront may arise from a simple misalignment rather than intrinsic high-order aberrations or complex phase structures.

By introducing the idea of fitting the centre during reconstruction, one is implicitly expanding the dictionary to include shifted versions of each basis function. Mathematically, one can express this as:

$$\mathbf{\Phi}_{expanded} = \phi_i(x - \Delta_x, y - \Delta_y) : \phi_i \in \mathbf{\Phi},\ (\Delta_x, \Delta_y) \in \mathbb{R}^2 \tag{4.20}$$

This expanded dictionary is infinite-dimensional, as it includes all possible shifts of each basis function. However, it's important to note that this expansion doesn't fundamentally change the nature of the problem – it remains an over-complete dictionary, just with an even higher degree of redundancy. The key insight is that by allowing for these shifts, one enables the reconstruction to find even sparser representations. A misaligned Zernike mode, which might require many coefficients in the original dictionary, can now be represented by a single coefficient in the expanded dictionary.

This adaptive approach not only allows for more accurate reconstruction of misaligned wavefronts but also provides valuable information about the optical system itself. The optimal (Δ_x, Δ_y) values can indicate misalignments in the optical setup or in the detection process, offering diagnostic capabilities beyond mere wavefront reconstruction. For bases that are not rotationally redundant, one can also use a rotation parameter, θ, to find the angle of rotation of the mode.

The technique presented here is simply named by the acronym: PROD (Phase Reconstruction using an Over-complete Dictionary). The following sections detail numerical and physical experiments demonstrating how this expanded dictionary approach, combined with L1-based sparse reconstruction, outperforms static reconstructions across a wide range of wavefront types.

4.2.3 Numerical Simulations

Construction of the Over-Complete Dictionary

Given a set of wavefront derivatives, with spatial dimensions (N_x, N_y), the over-complete dictionary is created by the synthesis of a number of modes from different bases. Added first is the pixel basis, with the introduction of $N_x \times N_y$ modes, allowing the technique to represent any discontinuities. Then, to provide an efficient way to represent the common low frequency features, a set number of Zernike modes are added, where the derivatives are found analytically [212]. Finally, with prior knowledge of the optical system being measured, one can simply add 'special' modes that are expected to be present in the wavefront, such as the optical vortex ($\phi = \theta$). The topological charge of the vortex will then be the modal coefficient that is found by the PROD. As has been described, the pixel basis is already sufficient to describe any wavefront. However, in doing so, it will also capture noise, and it does not provide useful physically interpretable modal coefficients; for example the order of a vortex.

Once the over-complete dictionary has been formed by the concatenation of the modal derivatives, a spatial transformer network is also initialised [154]. This is a module which performs a translation of the Zernike modes and a rigid transform of the special modes, with trainable parameters for the rotation angle, θ, and translations, (Δ_x, Δ_y), enabling the most efficient representation of the wavefront. The Zernike modes and the special modes are each given their own set of rigid transform parameters, to account for the potentially different sources of these distortions. The trainable modal coefficients and rigid transform parameters are then optimised by minimising the error with the measured wavefront derivatives whilst simultaneously minimising the L1 norm of the coefficients, according to Section 4.2.3.

In practice, at initialisation multiple sets of random coefficients and random (or grid) rigid transform parameters are created, and optimised in parallel. By then choosing the one that achieved the minimum cost, the robustness of the technique is increased, without

a dramatic increase in computational time due to the properties of GPUs. Due to the rigid transformation, the spatial domain which the over-complete dictionary is defined on is larger than the original derivatives, and depends on the upper bounds of the translation and rotation parameters; during optimisation, the predicted wavefront derivatives are simply cropped to the same size as the measured ones for the cost calculation.

Optimisation Framework

While L1 minimisation problems are traditionally approached using convex optimisation algorithms like Iterative Shrinkage-Thresholding Algorithms (ISTA)[213], Interior Point Methods (IPM) [214] or the Alternating Direction Method of Multipliers (ADMM) [215], here the gradient based optimiser Adam is used[137]. The complexity of the optimisation problem, particularly with the introduction of centre fitting, makes it potentially non-convex. Adam's ability to handle both convex and non-convex problems effectively makes it a suitable choice. Moreover, Adam's efficient GPU-accelerated implementation in frameworks like PyTorch can in practice outperform traditional convex optimisation algorithms without similar optimisation. The formulation of the sparse optimisation problem is,

$$\min_{\vec{c},\Delta_x,\Delta_y} |\vec{W} - \mathbf{\Phi}(x-\Delta_x, y-\Delta_y)\vec{c}|_2^2 + \lambda|\vec{c}|_1. \tag{4.21}$$

where $\vec{W}$ is the vectorised (flattened) measured wavefront (or more precisely its slopes), $\mathbf{\Phi}$ is the over-complete dictionary, $\vec{c}$ are the coefficients that are being optimised, Δ_x and Δ_y are the centre coordinates, and λ is a regularisation parameter. This approach combines the flexibility to handle potential non-convexity, the speed of GPU acceleration, and the sparsity-inducing properties required for the over-complete wavefront reconstruction problem. The following experiments were performed on a single NVIDIA GeForce RTX 3090 GPU.

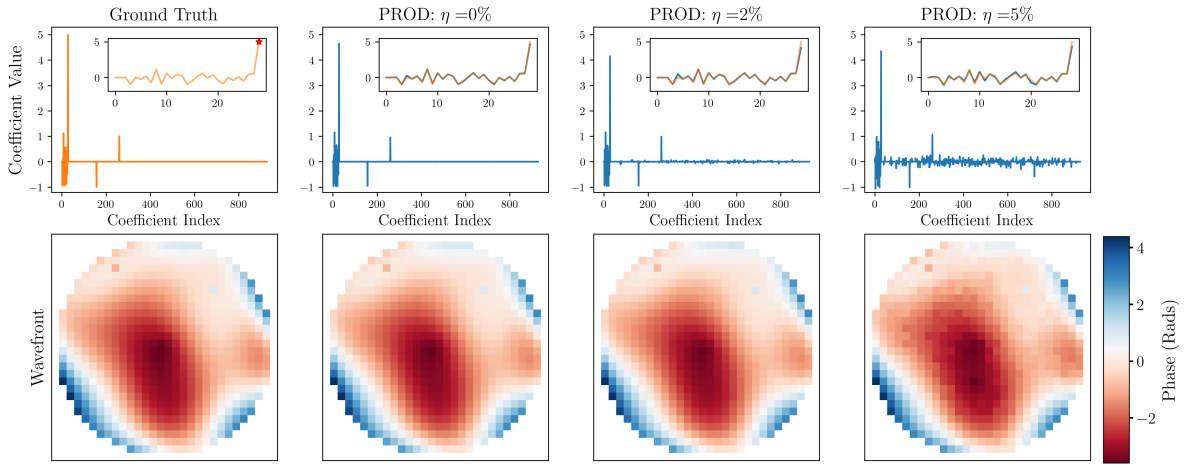


Fig. 4.11 The wavefront was synthesised by the random initialisation of 28 Zernike modes, and the Axicon mode. The top row displays the modal coefficient values, with each subplot showing a zoom of the first 29 coefficients – the Zernikes and Axicon (marked with a red star). The bottom row shows the predicted wavefront, with all plotted on the same color scale. One sees that even in the presence of significant noise (parameterised by a percentage, η, of the maximum value of the derivative), PROD is able to accurately extract the true coefficient values.

Results

Axicon Phase: Firstly, PROD is utilised to perform phase stitching of a wavefront not efficiently expressed in the Zernike basis: a simple linear radial phase, $\phi(r) \propto r$. Such a wavefront can be generated by an Axicon [216], a type of lens formed from a conical piece of glass. Due to its geometry, there is a linear dependence in the radial position and the amount of glass that the pulse has travelled through, resulting in the radial wavefront. When used to focus a Gaussian beam, it creates a Bessel-like beam which does not experience diffraction over a region of interest [217], making it of interest in a wide range of applications.

Here, a wavefront is created by randomly sampling Zernike coefficients from a Normal distribution $\sim \mathcal{N}(0, 0.5)$, up to the 28th mode, before adding the Axicon phase. To test the ability of the PROD technique to handle noise, 3 sampling scenarios are simulated. Normally distributed noise is added to the gradients, with a standard deviation equal to a percentage, $\eta \in [0\%, 2\%, 5\%]$, of the gradients maximum value. The learning rate

of the Adam optimiser was set to 3×10^{-3}, and the regularisation parameter was set to $\lambda = 7 \times 10^{-4}$. The results are shown in Fig. 4.11.

It is evident that as the noise level was increased, PROD was able to use the pixel basis in order to fit the noise, meaning that the predictions of the Zernike and Axicon mode remained fairly constant throughout the noise levels, proving the robustness of the technique.

Shifted Oblique Trefoil: This experiment demonstrates the benefits of the rigid transform aspect of the approach. The phase map is implemented as an off-centered Z_3^3 Zernike mode (oblique trefoil), with 2 pixels of shot noise. Considering the Zernike modes are defined from [-1,1], the new centre position is chosen as [0.2, 0.2]. Normally distributed random noise ($\eta = 2\%$) was added to the sampled derivatives, before the proposed approach was used to recover the wavefront from the derivatives. For comparison, the scikit-learn[218] implementation of Lasso regression [219] is used. This is a commonly implemented technique to solve Eq. (4.18), and is described in Appendix A. The technique is given the same over-complete dictionary, but possesses no ability to perform the transformation of modes. Lasso is parameterised by a L1 penalisation parameter, α, which controls the sparsity of the solution. This was adjusted, and the results are displayed in Fig. 4.12.

As alluded to in Section 4.2.1, a shifted Zernike mode introduces other lower order Zernike terms into the expansion. This is seen for the Lasso technique; as α is decreased, more low order Zernike modes are used to help represent the function. As the proposed approach uses a trainable rigid transform, it finds the most efficient representation using just the single Zernike mode.

4.2.4 Physical Experiments

An experiment was performed to measure an optical vortex. The raw data was captured at the University of Salamanca by Mr. Nils Weiße and Mr. Jakob Schröder (PhD students

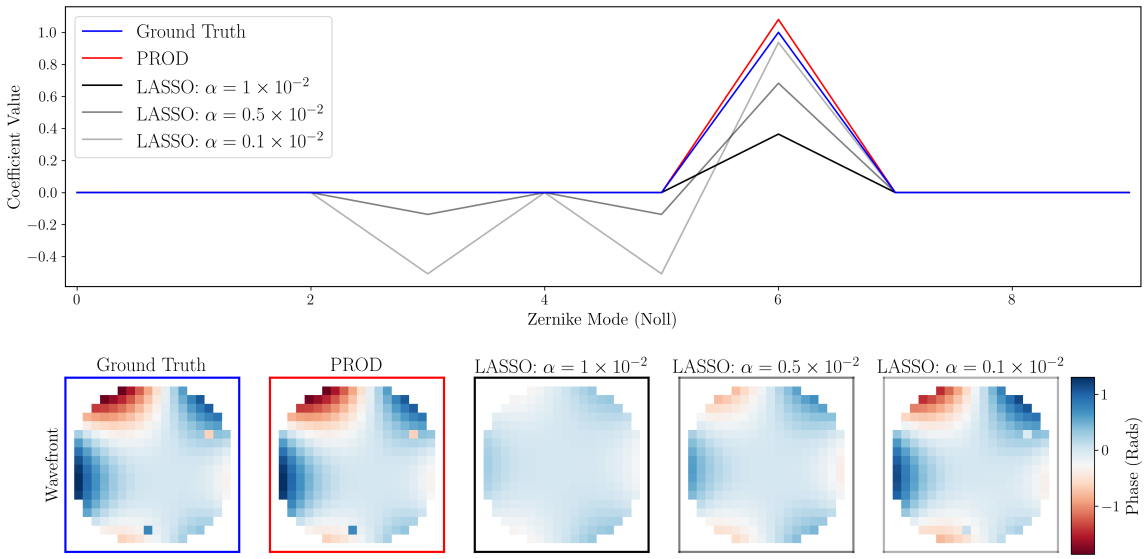


Fig. 4.12 Demonstrating the utility of PROD's rigid transform module to efficiently represent de-centered modes, in this case the Z_3^3 mode. Here PROD learns the de-center parameter, and identifies the single Zernike mode. Also displayed for comparison is the LASSO technique, which is given the same over-complete dictionary, but possesses no ability to transform the modes. All wavefronts are plotted on the same colour scale. If it's L1 penalisation parameter, α, is large, it cannot represent the wavefront accurately, whereas if α is small, it must utilise lower order Zernike modes to help represent the wavefront.

in the group of Dr. Andreas Döpp), using a Spectra-Physics Spitfire ACE laser system to generate ultrashort pulses with central wavelength, $\lambda_0 = 798$ nm, and a transform limited duration of 64 fs. To generate the vortex, a setup was used that has been thoroughly discussed in previous work [220, 221]. Firstly, a quarter-wave plate (QWP) was used to circularly polarise the incoming pulse, before it passed through a ($l = 2$) structured wave plate (s-wave plate), creating two circularly polarised vortices with opposing handedness. A further QWP was used to transform them into two linearly polarised vortices before a linear polariser was used to select one of them. The optical setup used here was not sensitive to spectral components, so the signal was integrated over time when captured. The wavefront measurement was then performed on a home-built Shack Hartmann sensor (focal length, $f = 14.2$ mm, and pitch, $\Lambda = 300$ µm) [132]. The microlens foci for the vortex are seen in Fig. 4.13. A reference microlens foci pattern was first captured, before the s-wave plate was introduced to the optical setup.

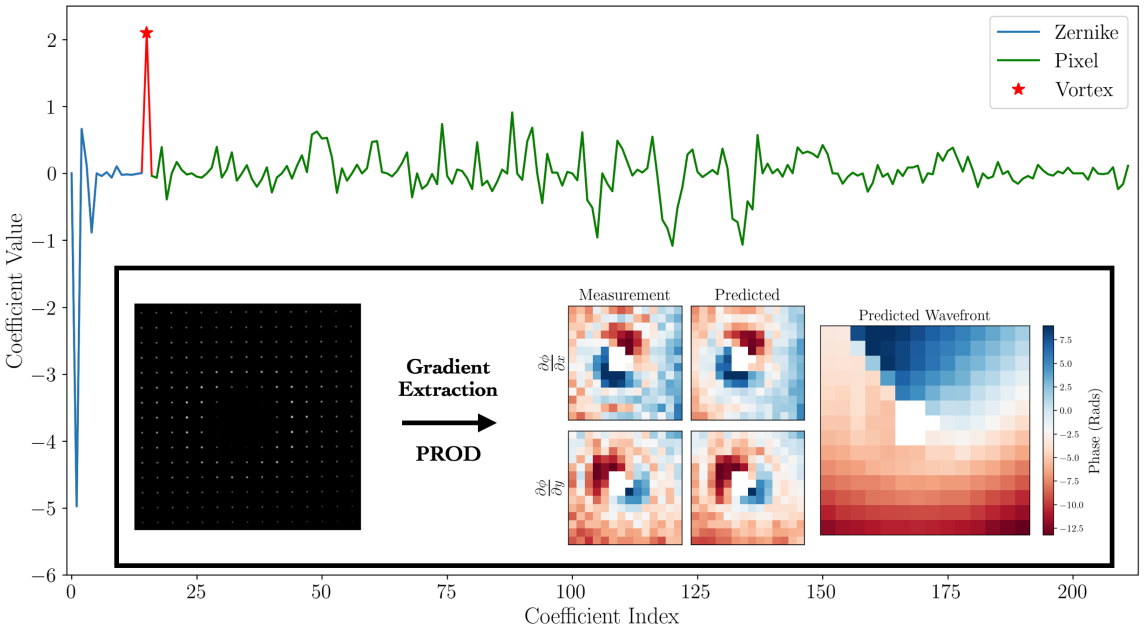


Fig. 4.13 The use of PROD to characterise an experimental optical vortex. Once the gradients are extracted from the Shack Hartmann pattern, the technique is used to find the modal coefficients for the vortex, and lower order Zernike modes. The coefficient for the vortex (the topological charge) was 2.09, in good agreement with the theoretical value for this setup of 2.

I performed the following wavefront analysis from this raw data using the PROD technique. The centroids of the microlens foci were extracted, before the wavefront gradients were found by the calculating the centre-of-mass difference between the vortex and the reference centroids. The PROD technique was then used to extract the modal coefficients and the wavefront. The results are displayed in Fig. 4.13. It is clear that PROD was able to correctly identify the 2nd order vortex, with a predicted modal coefficient of 2.09. This discrepancy of 5% relative to the theoretical value can be attributed to possible chromaticity in the vortex retarder (as seen in the following chapter), and to misalignment in the distance between the sensor and the microlens array, and imperfection in the vortex generation.

4.2.5 Conclusions and Outlook

Leveraging an over-complete dictionary and trainable rigid transforms, PROD has been introduced as a robust technique to reconstruct a wavefront from measurements of its

gradients. Where most existing techniques typically reconstruct in the pixel or the Zernike basis, there exist limitations with both however; the former overfits to noise and doesn't give physically interpretable coefficients, and there are some specialised wavefronts that cannot be efficiently expressed in the Zernike basis, such as the optical vortex. Exploiting one's prior knowledge about a measurement, PROD is able to extract the physically-meaningful modal coefficients for the modes that are within its over-complete dictionary. Furthermore, as its dictionary also incorporates the pixel basis, it can express noise without influencing the other modal coefficients, such as those for Zernikes and special modes.

In its presented form, a PROD reconstruction of a 20×20 wavefront takes on the order of seconds, meaning it is principally for post-analysis. However, if one assumes that all modes are centred and removes the rigid transform module of this technique, Lasso can be used to solve the L1 problem in a far shorter time (ms), allowing a real-time reconstruction for applications in adaptive optics. While this may be suitable in some cases, such as when one is certain that the system is centred, in the most general case this is not a suitable assumption, for example with a decentered vortex, and can lead to an under-sparse representation, and as a result, inaccurate modal coefficients.

This application of over-complete dictionaries and the principles of sparsity can also be applied to other problems in laser physics. For example, the reconstruction of the spectral phase from measurements can also be performed in different bases, and sometimes involves switching between bases [222, 223]. Furthermore, with sufficient computational resources, there is no reason why this technique cannot be extended to recent work in the spatiospectral regime, where the wavefront is considered in three dimensions [132].

This chapter has presented two methods that resulted in publications; firstly, CoordGate, to efficiently perform image denoising (deconvolution) for the case of spatially varying point spread function. This was developed to be part of a larger reconstruction pipeline, where snapshot compressive imaging reconstruction would reconstruct blurred hypercubes from the coded shot, before this technique could be used to remove the blur from the hypercube. However, as discussed in Chapter 5, considerations of the resolution required to

capture physically relevant information in laser characterisation led to a shift in approach, and CoordGate was ultimately not applied. Nonetheless, it is hoped that this method will prove useful to the broader scientific community. The other method presented here utilises overcomplete dictionaries to perform wavefront reconstruction, and is used in the final diagnostic in the following chapter.

Chapter 5

Single-shot Vector Field Measurement of Ultra-Intense Laser Pulses

As detailed in Chapter 2, in order to capture the vectorial field structure of a collimated ultra-intense laser pulse, one must capture the intensity and phase across two spatial dimensions, time/spectrum, and polarisation, resulting in a final dimensionality of $(2 \times 2 \times n_x \times n_y \times n_\omega)$; this poses a significant challenge considering the two-dimensional nature of current sensors. In the field of metrology, there is always a fundamental question regarding the necessary resolutions required to resolve the quantity of interest. Nearly all existing ultra-intense laser characterisation techniques measure the pulse in the near-field, where the intensities are lower which is necessary to avoid damage. From this perspective, it is perhaps intuitive that the principally desirable property of a measurement device is to resolve the near-field in as high detail as possible, spatially and temporally. Indeed, spatially-resolved spectral multi-shot methods typically capture the information on a CMOS sensor, which results in the thousands of spatial samples across the signal's domain, and typically these methods scan to resolve the spectral profile of the pulse in great detail.

However, it is interesting to note that the information that is actually relevant is the pulse structure at focus; this is what will be consequential for simulations and experimental applications. As has been described in Section 2.2, when focused by a lens that causes only

small angles from the optical axis, so that one remains in the Fresnel regime, the operation of focusing of laser pulse is captured by the Fourier transform [74]. This provides a simple relation between the range and resolution in each of the near and far fields, according to the Nyquist criterion. One can define a target focal volume of the measurement as a box covering $(-k_{x,max}, +k_{x,max}) \times (-k_{y,max}, +k_{y,max}) \times (-t_{max}, +t_{max})$, and can then calculate the respective resolution needed for a spatio-spectral measurement in the near field as

$$(\Delta x, \Delta y, \Delta\omega) = \left(\frac{\pi}{k_{x,max}}, \frac{\pi}{k_{y,max}}, \frac{\pi}{t_{max}} \right).$$

Both the spatial and spectral ranges of a laser pulse are naturally constrained in the near field, being the beam size D and the spectral width $\Delta\omega_{pulse}$, which then leads one directly to the required number of measurements in this domain, i.e. $n_x = D/\Delta x$, $n_y = D/\Delta y$, and $n_\omega = \Delta\omega_{pulse}/\Delta\omega$. An important question revolves around how large to make this focal volume (the corresponding values of $k_{x,max}, k_{y,max}$ and t_{max}). It is true that with this approach, any frequencies that are focused outside the box may cause errors due to aliasing; thus it is important to capture the vast majority of the energy of the pulse within the box. This is made easier by the fact that for an ultra-intense laser pulse to satisfy its own definition (to be ultra-intense), the signal must be tightly compressed both spatially and temporally in the far field.

One may apply these principles to a real facility, for example the ATLAS-3000 petawatt-class laser at the Center for Advanced Laser Applications, which generates pulses that have a flat-top spatial profile in the near field with diameter $d = 28$cm, a pulse duration of $T = 29.6$ fs, and a spectrum between $750 - 850$ nm – corresponding to frequencies, ω, in the range $2.20 - 2.52$ rad PHz (and therefore $\Delta\omega_{pulse} \approx 0.32$ rad PHz). According to the principles of Fourier optics, the corresponding spatial far-field profile will be the Airy disk. For an ideal pulse, 95% of the energy is contained in the region up to the 5^{th} ring, which corresponds to a $k_{\max} = 112.2$ rad m^{-1} (see Appendix B). Similarly, one can define the temporal region of interest to be a multiple of the pulse duration, so that the majority

of the energy is contained ($t_{max} \sim 100$ fs). In order to resolve the spatiotemporal profile at the focus, one then uses the Nyquist sampling relation to find the required resolutions in the near field,

$$\Delta x = \frac{\pi}{k_{max}} = 0.028 \text{ m} = \frac{d}{10}, \tag{5.1}$$

$$\Delta\omega = \frac{\pi}{t_{max}} = 1.57 \text{ rad THz} = \frac{\Delta\omega_{pulse}}{9.5}. \tag{5.2}$$

This reveals the interesting property of ultra short laser pulses, namely that the region of physical interest can be resolved with a relatively coarse resolution; within the previously defined focal volume, both the spatial and spectral extent of the near-field need to be sampled around 10 times. In-fact, if one captures finer details in the near field with a higher resolution, the extra information does not appear in the focus at the far field, but rather is diffracted to higher orders. One notes that increasing the resolution of the far-field is as simple as zero-padding the near-field before numerical propagation, with the prior knowledge that it is constrained by the limited domain of the pulse.

Another similar insight emerges from viewing this property from the perspective of signal processing; in particular from the Wiener-Khinchin theorem, where the high concentration of energy at focus necessitates a large scale length in the autocorrelation function of the near-field. As shall be demonstrated in the proceeding discussion, this fundamental property of smoothness in ultra-intense laser pulses provides additional conditioning to the reconstruction problem, giving the prior knowledge to select smooth solutions, enabling the capturing of the vector field in a single shot.

The preceeding analysis is derived from focusing in the Fresnel regime, which accordingly gives the principles of Fourier optics. In some applications (such as ion acceleration), one may wish to focus the laser pulse tightly, with large angles relative to the optical axis. As long as one can make assumptions about the ultra-intense laser pulse's power spectrum (the volume the pulse *would* occupy when focused in the Fresnel regime), they can still infer properties of the underlying coherence structure in the near field. Then one may

use a different propagator (such as Rayleigh-Sommerfeld) in order to study the focus of a pulse under strong focusing.

The following section describes how this information can be leveraged to design RAVEN, a measurement device for the *Real-time Acquisition of Vectorial Electromagnetic Near-fields*, that uniquely encodes the spatio-spectral vector field in the near field and a software pipeline to decode that information. This is followed by a series of experimental results using the ATLAS-3000 petawatt laser in Garching, demonstrating the full capabilities of the designed device.

5.1 Technique

The design can be broken into the two steps of a *physical encoding* of the relevant information, followed by a *computational decoding* to reconstruct the pulse. Various steps of this reconstruction process required a *calibration* of the device, and these three steps are described in this section.

5.1.1 Physical encoder

The previously introduced sampling rates are far smaller than is typically aimed for in ultra-intense laser characterisation, and provide an opportunity to trade off the spatial resolution of the sensor in order to capture the information of different dimensions. Here an optical system is proposed, shown in Fig. 5.1a, to encode the phase, spectral and polarisation information into the intensity before it is captured by the sensor.

The design begins with the simple Shack Hartmann sensor [36] (Thorlabs MLA150-7AR, $f_{ml} = 5.2$ mm). As described in detail in Section 2.4.2, this works by means of a microlens array, which locally samples the wavefront in many locations and encodes the average phase gradient over each sub-aperture into the resulting shifts of the focal spots. This averaging of the local wavefront proves crucial as it both reduces aliasing from higher frequency components and leads to a better signal-to-noise ratio compared to other phase

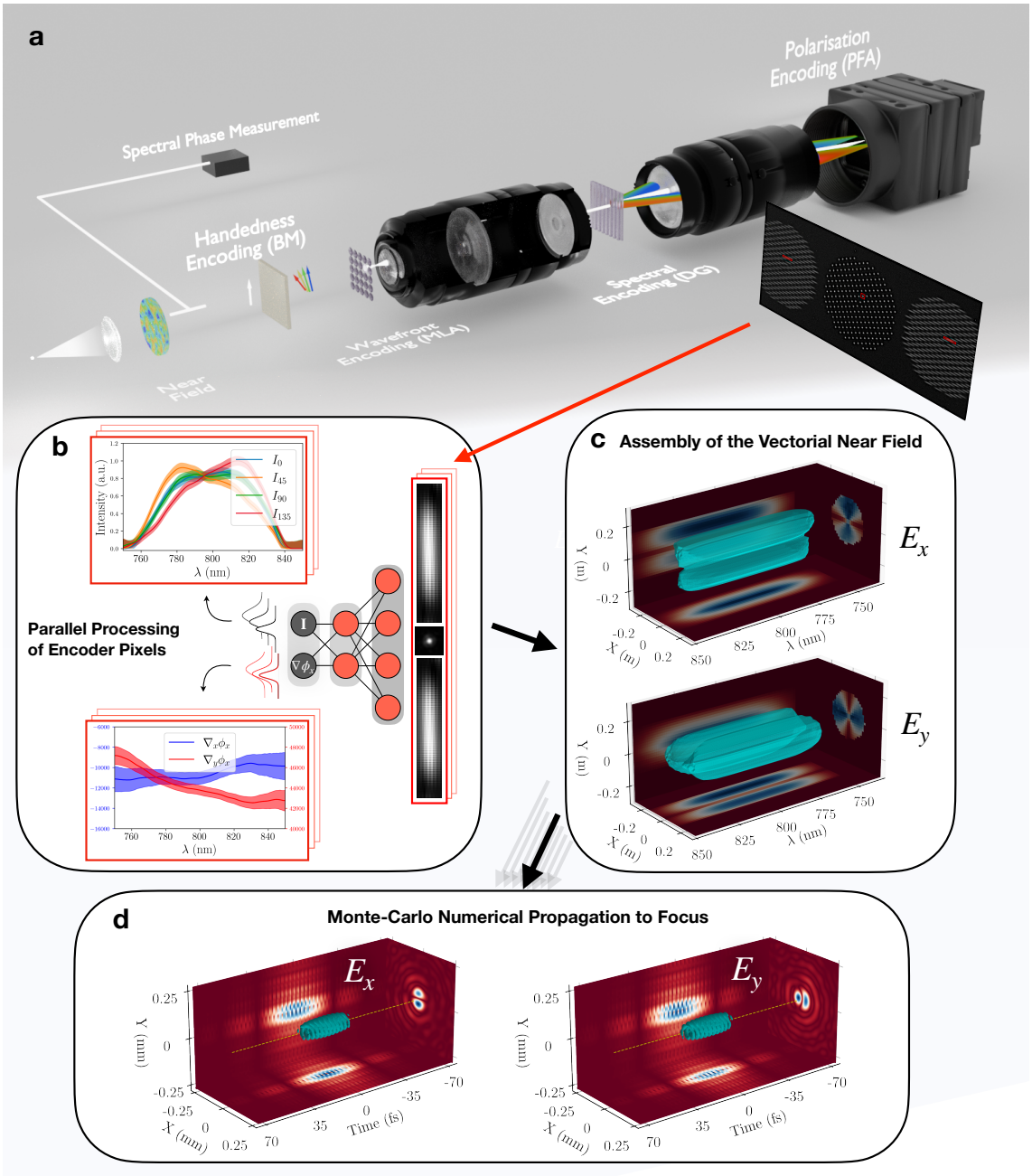


Fig. 5.1 **The RAVEN technique to perform single-shot measurements of vector fields.** **(a)** The optical setup encodes the four dimensional vector field onto a two-dimensional intensity measurement (not to scale). The collimated beam is split, with one part used to measure the spectral phase and the rest passing through a birefringent medium (BM), creating a relative spectral chirp between the polarisations. Next, a microlens array (MLA) encodes the wavefront. The resulting pattern is imaged by a 4f system with a diffraction grating (DG) placed near the Fourier plane, providing the spectral encoding, and a polarisation filter array (PFA) is used on the sensor. **(b)** Extraction and processing of the encoding pixels. At each point in space, the spectral and polarisation information is recovered, with uncertainty estimates. **(c)** Aggregation of the information from the encoder pixels into the vectorial near field. **(d)** Propagation to focus.

encoding methods, e.g. the pinhole-based Hartmann sensor. In order to extend this to a hyperspectral setting, the plane of the foci is imaged using a 4-f system, with a transmissive diffraction grating (Edmund Optics #46-074, 110 lines/mm) placed in the Fourier plane; this is shown in Fig. 5.2. The lenses used in the 4-f system were the MVL50TM23 model ($f_l = 50$ mm), also from Thorlabs. Naturally, slight deviations in the distances between optical components may arise from alignment inaccuracies in the experimental setup. The resulting diffraction pattern at the sensor is composed of the fundamental order, as well as the ± 1 orders, resulting in a kind of sparse tomography. Crucially, the ± 1 orders obtain chromatic dispersion, thereby encoding the spectral information in streaks at the detector. Analogous to tomographic reconstruction, these streaks are projections of the hyperspectral intensity, while the hyperspectral phase causes shifts in the streaks. One notes that it is necessary to have both of the ± 1 orders, as with just one there exists an ambiguity between spectral intensity and phase – this can be seen in Appendix C. A rotation was applied to the microlens array in order to ensure that there was no overlap between the streaks for neighbouring spatial locations.

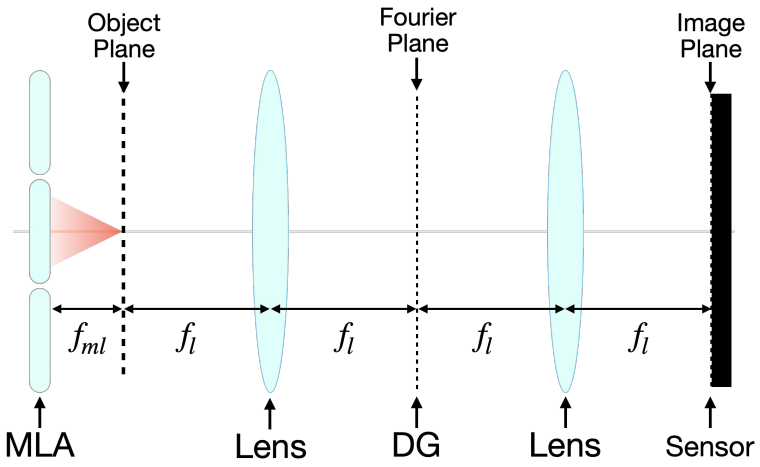


Fig. 5.2 A further diagram of the optical set up in RAVEN. A microlens array (MLA), with focal length f_{ml}, causes a pattern of foci to form in the object plane. This is imaged onto a sensor with a 4-f system, using two lenses with focal length f_l. A diffraction grating (DG) is placed in the Fourier plane to disperse the foci into streaks.

To fully resolve the polarisation state, the Stokes parameters must be captured, restated from Section 2.1.1 as,

$$\vec{S} = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} (E_{x0})^2 + (E_{y0})^2 \\ (E_{x0})^2 - (E_{y0})^2 \\ 2E_{x0}E_{y0}\cos(\delta) \\ 2E_{x0}E_{y0}\sin(\delta) \end{bmatrix}, \tag{5.3}$$

where E_{x0}, E_{y0} are the amplitudes of the orthogonal field components, and $\delta = \delta_x - \delta_y$ is the relative delay between the components. Here a camera is used which contains a polarisation filter array (in particular, a Lucid Vision PHX050S-PC, based on the Sony IMX250MZR chip). This consists of a mosaic of 2×2 pixels, containing linear polarisers with each of the angles $\in [0°, 45°, 90°, 135°]$, which gives a measurement vector with direct access to I_H, I_V, I_D and I_A for each spatial location. The measurement vector can be written as,

$$\vec{M} = \begin{bmatrix} I_H \\ I_V \\ I_D \\ I_A \end{bmatrix} = \begin{bmatrix} (E_{x0})^2 \\ (E_{y0})^2 \\ I_H + I_V + 2E_{x0}E_{y0}\cos(\delta) \\ I_H + I_V - 2E_{x0}E_{y0}\cos(\delta) \end{bmatrix}, \tag{5.4}$$

and thus, the first three Stokes parameters are found easily,

$$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} I_H + I_V \\ I_H - I_V \\ I_D - I_A \end{bmatrix}. \tag{5.5}$$

In order to resolve S_3, one must find δ. By rearrangement of the formulas for the measurements, this value can be retrieved up to a sign ambiguity due to the oddness of the Cosine function,

$$\pm\delta = \arccos\left(\frac{I_D - I_A}{2\sqrt{I_H I_V}}\right). \tag{5.6}$$

Intuitively, this ambiguity pertains to the chirality, or handedness, of the polarisation (meaning whether it is right-handed or left-handed cannot be discerned). In the ellipse representation, the magnitude of the ellipticity angle χ can be determined, but not its sign. In RAVEN, the ambiguity is removed by a chromatic birefringent medium.

One may apply a characterised phase delay, δ_B, to one of the orthogonal polarisations by means of a birefringent medium. The electric field vector at the detector is found by multiplication with the Jones matrix for a birefringent medium:

$$\vec{E'} = \mathbf{Q}\vec{E} = \begin{bmatrix} e^{i\delta_B} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} E_{x0}e^{i\delta_x} \\ E_{y0}e^{i\delta_y} \end{bmatrix} = \begin{bmatrix} E_{x0}e^{i(\delta_x+\delta_B)} \\ E_{y0}e^{i\delta_y} \end{bmatrix}. \tag{5.7}$$

The new measurement vector is given as,

$$\vec{M'} = \begin{bmatrix} I'_H \\ I'_V \\ I'_D \\ I'_A \end{bmatrix} = \begin{bmatrix} (E_{x0})^2 \\ (E_{y0})^2 \\ I_H + I_V + 2E_{x0}E_{y0}\cos(\delta + \delta_B) \\ I_H + I_V - 2E_{x0}E_{y0}\cos(\delta + \delta_B) \end{bmatrix}. \tag{5.8}$$

The effect of the addition of the birefringent medium is elegantly displayed in the ellipse framework, as seen in Fig. 5.3.

Considering this monochromatic case, the birefringent medium always works to make the resultant ellipse more right-handed, by adding a delay, δ_B. There still exists a similar ambiguity in this setting. When a pulse with an initial polarisation delay of δ_0, passes through the birefringent medium, the total delay is, $\delta_M = \delta_0 + \delta_B$. When the ambiguous measurement passes through the birefringent medium, its polarisation delay must be $-\delta_M$ (thereby the ellipse has the same shape). One can then calculate the initial delay of the

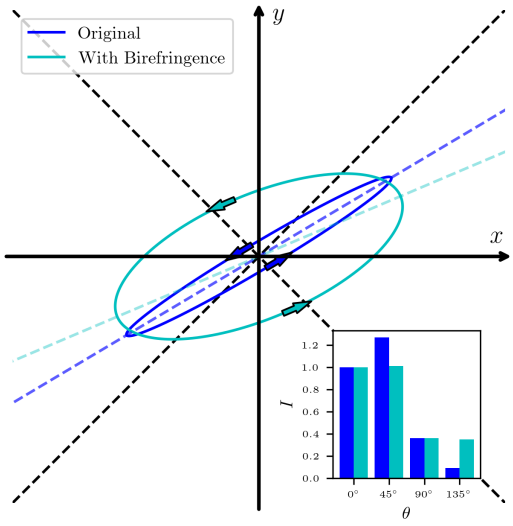


Fig. 5.3 Demonstrating the change in polarisation ellipse upon the addition of a birefringent medium. The subplot displays the corresponding intensity values that are measured by the polarisation filter array (I_H, I_D, I_V, I_A). The birefringent medium adds a phase shift of $\pi/4$.

ambiguous measurement, $\delta_0^{(A)}$, by,

$$\delta_0^{(A)} + \delta_B = -\delta_M \tag{5.9}$$

$$\delta_0^{(A)} = -\delta_0 - 2\delta_B \tag{5.10}$$

In general, this means there are still two possibilities for the original ellipse as shown in figure Fig. 5.4, where it could either be a left-handed ellipse that was made more circular by the birefringent material, or a right-handed ellipse that was made less circular.

At this point it seems that not much has been gained with the addition of the birefringent medium, as there still exists a binary ambiguity between left and right-handed polarisation. However, this is changed upon consideration of the hyperspectral light, and the prior assumption of smoothness. A typical birefringent medium is chromatic, so that the phase shift between polarisations is spectrally dependent, $\delta_B = \delta_B(\lambda)$. This is visualised on the ellipse in Fig. 5.5.

According to the smoothness of the initial field, one may assume that the spectral variation of the polarisation delay, $\left\| \frac{d\delta_0(\lambda)}{d\lambda} \right\|$, is small and then choose a birefringent medium

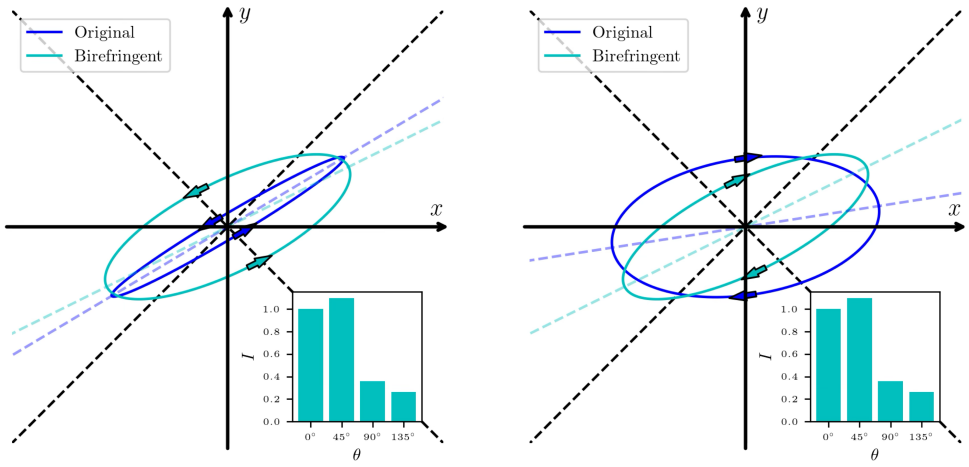


Fig. 5.4 Two original pulses that, once the birefringent material is applied, yield the same shaped ellipse (one is left-handed and one is right-handed), and thus the same measurements, demonstrating the ambiguity.

so that,

$$\left\| \frac{d\delta_0(\lambda)}{d\lambda} \right\| \ll \left\| \frac{d\delta_B(\lambda)}{d\lambda} \right\|. \tag{5.11}$$

Consequently, the field at the detector will have a polarisation delay that has a spectral variation in the same direction as the birefringent medium. When a measurement is taken, the two ambiguous solutions for the delay at the detector are, $\delta^{(\pm)}(\lambda) = \pm\delta_M(\lambda) = \pm(\delta_0(\lambda) + \delta_B(\lambda))$. The two solutions are differentiated by the sign of spectral variation, $\frac{\partial\delta^{(\pm)}(\lambda)}{\partial\lambda}$, and the correct solution is found easily, as the one that has a spectral variation in the same direction as the birefringent medium. The initial field delay, $\delta_0(\lambda)$, is then found by subtracting the known delay of the birefringent medium.

To assess the sensitivity that this permits in the spectral variation of the delay of the initial field, this condition will select the correct solution, as long as,

$$\left\| \frac{d\delta_0(\lambda)}{d\lambda} \right\| < \left\| \frac{d\delta_B(\lambda)}{d\lambda} \right\|. \tag{5.12}$$

A typical birefringent crystal, such as a half-wave plate, adds a delay dependent on wavelength, $\delta_B(\lambda) = \omega(\lambda)\tau = \frac{2\pi\tau c}{\lambda}$. Consequently, choosing a larger delay, τ, will increase the technique's sensitivity in the spectral variation of the polarisation delay. One notes

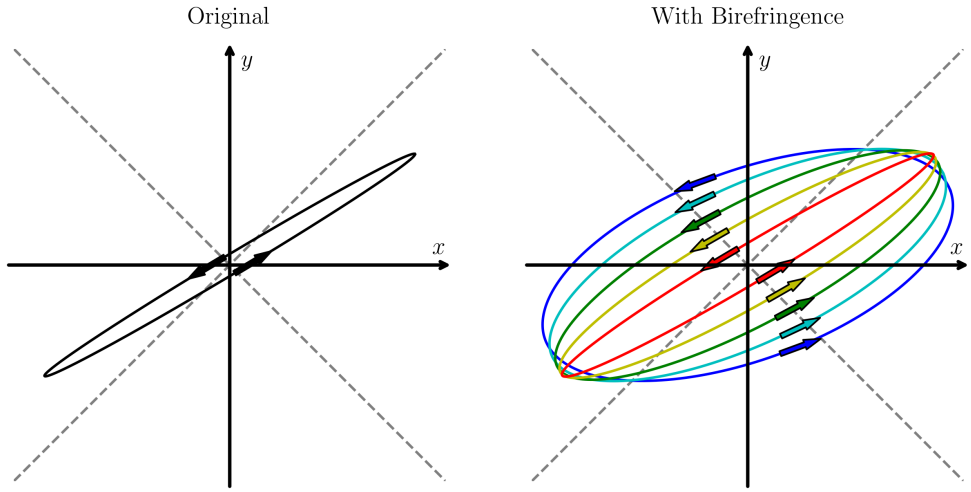


Fig. 5.5 An original pulse with the same ellipse for all wavelengths. When passing through the birefringent medium, each wavelength obtains a different phase shift.

that this benefit of increasing the delay of the medium only increases until a point. Once the delay becomes large, there will be an aliasing effect due to the binning of wavelengths. While this may be technically possible to calibrate away, it adds a risk of overfitting.

One final point is to mention a slightly different intuition behind the process of removing the ambiguity. From the two possible resultant delays, one can find the two possible initial delays by subtracting $\delta_B(\lambda)$. Of course, one of these solutions will be the true, desired initial delay, δ_0. The other solution is,

$$\delta_0^{(A)}(\lambda) = -\delta_0(\lambda) - 2\delta_B(\lambda) \tag{5.13}$$

Considering that one assumes that the spectral variation of δ_0 is small, it follows that $\left\| \frac{d\delta_0(\lambda)}{d\lambda} \right\| \ll \left\| \frac{d\delta_0^{(A)}(\lambda)}{d\lambda} \right\|$. Thus, to select from the two possible initial solutions, one can apply total variation along the spectral axis, and select the solution that has a smaller value. This is the approach used in the vortex measurement in the results section of this chapter.

One notable limitation of this current prototype of RAVEN is the relative dynamic range in the diffraction orders. The energy diffracted to each order by the diffraction grating is $\sim 28\%$ in the fundamental and $\sim 25\%$ in the ± 1 orders. As the ± 1 orders obtain dispersion, the intensity values are significantly lower. When taking measurements

this leads to a problem, of either taking a measurement that isn't over-exposed in the fundamental but has low signal in the diffracted orders, or allowing the fundamental to be overexposed. Here the latter was used due to the fact that the ± 1 orders contained more information, and a preprocessing network was used to remove the overexposure effect from the fundamental (as is described below). This is something to correct in future generations of RAVEN.

5.1.2 Software decoder

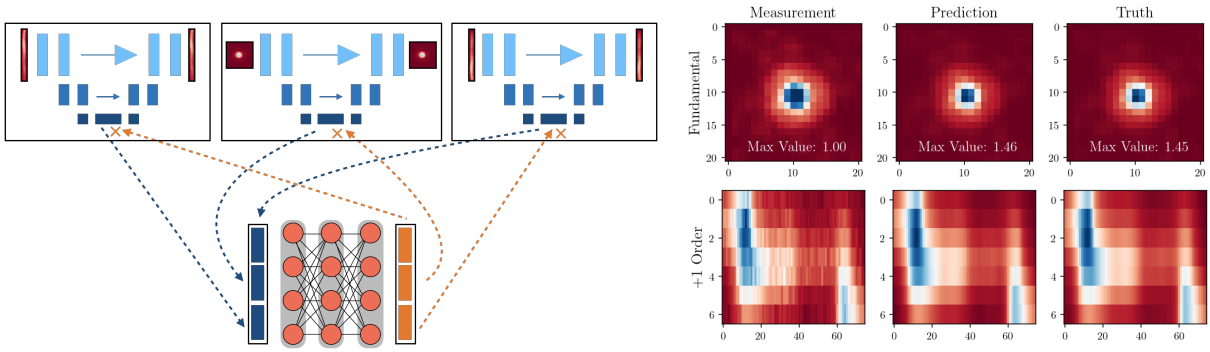
Due to the unique encoding achieved by the optical setup (Appendix C), the system is theoretically bijective, meaning that there exists a one-to-one relation between each possible field and the corresponding measurement. However, real-world measurements invariably contain noise, which complicates the reconstruction process. While convex optimisation techniques could be employed, they are typically relatively slow and less adept at handling noise and uncertainty estimation. Thus, to achieve a real-time diagnostic capability that matches the repetition rate of ultra-intense laser systems, neural networks are employed. This approach not only provides rapid reconstruction (taking $<$0.1s for the near-field) but also excels at denoising the data and estimating uncertainties in the reconstruction. For the following discussion, the spatiotemporal indices $(x, y, \lambda/t)$ are suppressed for brevity as they apply to every variable – conversely, where indices are shown, they are the sole indices.

The reconstruction process begins by the extraction, and subsequent analysis of the *encoder pixels*. From a raw measurement image, the triplet of streaks for each spot (the fundamental and the ± 1) can be obtained by first using a simple peak finder (`skimage.feature.peak_local_max`) to identify and extract the peak in the fundamental order, before using calibrated dispersion curve to extract rectangular patches around the dispersed spots in the ± 1 diffraction orders. This process is demonstrated in Fig. 5.1b. Each of these encoding pixels are then normalised by the integrated intensity in the fundamental

before being analysed in parallel by a deep neural network, which makes a prediction for the intensity in each polarisation axis, $\mathbf{I} = [I_0, I_{45}, I_{90}, I_{135}]$, and the derivatives of the wavefront of one of the polarisations, $\boldsymbol{\nabla\phi_x} = \left[\frac{\partial\phi_x}{\partial x}, \frac{\partial\phi_x}{\partial y}\right]$, at each wavelength, λ. Analysing the encoding pixels separately greatly reduces the dimensionality of the input space of the reconstruction algorithm, and consequentially enables a more robust reconstruction. Crucially there is no possibility for the neural network to overfit to spatial wavefront patterns, as they are analysed separately, allowing for a greater degree of trust in the predictions. Furthermore, the encoder pixels for each spatial location can be put into the batch dimension of the neural network and analysed in parallel, meaning that the process is fast (~ 0.1 s to reconstruct the near-field).

The structure of the network consisted of two parts. A first convolutional ('preprocessing') network was used to remove the measurement distortions (noise and overexposure) from the triplet, whilst also predicting an estimate of the noise level. The architecture of the preprocessing network involved three U-Nets [148], to correct the spot's projection in each of the three orders. However, it is intuitive that this denoising should not be independent for all three orders. Thus, an attention mechanism was introduced to allow information to propagate between the U-Nets. At the lowest level (the smallest spatial extent) of the U-Nets, the layers were flattened and concatenated with the same layer from the others. Then this tensor was fed into a fully connected neural network, yielding 'coefficients'. These were then multiplied by the U-Net's lowest layer before the upsampling continued. A diagram of the preprocessing network is displayed in Fig. 5.6, along with results for a synthetic measurement that was both noisy and overexposed.

Following the preprocessing, the second part of the network made predictions of the intensity and phase gradients. Due to the low dimensionality of the input and output, a simple fully connected network was used. The network was also given the spatial location of the specific encoder pixel, so that it could learn the spatially varying intensity response of the RAVEN hardware. Here the principles of variance networks were utilised in order to estimate the network's error in its predictions. Churchill Eisenhart, former Chief of the



(a) Preprocessing Network

(b) Preprocessing results.

Fig. 5.6 a) The preprocessing network. The fundamental and ±1 orders are analysed in individual U-Nets, while attention is used to allow the networks to communicate. b) Performance of the preprocessing network. For compactness, here is shown only the fundamental and +1 orders for a single polarisation.

Statistical Engineering Laboratory, once said *"A reported value whose accuracy is entirely unknown is worthless"* [224]. The reconstruction method here presents a full treatment of the uncertainty in the prediction.

To train a neural network for regression, the most commonly used loss function is the mean squared error. While this indeed allows the network to learn efficiently [225], it results in predictions that are not accompanied by any prediction of confidence. However, in a physics setting one also needs to know about the uncertainty in order to quantify the precision of a measurement. Thus, if instead one assumes a prior distribution on the uncertainty of the prediction (e.g. Gaussian), then a network can be trained to predict the parameters of the distribution, by maximising the likelihood of its predictions – in actuality, minimising the negative log likelihood (NLL) [226, 227]. This prediction of the variance of the distribution allows the network to signify when it is unsure (perhaps due to a particularly noisy input). For example, training with this loss function will heavily punish a prediction that is wrong and made with low variance, versus one that is made with high variance. In this work, the prior distribution on the phase derivatives is chosen to be Gaussian so that a mean and variance is predicted. The NLL for this distribution is as follows;

$$\mathrm{NLL}(\mu, \sigma^2, y) = \frac{\log(2\pi\sigma^2)}{2} + \frac{(\mu - y)^2}{2\sigma^2} \tag{5.14}$$

This distribution is less appropriate for the intensity predictions however, due to the fact that it can take values below 0. Thus, training with the NLL for the standard Gaussian biases the network to predictions with more positive values. Instead, for the first time to the author's knowledge, here the folded Gaussian distribution is utilised for this purpose [228],

$$\mathrm{NLL}(\mu, \sigma^2, y) = \frac{\log(2\pi\sigma^2)}{2} - \log\left(\frac{-(\mu - y)^2}{2\sigma^2} - \frac{(\mu + y)^2}{2\sigma^2}\right) \tag{5.15}$$

This was implemented using the numerically-stabilised `torch.logsumexp` function, and resulted in the subsequent reduction of the NLL from -6.5 to -7, when compared to the standard Gaussian. This process is shown in Fig. 5.7, as well as Fig. 5.1. The final network made predictions for 101 spectral channels between 750 nm and 850 nm, and estimates for the mean and standard deviations of the distributions, so the network's output had a shape: $(\text{batch size}, 2, 101, 6)$. The training of this network is described in Section 5.1.3.

The vectorial near field is then formed by the synthesis of the encoding pixels. The first thing to note is that the derivative predictions are in the horizontal and vertical directions, relative to the camera. As was previously mentioned a rotation was applied to the microlens, to allow for the maximum size of the streaks without overlap. The usual format for the input to zonal phase stitching methods is the horizontal and vertical derivatives, relative to the grid which in this case is the microlens. Therefore, the derivatives must be rotated by the angle of rotation of the microlens array, so that they are horizontal and vertical relative to the microlens. This is done simply at each microlens spot by the following formula:

$$\begin{bmatrix} s_x^{\mathrm{ml}} \\ s_y^{\mathrm{ml}} \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} \begin{bmatrix} s_x^{\mathrm{cam}} \\ s_y^{\mathrm{cam}} \end{bmatrix}, \tag{5.16}$$

where s^{ml} and s^{cam} are derivatives relative to the microlens and camera respectively, and θ is the angle of rotation of the virtual microlens array.

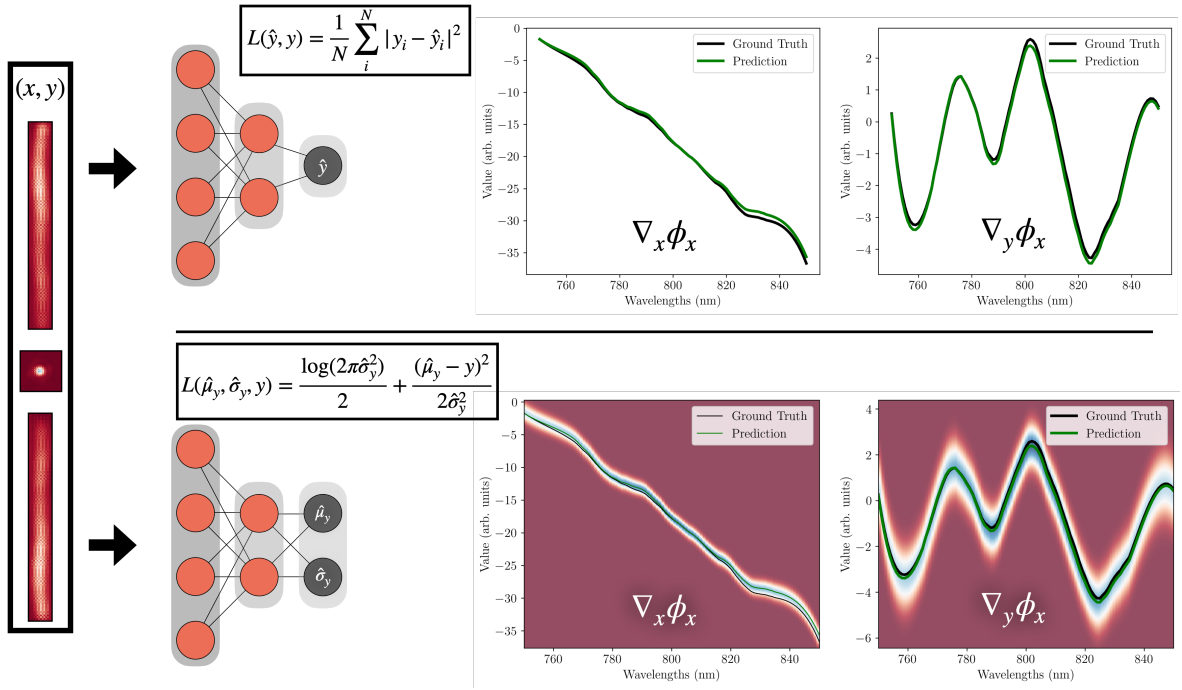


Fig. 5.7 Displaying the difference between traditional neural networks and variance networks. Top row is the output of a network makes a point estimate by minimising the mean squared error. Bottom row is a variance network, which minimises the negative log likelihood, and leads to a prediction of a distribution. Predictions for the intensity components are also made, but not shown here.

From the hyperspectral phase derivatives, a number of proceeding steps can be applied depending on the desired result. For example, the modal spatio-spectral wavefront retrieval method (Section 2.4.6) can be applied to directly recover the values of spatiotemporal couplings. Alternatively, the derivatives can be stitched together utilising the zonal approach [229], and the least squares (pseudoinverse) solution, yielding ϕ'_x. Due to the fact that both these methods involve a simple matrix multiplication, the variance in the phase gradients can also be propagated to the final result. For example, the phase is obtained from its stitched slopes via a linear equation, $\phi = [\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T\mathbf{s}$; thus the variance in the phase gradients can also be propagated simply using the element-wise squared pseudoinverse matrix, $\sigma_\phi^2 = ([\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T).^2\sigma_s^2$. In order to obtain the spatiotemporal phase, the spectral phase (which is measured separately with a WIZZLER) must also be added: $\phi_x = \phi'_x + \tilde{\phi}(\lambda)$. The polarisation delay, δ, is extracted from the intensity predictions, as described previously, so that ϕ_y is then found as $\phi_y = \phi_x - \delta$, without the

need of a separate spectral phase measurement in the other polarisation direction. The intensity predictions are re-multiplied by the normalisation constant. The result of this analysis is the complete near field with uncertainty estimates, which can be obtained at a rate of > 10 Hz.

Monte Carlo Propagation and Covariance Estimation

It is the field at focus which is of physical significance. Thus, the near-field is numerically propagated to the far field using Fourier optics. Being a linear system, there has been work on propagating uncertainties through Fourier transforms [230, 231]. However, a key issue is that in the case of RAVEN the uncertainty is predicted on the amplitude and phase rather than the real and imaginary parts of the field. This means the error propagation is actually nonlinear, and proved to be unstable when an attempt was made to implement it. Therefore, to propagate the uncertainty through the far-field propagation, Monte-Carlo sampling is used. For a number of samples, the intensity and phase are sampled from their distributions in the near-field, before the resulting field is propagated to the focus. There, statistics for quantities of interest (eg. peak intensity or pulse duration) are collected, such as the mean and standard deviation.

An important detail is that the neural network does not predict a full covariance matrix for the intensities and phase derivatives, but instead treats each spectral component independently, outputting a variance for each. Implicitly this assumes a diagonal covariance matrix; however, one knows that the neighbouring spectral components are correlated (from the prior knowledge of smoothness), and thus these variances are overestimated. Another way of seeing this, is while the network predicts 101 discrete wavelengths (1 nm steps between 750 nm and 850 nm), the true resolution is lower, as the spots blur into each other in the streaks, thus causing the correlation.

In order to perform sampling from the intensities and phase derivatives, the variance estimates must be converted into an effective covariance matrix. To account for this effect, consider the initial spectral signal prediction (e.g. phase derivative), to have mean values

described by vector μ_y, and each spectral component is considered to be independent, with a vector of variances, σ_y. Here one desires to incorporate the knowledge of the correlation between spectral components, so that they can generate a covariance Σ matrix, and subsequently sample. A correlation matrix is generated, $\mathbf{C}$, using Gaussian's with a standard deviation proportional to the measured microlens foci width. The correlation matrix is then found by $\mathbf{\Sigma} = \sigma_y \mathbf{C} \sigma_y^T$. Cholesky decomposition is subsequently performed in order to obtain $\mathbf{L}$, where $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T$. One can now sample from this covariance, by $\tilde{y}^{(n)} = \mu_y + \mathbf{L}z$, where $z \sim \mathcal{N}(0, 1)$ [225].

5.1.3 Calibration

Device Characterisation

Various parts of the reconstruction process described in Section 5.1.2 require the characterisation of the experimental device. As has been discussed, the vector field is to be reconstructed on the spatial domain of the microlens foci spots. To also encode the spectral information, each spot was also dispersed in the ± 1 orders via the use of a diffraction grating. Here, the hyperspectral vector measurement was evaluated at each microlens spot individually. Thus, to evaluate the vector field in each spatial position, it was necessary to extract the microlens spot in each of the three diffraction orders.

The peaks in the fundamental order were identified with a peak finder, and the streaks in the ± 1 orders were then collected using a calibrated dispersion curve of the diffraction grating. In order to perform this calibration, the microlens array was replaced with a pinhole, which produced a single spot in each of the three diffraction orders. Light from an ultra-short oscillator (Rainbow Femtosecond) was passed through a tunable triple grating monochromator (Optics Focus 7ISW301), and used to illuminate the RAVEN device with a single colour (1 nm bandwidth) at a time. The distance between spot's location in the ± 1 orders and the fundamental was measured for each wavelength, yielding the dispersion curve, which is shown in Fig. 5.8.

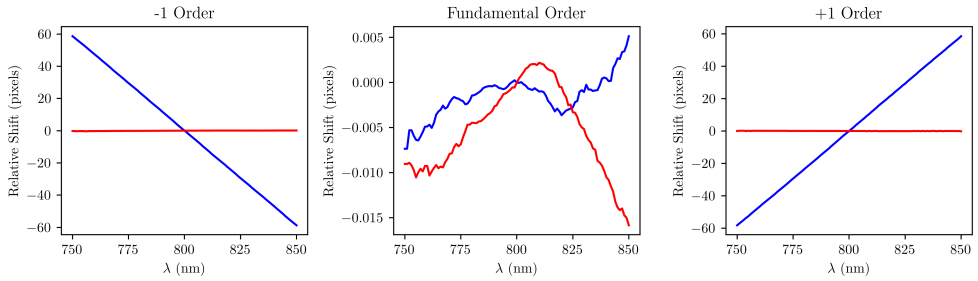


Fig. 5.8 Displaying the measured dispersion curves, relative to the central wavelength ($\lambda_0 = 800$ nm). A pinhole was illuminated by approximately monochromatic light and the shift of the spot was measured for each wavelength. One sees stability in the fundamental order with a pixel shift of far less than a pixel across the wavelengths, and the dispersion curve is symmetric across the $+1$ and -1 diffraction orders. The distance between the spot in the fundamental and the diffraction orders for 800 nm was ± 867 pixels for the ± 1 orders respectively.

A further quantity that it was necessary to characterise was the devices' spatiospectral intensity response, across the full range of the sensor. This is then used in the virtual diagnostic, presented in Section 5.1.3, to make synthetic measurements that are as close to reality as possible. This was achieved similarly to the dispersion curve. Again, the microlens array was removed, and this time replaced with a slit, that allowed the 0 and ± 1 orders to be illuminated with a small gap between them. Again using the monochromator, an image was recorded for each wavelength resulting in a cube. The cube was smoothed spatially by Gaussian blur to remove sharp features, before the fundamental was shifted to overlay the ± 1 orders using the calibrated dispersion curve. Finally, the intensity in the ± 1 order was divided element-wise by the intensity in the fundamental to give the device's spatial-spectral response. This process was done on a polarisation-pixel basis, in order to obtain this spatial-spectral response for each polarisation channel (due to the possible dependence of the grating), and is shown for one polarisation channel in Fig. 5.9. Furthermore, the polarisation dependence of the grating was also tested by the comparison of an image of the fundamental order and an image when the grating was removed. It was found that the change in polarisation content was smaller than 1%.

The final object to be captured was a measurement cube using the microlens array where the monochromator was used to capture each wavelength slice of the cube. This

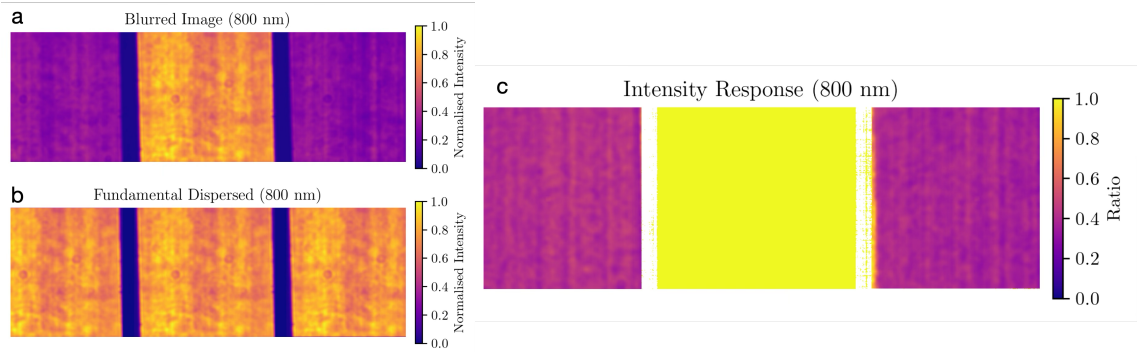


Fig. 5.9 Displaying the characterisation of the devices' spatiospectral intensity response. Monochromatic light was used to illuminate a slit (meaning the fundamental and the ±1 orders didn't overlap for any wavelength) which was imaged by the device. The image for each wavelength was first blurred with a Gaussian kernel to remove sharp features, shown for 800 nm in a). Then the fundamental order was extracted and the calibrated dispersion curves were used to shift it over the ±1 orders, shown in b). The initial blurred image was then divided by this dispersed fundamental order (i.e., a ÷ b), in order to give the grating intensity response, in c). This process was repeated for each wavelength and done individually for each polarisation channel.

cube was to be used in the virtual diagnostic, as well as a measurement to test the neural network.

Virtual Diagnostic

In order to train the neural network it was necessary to generate synthetic data, to be able to create a high amount of variability in the training data, thus allowing the network to generalise. For the network to make accurate predictions on real data, it is essential to close the "reality gap" – the difference between the simulated data used for training and real data. Due to the fact that the neural network worked on each spatial location individually, it is possible to simply generate training data consisting of single spatial locations (triplets). However, here the full measurement process is simulated, in order to try and achieve more reliable consistency with reality.

The process to simulate a hypercube occurred as follows:

1. The fundamental order of the measured cube was taken, the spots were isolated and normalised spectrally.

2. For each microlens spot, a hyperspectral amplitude and phase derivatives were generated for the x and y polarisations, as well as a spectral relative phase delay between the polarisations (corresponding to handedness). These generated values were chosen to be very diverse. These were applied to each spot by simple multiplication with amplitude and shift according to the derivative.

3. The effects of the birefringent medium and polarisation filter array were simulated, according to the formulas in Appendix C. These were also applied to the generated hyperspectral amplitudes and polarisation-relative phase differences to yield the hyperspectral intensities in each measurement orientation (0, 45°, 90°, 135°). This also included the leakage of energy between the polarisation channels, according to the manufacturer's specifications (an extinction ratio ≈ 100 [232]).

4. The cube was integrated spectrally.

5. To ensure the simulation process was not biased to certain polarisation orientations, the final image was randomly rolled in each spatial direction.

Training Data Generation

This process generated clean synthetic measurements. Some augmentation steps were also performed, again to match the data to the real experimental data. Firstly, Poisson noise was simulated. For each data sample, the signal was scaled (i.e., the effective photon count was adjusted) to produce varying noise significance, before being renormalised. The other distortion to be simulated was overexposure. As mentioned, the signal values in the fundamental were significantly larger than those in the ± 1 orders. In order to maintain sufficient levels of signal in the diffracted orders, sometimes the fundamental became overexposed. Here, this was simulated by thresholding the signal at a randomly chosen value. For both these augmentation steps the clean original signal was also stored, in order to make training pairs for the preprocessing network.

It was tested that the simulation process could generate images that resembled real measurements. Once synthetic measurements had been generated, the same peak extraction process was used as that which is used on real measurements, guaranteeing consistency, before being fed to train the network. In total, 1000 cubes were generated with approximately 250 encoder pixels each. The network was trained for 150 epochs on an Nvidia A100 GPU, using the Adam optimiser, which is when it showed qualitative convergence.

Reference & Implementation

During an experiment, the beam is transported from the area of focus in an experiment to the RAVEN diagnostic using an additional optical setup which may cause a new set of aberrations, which do not exist in the actual experimental focus. In order to remove the effects of these components, a reference must be taken. Generally this is achieved by adding a pinhole at the focus of the laser to spatially filter the pulse and to provide a flat wavefront before it passes through the additional optical setup. The wavefront that RAVEN measures in this case is subtracted from subsequent measurements. If one is simply interested in the relative change of STCs, for example in some dynamic event, then a reference is not necessary.

A python library was designed to handle the entire measurement process. Once a reference measurement is taken, the spot locations of the fundamental order are extracted and a virtual microlens array is fitted using a spatial transformer network [154]. This virtual microlens array contains equal spacing between all spots according to the parameters of RAVEN, and is used as the grid on which to perform zonal phase stitching. For subsequent measurements, the spot displacements are calculated with respect to the reference measurement spots, before being put onto the virtual grid.

5.2 Measurements

The following experiments were performed at the Center for Advanced Laser Applications (CALA) in Munich, on the ATLAS-3000 petawatt class laser. This facility contains a number of chambers where experiments can be performed with interests ranging from ion acceleration, basic physics and electron acceleration. In the following, the device was installed in the ETTF chamber (as seen in the pictures of Appendix D), which is usually used for electron acceleration.

5.2.1 RAVEN Installation

For usual experiments, the beam is focused into a gas jet using a spherical mirror with a focal length of f_1 =10 m, in order to perform laser wakefield acceleration. To perform measurements with RAVEN, the gas jet was not used and the beam was allowed to re-collimate after the focus using a lens, before being demagnified onto RAVEN's sensor. This demagnification was achieved by first using an existing set up with an achromatic doublet lens f_2 =0.5 m, which was used to image the beam onto an installed FALCON diagnostic. RAVEN requires a slightly smaller input beam size, which was achieved with a beam splitter followed by a 3:5 re-imaging and de-magnifying telescope (achromatic lens and a planar achromatic microscope lens). Having RAVEN and FALCON installed simultaneously was useful for comparison in verification measurements, as is discussed subsequently. During all measurements, the beam was attenuated prior to compression using a reflective attenuator.

The spectral phase was measured in approximately the spatial centre of the beam using self referenced spectral interferometry (a WIZZLER device), yielding a temporal pulse duration of 29.6 fs. To optimise the experimental workflow, these measurements were conducted immediately following the RAVEN measurements, rather than simultaneously. This approach was justified by the observed high shot-to-shot stability of the spectral phase. Consequently, the intensity measurements presented in the following sections incorporate

the spectral phase data from a shot taken in close temporal proximity. It is worth noting that this sequential measurement strategy was adopted to efficiently utilise available beam-time and does not reflect any inherent limitation of the RAVEN technique. Future implementations could readily incorporate simultaneous spectral phase measurements if deemed necessary for specific experimental objectives.

As has been discussed, a referencing step is included (where a pinhole is placed in the focus), which allows the removal of any linear aberrations caused by the part of the setup after the focus (i.e. the re-collimation lens). It should be noted that this pinhole-referencing technique does not work for nonlinear effects. In the case that the beam is too intense, the propagation through the lens will cause an intensity-dependent aberration, that will cause clear differences between the pinhole and standard measurements.

There are several optical elements installed after the focus (collimating lens, a beam splitter for another diagnostic setup, windows etc.) which together amount to a few centimeters of fused silica. In the experiment, the beam is attenuated using a reflective attenuator and a series of filters. Together, these elements provide a final attenuation by $10^9\times$, and therefore the transmitted beam energy is on the order of tens of nJ, resulting in an initial intensity of 10^6 W/cm^2 at the first lens, which is well-below the threshold for non-linear effects to take place. A simple calculation of the B-integral confirms this:

$$B = \frac{2\pi}{\lambda}\int n_2^{(\text{silica})} I(z)dz \approx \frac{2\pi}{800\text{ nm}} \cdot 2.4\times10^{-16}\text{ cm}^2/\text{W}\cdot 10^6\text{ W/cm}^2\cdot 0.05\text{ m} \approx 10^{-4}, \tag{5.17}$$

where the value of n_2 for silica at $\lambda = 800$ nm is taken from A. J. Taylor *et al* [233]. Thus, this nonlinear effect will cause a very small effect on the measured wavefront, and it is deemed negligible.

5.2.2 Comparison with FALCON

With any new diagnostic, it is important to compare it to existing, proven techniques. To perform such an experiment here, simultaneous measurements of the laser pulse were taken with RAVEN and the FALCON technique [132], using a beam splitter. This latter method involves the use of a Shack Hartmann sensor and a motorised filter wheel. By taking eight shots of the laser, each with a different narrowband filter spaced between 770 nm and 840 nm, multiple Shack Hartmann patterns are obtained across the spectral range of the laser pulse. The wavefront gradients are extracted, before modal spatio-spectral wavefront retrieval is used to reconstruct the wavefront in terms of spectrally dependent Zernike modes, as described in Section 2.4.6. RAVEN is superior to FALCON in that it can obtain the laser pulse in a single shot, as well as including polarisation information, yielding the full vector field. For this comparison test however, the laser used was ATLAS 3000, which is linearly polarised. FALCON has been verified extensively [132], making it a suitable benchmark for RAVEN. This experiment was carried out at a lower pump energy (15 J before compression) than full capacity. Once the beam line was

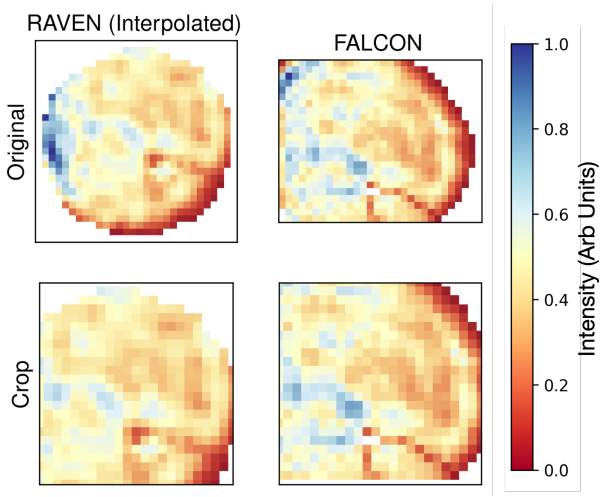


Fig. 5.10 The intensity of the field at 800 nm, captured with both RAVEN and FALCON.

aligned, a reference was taken for both methods and the measurements were taken. In the case of FALCON, a separate measurement was taken for each narrowband filter (for both the reference and the main measurement). As RAVEN is a single-shot method and

measurements were being taken simultaneously with FALCON, RAVEN performed eight full laser characterisations for each complete FALCON characterisation (which required laser shots for each of FALCON's eight bandpass filters), and the final result was the average of these measurements, increasing the signal-to-noise ratio. The intensity and wavefront were extracted from each method. The presence of a pick-off mirror (used for the measurement of the spectral phase) was used to spatially align the measurements for each method, ensuring that the same part of the beam was being analysed in each. This is shown in Fig. 5.10.

Then, in this region of interest, the wavefronts were compared as shown in Fig. 5.11, demonstrating general qualitative agreement. This is backed up by the application of the spatio-spectral wavefront retrieval algorithm, allowing the extraction of the coefficients for the wavelength-dependent Zernike coefficients. These are also shown in Fig. 5.11c, with both methods identifying pulse front tilt as the dominant mode.

To quantify the differences between the spatiotemporal wavefronts, the RAVEN phase map was linearly interpolated to match the size of the FALCON map using `scipy.interp2d`, and the root-mean-square error was calculated, in units of waves, by,

$$\mathrm{RMS} = \sqrt{\frac{1}{2\pi N_x N_y N_\lambda} \sum_i^{N_x} \sum_j^{N_y} \sum_k^{N_\lambda} \left[\phi_{RAVEN}(i,j,k) - \phi_{FALCON}(i,j,k)\right]^2}, \tag{5.18}$$

yielding a final result of 0.091. This difference is also seen qualitatively between the measurements of the wavefront performed by FALCON and RAVEN. One characteristic of this is the greater smoothness in the measurement of RAVEN, which may be due to the averaging over multiple measurements which reduces the effects of shot-to-shot fluctuations. Any remaining differences may be attributed to the fact that these are two separate devices in two separate arms, and thus, may be imaging slightly differing planes.

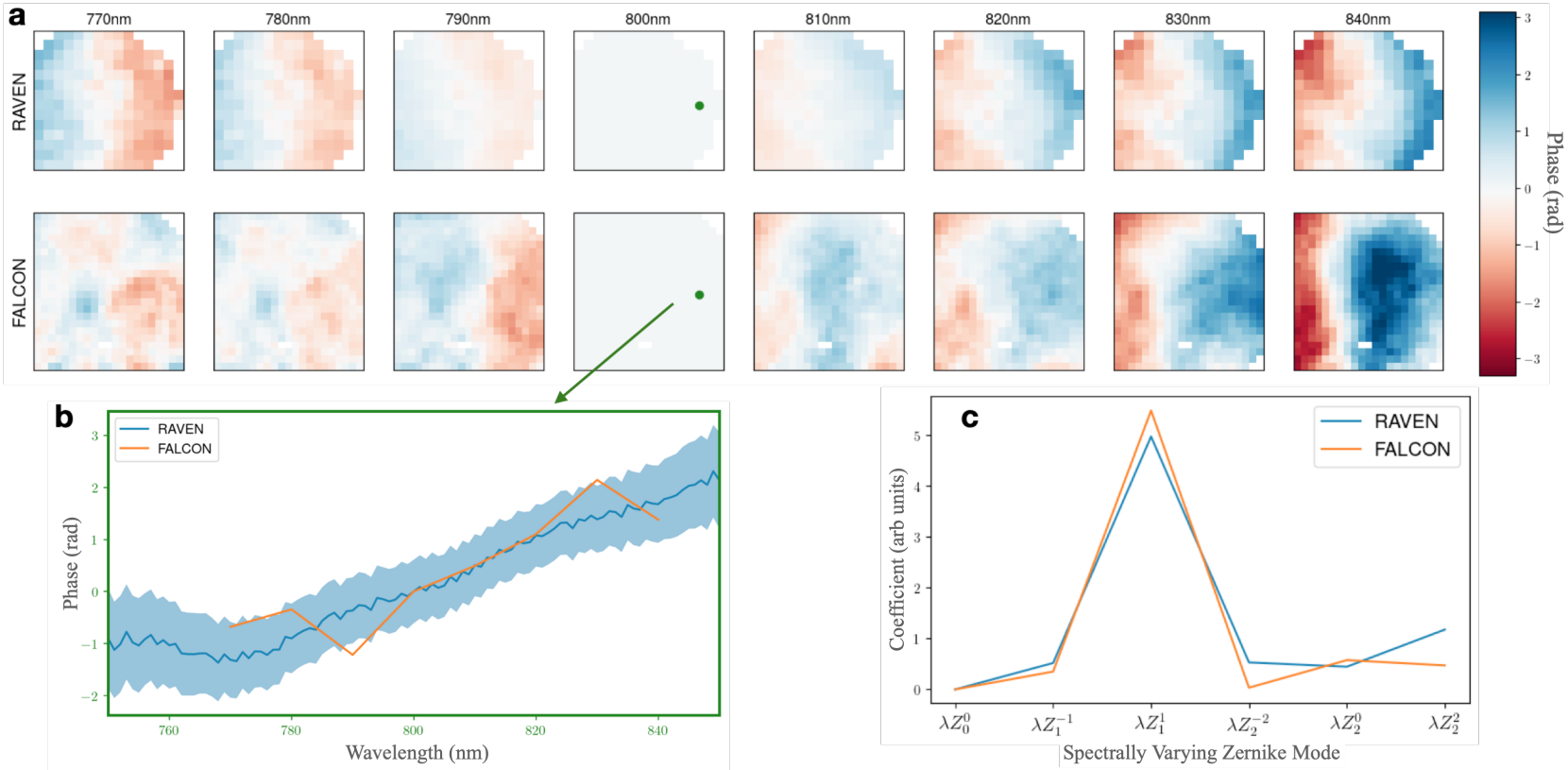


Fig. 5.11 The spatiospectral wavefront of the field, captured with both RAVEN and FALCON. **a.** The wavefront displayed at each of the wavelengths that FALCON measures at. **b.** The spectrally varying wavefront at a single spatial location, along with the measured uncertainty estimate for the RAVEN technique. **c.** The decomposition of the spatiospectral wavefront into the spectrally-varying Zernike basis.

5.2.3 Single-shot spatiotemporal Measurement of Ultra-Intense Laser Pulses

Upon the verification of RAVEN using the FALCON technique, the laser's power was increased to its maximal operational capacity for the day (35 Joules due to the presence of hot spots), and an attempt was made to characterise the individual pulses. From the initial measurements, it was found that the beam had significant residual pulse front tilt. Extracting the coefficient from the spatiotemporal wavefront yielded values of $\alpha_x = -0.061\,\mu\text{rad/nm}$ and $\alpha_y = -0.023\,\mu\text{rad/nm}$ for the pulse front tip and tilt respectively, which are of a similar order as previous measurements performed on the attenuated form of ATLAS-3000 [132]. A plot of the pulse's structure is shown in Fig. 5.12.

According to Section 2.4.7, the reduction in peak intensity due to pulse front tilt is found by,

$$\chi_{\text{intensity}} = \frac{1}{1 + \left(\frac{\alpha w_\lambda D}{0.84\lambda_0}\right)^2}. \tag{5.19}$$

By fitting a Gaussian to the spectra to find the spectral waist ($w_\lambda \approx 30\text{nm}$), and inserting numbers from the experiment ($\alpha_x = 0.061\text{urad/nm}$, $\alpha_y = 0.023\text{urad/nm}$, $D = 27\text{cm}$, $\lambda_0 = 800\text{nm}$), one arrives at a peak intensity reduction of 35% and 8% due to the pulse front tilt in the x and y directions respectively.

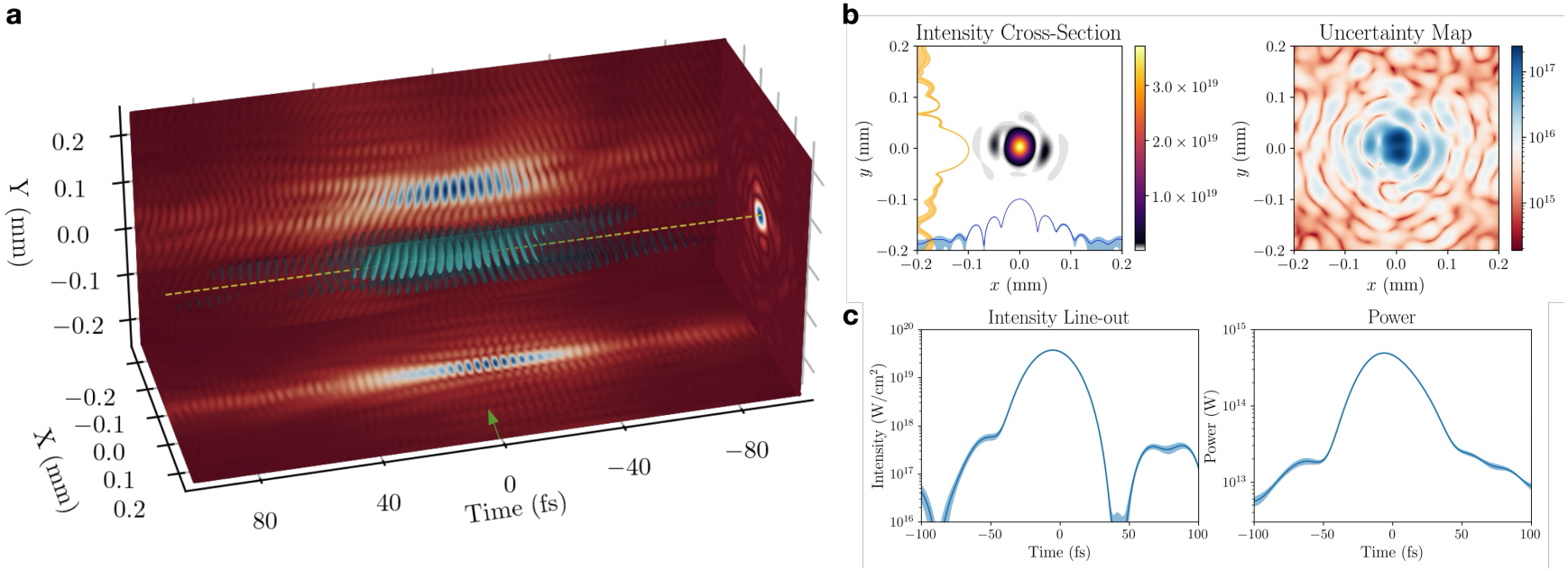


Fig. 5.12 **A single-shot measurement of the spatiotemporal field of a petawatt-class laser pulse, showing residual pulse front tilt.** **a**, The electric field of an ATLAS-3000 laser pulse at focus. The visualisation shows the iso-surface of the field, where the carrier frequency has been reduced by factor of 2 for clarity. **b**, A temporal slice of the intensity at the centre of the beam (marked by the green arrow), along with its uncertainty estimate. Also included in the intensity slice are spatial line-outs through the centre of the cross-section, shown in log-scale. **c**, The intensity line-out through the centre of the beam, shown by the yellow line, and the power found by integrating the intensity over space in the transverse dimensions.

Grating Realignment

In order to mitigate this loss of intensity, the information provided by RAVEN was then used to remove the PFT from the beam, by the realignment of the final grating in the compressor. Its angle of rotation was adjusted incrementally, while RAVEN was used to monitor the PFT in both directions. The results are displayed in Fig. 5.13. The quoted change in the grating angle is found by multiplying the number of steps of adjustment by the motor's quoted angular step size. However, it is clear from the plot that there were a number of irregularities in the motor's performance, including significant hysteresis, the coupling of the x and y axis (potentially due to relative rotation between the grating and RAVEN), and inconsistent changes of PFT for set number of adjustment steps. This highlights a significant benefit of having a diagnostic to monitor the PFT directly.

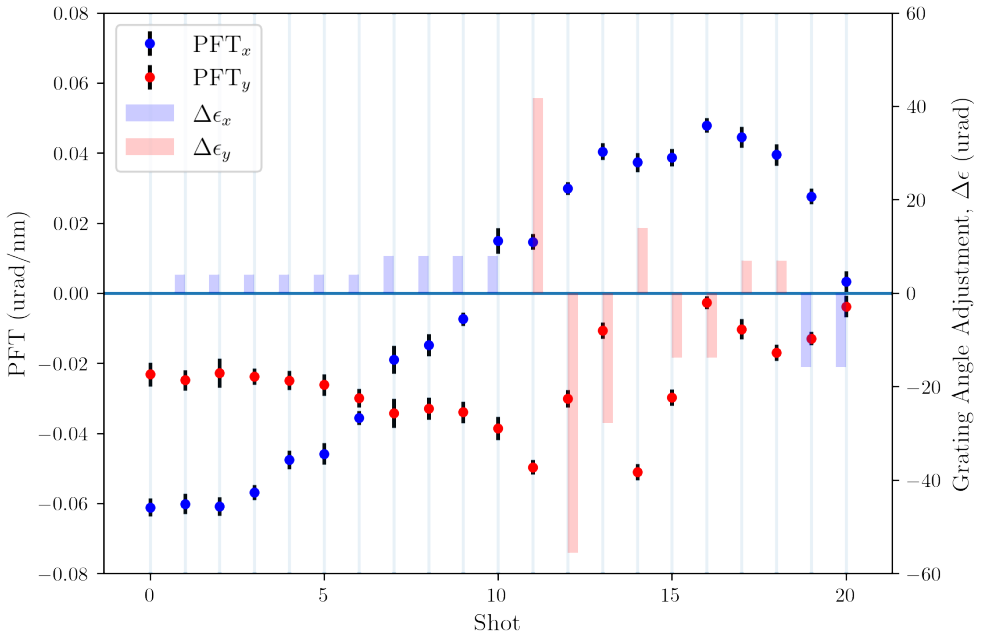


Fig. 5.13 Showing the process of realigning the grating. For each shot, RAVEN measured the PFT and then the angle of the final grating in the compressor was changed with the aim of minimizing the measured PFT. Each point shows the measurement of PFT, where blue and red are the $\hat{x}$ and $\hat{y}$ directions respectively. The bars then show the incremental change in grating angle (on the scale on the right).

The accumulated amount of angle adjustment to remove the pulse front tilt in both directions was $\epsilon_x = 23.6\,\mu\text{rad}$ and $\epsilon_y = 41.6\,\mu\text{rad}$. The corresponding angular chirp, α_x

can be calculated as,

$$\alpha_x = 2N\epsilon_x \frac{\tan(\beta)}{\cos(\nu)}, \quad \beta = \arcsin\left(\sin(\nu) - N\lambda_0\right) \tag{5.20}$$

where N is the grating constant (1480 mm^{-1}), ν is the angle of incidence (50°), and β is the grating angle [134]. Thus, the correction to the angular chirp was, $\alpha_x = 0.049\,\mu\text{rad/nm}$ and $\alpha_y = 0.089\,\mu\text{rad/nm}$. One sees these values are in the right order of magnitude to what was measured by RAVEN, and the difference is due to the aforementioned uncertainties (hysteresis) in the motor's performance. One notes that the real-time PFT measurements that are provided by RAVEN could be used as feedback for a regular automated alignment process.

The resulting spatio-spectral field at focus, after realignment, is shown in Fig. 5.14a, while a cut at $t = 0$ is shown in Fig. 5.14b. Note that the relative uncertainty at the peak intensity is two orders of magnitude lower than the signal level. Fig. 5.14c shows the temporal intensity of the pulse, both at the focus position, $(x = 0, y = 0)$, and integrated over space. The pulse duration of the former was found to be 29.8 ± 0.2 fs, in accordance with the previously quoted value measured by the WIZZLER. The spatiotemporal Strehl ratio of the laser was calculated, which is the reduction of the peak intensity due to the spatiotemporal wavefront (Section 2.4.4). The absolute ratio was 0.81 and the STC-isolated ratio (taken by removing the spectrally averaged wavefront from the pulse) was 0.93. These values are similar to those quoted at other Petawatt facilities [115]. The quoted intensity and power measurements in Fig. 5.14 also incorporate the combined efficiency of the compressor and beam transport system of ∼70%.

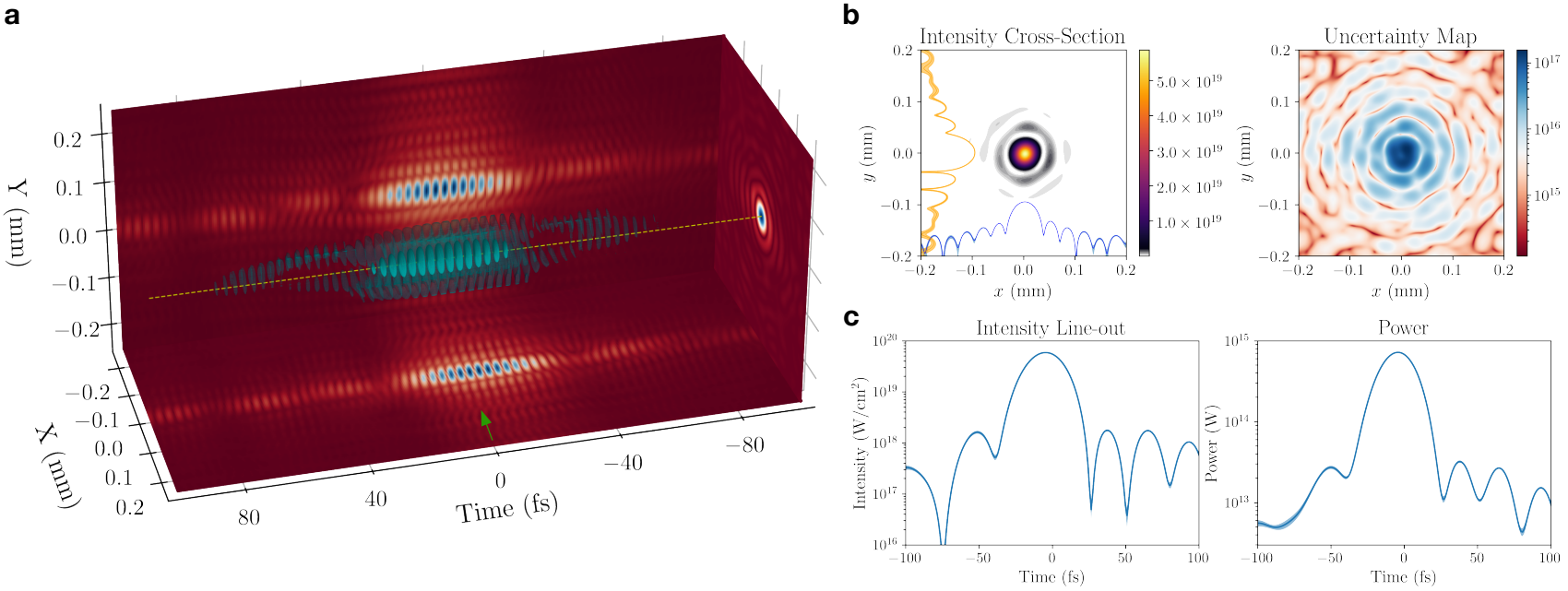


Fig. 5.14 **Single-shot measurement of the spatiotemporal field of a petawatt-class laser pulse, after grating realignment.** **a**, The electric field of an ATLAS-3000 laser pulse at focus. The visualisation shows the iso-surface of the field, where the carrier frequency has been reduced by factor of 2 for clarity. **b**, A temporal slice of the intensity at the centre of the beam (marked by the green arrow), along with its uncertainty estimate. Also included in the intensity slice are spatial line-outs through the centre of the cross-section, shown in log-scale. There is a clear increase in peak intensity when compared to Fig. 5.12. **c**, The intensity line-out through the centre of the beam, shown by the yellow line, and the power found by integrating the intensity over space in the transverse dimensions.

Dynamics

After showing that RAVEN was able to resolve individual laser pulses, an experiment was performed to monitor, in real-time, the STC content of the laser. Here, the ATLAS-3000 laser was operating at 1 Hz repetition rate, whilst the energy of the pump energy was changed approximately every 20 minutes by changing the pump laser configuration. For each laser shot, RAVEN was used to acquire the pulse's structure. Then both the STC content and the STC-isolated spatiotemporal Strehl ratio were calculated, allowing the long term dynamics of these changes to be studied. For these measurements, a reference was not taken, which means the Strehl ratio includes the STCs induced by the optical system used to transport the pulse to RAVEN, and is thus reduced compared to the true value. However, the dynamics and relative change of STC coefficients can still be studied. Fig. 5.15 displays the evolution of the spatiotemporal Strehl ratio, along with the two STCs with the highest Pearson's correlation coefficient with this ratio – pulse front tilt in the horizontal direction (PFT_x) and pulse front curvature (PFC). It is apparent that the hyperspectral Strehl ratio generally improves with pump energy; intuitively this is due to the fact that the compressor is aligned at one pump energy for which PFT is minimised with an inverted field autocorrelator [134]. Of all the STCs, it was found that pulse front tilt was the most strongly correlated to the peak intensity ($\rho = 0.92$).

Also highly correlated was pulse front curvature, which displayed the property of stabilising over relatively long time periods. This can be seen especially in the case when all the pump lasers were turned on, after they had all been off ($T = 20$ minutes). An explanatory candidate for this effect is the inhomogeneous heating of the gain medium as the pump energy was adjusted, and the subsequent change in thermal lensing [234]. This lensing is in itself chromatic and will thus result in different pulse front curvature. To my knowledge, these subtle dynamics have not been previously observed, even though they have non-neglible influence on the focused intensity, characterised by the spatiotemporal Strehl ratio. At $T = 98$ minutes, one Gaia pump laser was switched off, and another from

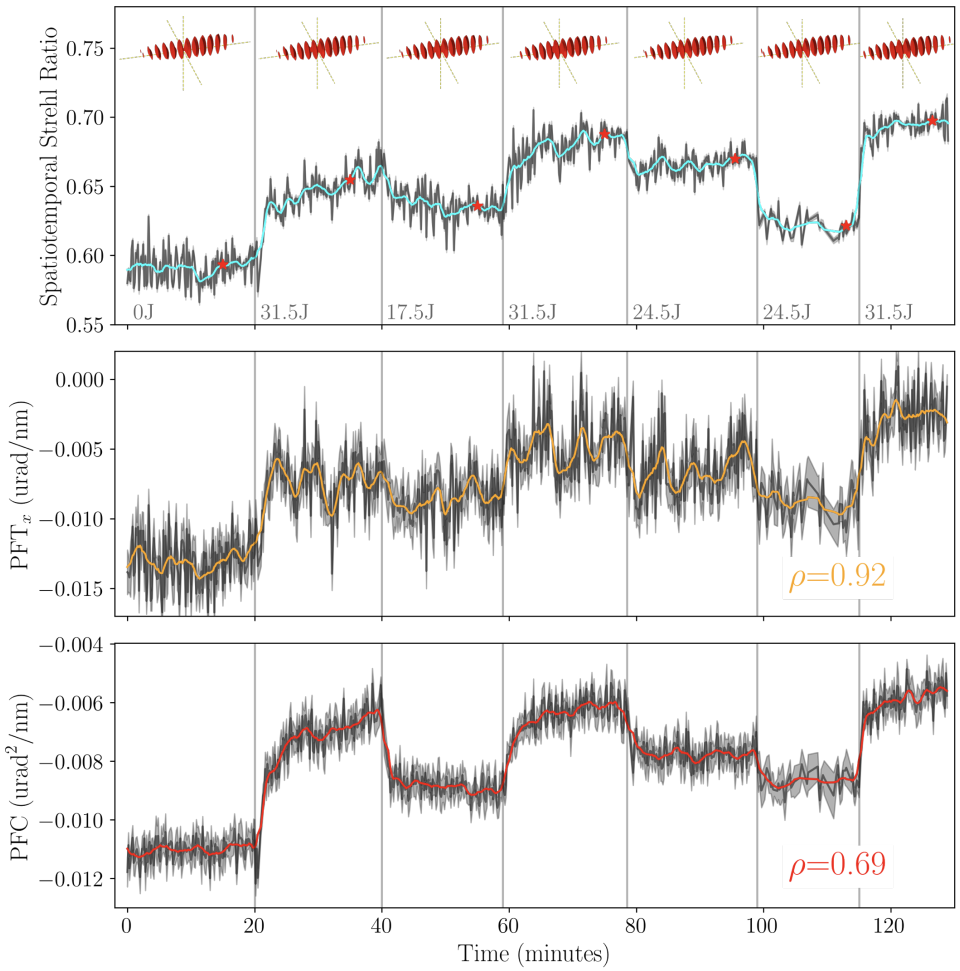


Fig. 5.15 **The evolution of the spatiotemporal Strehl ratio and STCs, with changes in the laser pump energy.** At each vertical grey line, the pump energy was changed, as denoted at the bottom of the spatiotemporal Strehl ratio subplot. Each red star denotes the temporal location where the field is shown in the 3D plot above. This measurement was not referenced, and thus, the spatiotemporal Strehl ratio includes contributions from STCs of the optical set up used to transport the pulse to RAVEN.

a second bank was switched on. As this separate bank had only one active Gaia, it is likely that it had not experienced as much heating, and therefore the total thermal lensing was smaller (and thus it was overly compensated).

5.2.4 Vector Field characterisation of Optical Vortices

To demonstrate the full vectorial capability of RAVEN, an experiment was performed to measure a circularly polarised optical vortex. The optical vortex is a type of pulse that

has the intriguing property of possessing orbital angular momentum (OAM), manifested by a spiral wavefront, which leads to the formation of the characteristic 'donut' focus upon propagation. These attributes make it of high interest in many experiments in the high-intensity regime [235, 54, 236]. Existing methods to characterise such fields have typically been developed on high repetition-rate systems, requiring a scan with many laser shots [70, 71].

The vortex retarder, also known as a spiral wave-plate, is an optical element designed to produce a beam with non-zero OAM, which is seen in the form of a spiral wavefront,

$$\phi_m(\theta) = m\theta, \tag{5.21}$$

where $\theta = \cos^{-1}(y/x)$ and m describes the orbital angular momentum. This wavefront is realised by use of a Liquid Crystal (LC) polymer, with its orientation (referred to as the fast-axis) rotating continuously over the optic. At each spatial position, the LC causes a delay to each polarisation component of the pulse, depending on the relative orientation of the polarisation and the fast axis. The fast-axis is rotated across the optic so that a delay depending on the azimuthal coordinate is applied, resulting in a $m2\pi$ phase shift around the full circle. Where the s-polarisation obtains a long delay, the p-state will obtain a short delay and vice versa. This leads to a relative phase shift in the vortices between different polarisations. To see this mathematically, consider the Jones matrix description of a vortex retarder [237, 238]:

$$\mathbf{M}_s(\phi) = \begin{bmatrix} \cos(m\phi) & \sin(m\phi) \\ \sin(m\phi) & -\cos(m\phi) \end{bmatrix} \tag{5.22}$$

When acting on an incoming circularly polarised beam, a vortex wavefront can be factorised out by means of Euler's formula,

$$\vec{E}_{out} = M_s \begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} \cos(m\phi) + i\sin(m\phi) \\ \sin(m\phi) - i\cos(m\phi) \end{bmatrix} = e^{im\phi} \begin{bmatrix} 1 \\ -i \end{bmatrix}, \tag{5.23}$$

where the original handedness of the circularly polarised light is also flipped.

Due to beam-time constraints and the cost of large aperture optics, this experiment was performed on an ultrafast oscillator. However, it should be emphasised that there is no reason why this experiment would not work in the same way on a PW class laser. The oscillator used was the Femtolasers Rainbow, which outputs a linearly polarised beam, with pulses of $<$ 7fs duration, and $\sim$1nJ. This shorter duration, relative to ATLAS, means a larger bandwidth, and so the pulses were passed through a set of low and high bandpass filters to create a more confined spectral range (of 100 nm centered around 800 nm), better approximating the ATLAS-3000 facility. This was done primarily because the specific prototype of RAVEN was designed for this bandwidth, and a higher bandwidth will cause the streaks to begin to overlap.

To create the initial circularly polarised beam, a quarter-wave plate (Thorlabs: WPQ10E-780) was added to the beam path, with its fast axis at a 45° angle to the polarisation orientation of the incoming pulse. This led to all four polarisation filter channels receiving the same intensity. This circularly polarised pulse then passed through a $m = 2$ vortex retarder (Thorlabs: WPV10-780). The full apparatus is shown in Appendix D: Fig. D.4 and Fig. D.5.

In order to resolve the handedness of the resulting field, according to the previous arguments in Section 5.1.1, a half-wave plate (Thorlabs: WPH10ME-808) was added before the microlens array of the camera. This object added the spectrally varying delay between polarisations, allowing the for the disambiguation of the two possible solutions. Its fast-axis was aligned with the 0° filter in the polarisation filter array, thus creating a π

phase difference between the 0° and 90° components at 808 nm. Of course the quarter-wave plate was already chromatic, but here the additional half-wave plate was used to be able to actually measure the spectral variation and quality of the circular polarisation added by the quarter-wave plate. In general, it is better to use a birefringent medium that creates a larger spectral variation in polarisation delay – for example, a multi-order half-wave plate. However, confronted with time limitations, here a zero-order half-wave plate was used, and its success in removing the handedness ambiguity is demonstrated.

Two single-shot measurements were taken: a reference (by removing the vortex retarder), and one with the vortex retarder. Predictions of the intensities and wavefront derivatives were made using the neural network. From there, the Stokes parameters were extracted from the intensity measurement, and the wavefront of the 0° polarisation component was found. These two quantities are then combined to form the vector field. No spectral phase measurement was taken, so subsequent results assume a flat spectral phase when averaged over the spatial extent of the beam.

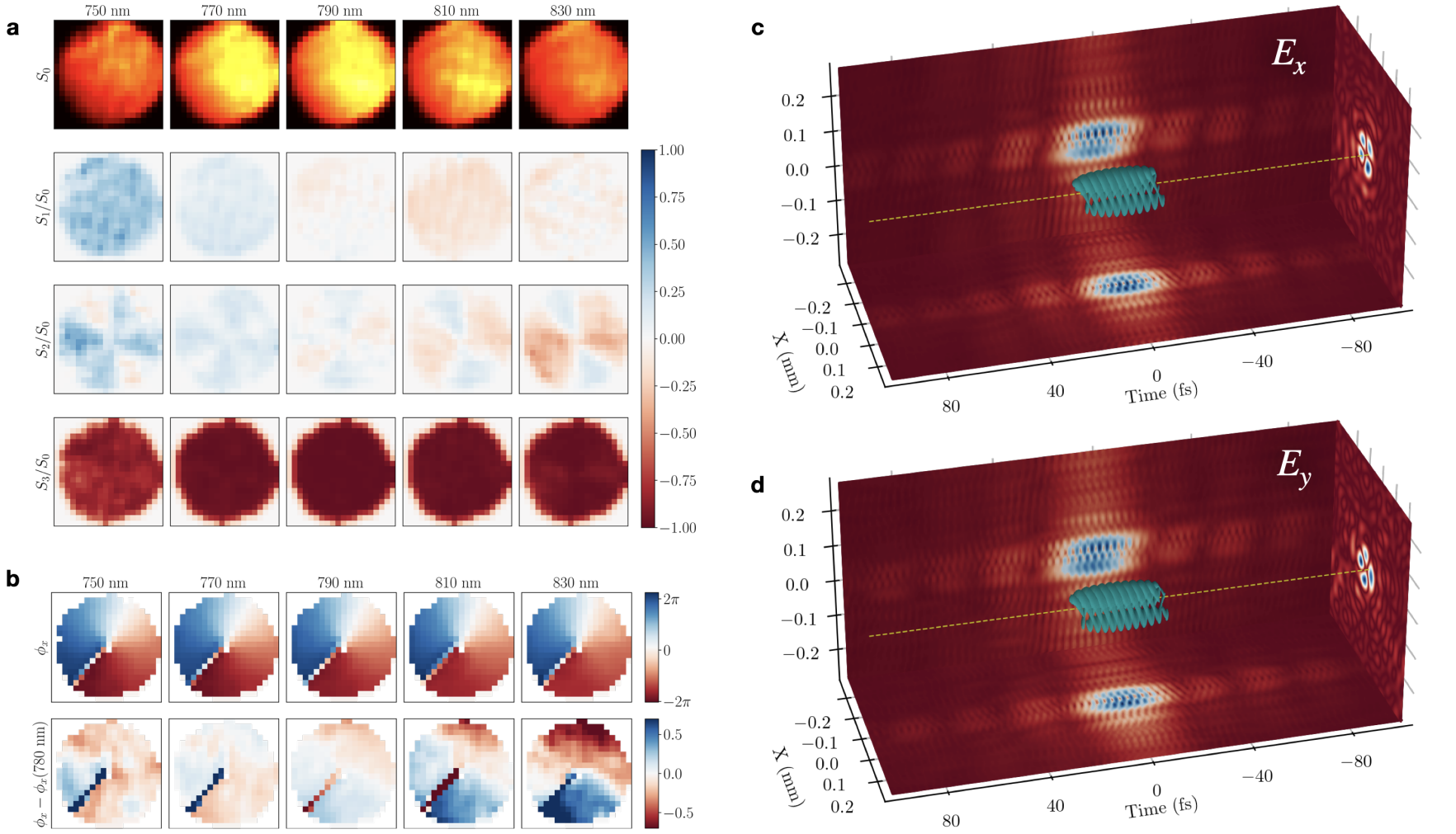


Fig. 5.16 **Single-shot measurement of the spatiotemporal vector field of an optical vortex.** **a**, The spatio-spectral Stokes parameters of the vectorial near field. **b**, The spiral wavefront of the optical vortex. Also shown is the relative change from the design wavelength of the vortex retarder (780 nm), displaying the chromatic properties of this object. **c,d**, The field at focus, for both E_x and E_y. The carrier frequency has been reduced by a factor of 2 for clarity, and the spatial units have been scaled to the ATLAS-3000 system (27 cm diameter in the near-field).

The results in Fig. 5.16 demonstrate the spectral dependence of the resultant vector field, as expected, due to the chromatic properties of the quarter-wave plate and vortex retarder used to generate it. In the case of a perfect left-handed circularly polarised beam [239], the final three normalised Stokes parameters have the values:

$$\frac{1}{S_0}\begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}.$$

Clear deviations from these values are seen as one moves away from the central wavelength of the quarter-wave plate, due to the fact that it is designed to create circularly polarised light at its design wavelength.

In order to extract the handedness of the circular polarisation, one can use the argumentation in Section 5.1.1. Firstly, the magnitude of the delay is found by,

$$\pm(\delta) = \arccos\left(\frac{I_D(\lambda) - I_A(\lambda)}{2\sqrt{I_H(\lambda)I_V(\lambda)}}\right) - \delta_{HWP}(\lambda), \tag{5.24}$$

where δ_{HWP} is the pre-calibrated delay imparted by the half wave-plate. Here, one can then exploit the prior knowledge of spectral smoothness, and choose the solution with the minimum total variation (TV). Thus, the chosen solution is;

$$\delta(\lambda) = \underset{i\in[+,-]}{\operatorname{argmin}}\left(\mathrm{TV}(\delta_i(\lambda))\right). \tag{5.25}$$

Both solutions are displayed in Fig. 5.17, with solution 2 being the one that minimises the total variation, and is chosen. Also displayed for verification, is the angle recorded for a further measurement of the QWP (without vortex or HWP), demonstrating the close likeness with the chosen solution, whilst any deviations are due to the vortex plate. Finding the magnitude of this delay was crucial for determining the sign of the fourth Stokes parameter, S_4.

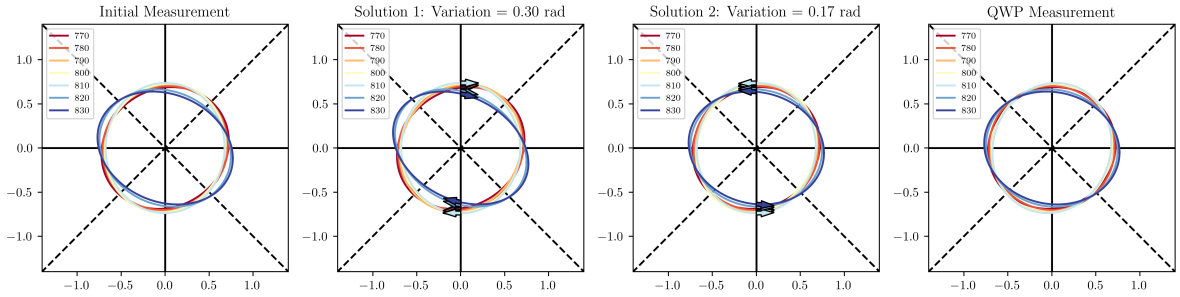


Fig. 5.17 Each plot shows the polarisation ellipses for a number of wavelengths. The far left shows the raw measurement, and the second and third plots show the two corresponding ambiguous solutions (found by subtracting the pre-calibrated delay from the birefringent medium). The solution with the minimum total variation is chosen. The final plot shows the polarisation ellipses for an additional measurement that was taken without the birefringent medium, showing qualitative agreement with the selected solution.

Extraction of the Wavefront

The perfect optical vortex has a smooth spiral wavefront in three dimensions. However, when projected onto two dimensions, there exists a discontinuity due to the wrapping of the phase. Furthermore, the vortex retarder is designed to work at a certain wavelength, and deviating away from this wavelength will create a discontinuity in the wavefront, even in three dimensions. This creates a difficulty in using the typical zonal phase stitching methods, which are solved by least squares, and smooth out the discontinuity, thereby destroying the vortex wavefront. This was, of course, part of the motivation of developing the PROD (phase reconstruction using an overcomplete dictionary) technique in Section 4.2.

In order to extract the spatiospectral vortex, the following steps were taken,

1. PROD was used to find the vortex modal coefficient and affine transform parameters for the derivatives of the 800 nm wavefront.

2. It was then used to find the coefficients for all other wavefronts, using the learnt coefficients of the 800 nm wavefront as the initialisation.

3. For each wavelength, the fitted vortex derivatives were subtracted from the wavefront derivatives, and the remainder was reconstructed using standard zonal phase stitching (including the uncertainty estimate).

4. The fitted vortex was added to this result from the zonal phase stitching.

The fitted OAM order for each wavelength is shown in Fig. 5.18, while the wavefronts are seen in Fig. 5.16(b). One notes that an uncertainty estimate could have been provided for these OAM values by Monte-Carlo sampling the measured near-field and fitting an OAM value to each sample of wavefront derivatives, before finding the variance over the samples. This calculation was not performed here due to computational and temporal restrictions.

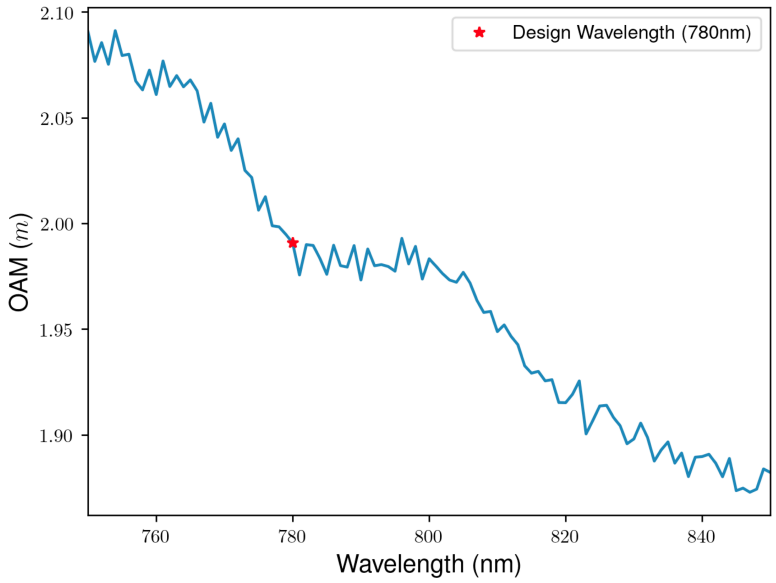


Fig. 5.18 The fitted OAM coefficient for each wavelength by using the PROD technique. This clearly demonstrates the chromaticity of the vortex retarder.

5.3 Discussion

This chapter has introduced RAVEN, a technique capable of single-shot measurements of the spatiotemporal vector field of ultra-intense laser pulses. In a paradigm shift from standard laser characterisation devices, the development of RAVEN began with systematic

considerations of the quantity of interest, the far-field, which informed every subsequent design choice. The result is a method of superior capability and trustworthiness, providing a number of crucial benefits over previous work. Where most existing methods require hundreds of shots, RAVEN achieves a complete spatiotemporal characterisation of a laser pulse in just one, including full polarisation information, resulting in significant time savings. Thanks to its fast reconstruction time, the method allows for online measurement and optimisation at ultra-intense laser facilities with repetition rates $\sim 1\,\mathrm{Hz}$. This was demonstrated in its application to the ATLAS-3000 system, in the first single-shot spatiotemporal vector field characterisation of a PW-power laser pulse. Importantly, RAVEN facilitates the detection of phenomena that were previously inaccessible. This includes the shot-to-shot fluctuations of the spatiotemporal wavefront of the laser pulse, but also with its complete treatment of the polarisation state of the pulse, vectorial ultra-intense laser pulses can now be studied in real time.

For future development of diagnostics based on the principles of RAVEN, there are a number of improvements that one could make, which largely centre around the grating. The most obvious issue with this current iteration is the grating's diffraction efficiencies; the grating had approximately equal efficiency in the fundamental and ± 1 orders ($\sim 28\%$ in the fundamental and $\sim 25\%$ in the ± 1 orders), but the dispersion in the ± 1 orders caused the intensity values to be much lower than the fundamental. This was worked around by overexposing the fundamental in order to get sufficient signal in the other orders, and then using a preprocessing network to recover the original fundamental order, but it would be much more simple and trustworthy to instead use a custom grating that could send the desired signal levels to each order.

Another desirable property of the grating would be that it also created copies in the vertical direction (and therefore with dispersion in the vertical direction). With the version of RAVEN presented in this chapter, ± 1 diffraction orders are created along the horizontal direction which are also dispersed horizontally. This led to a greater precision in the wavefront derivatives in the vertical direction, $\frac{\partial \phi}{\partial y}$, when compared to the horizontal, $\frac{\partial \phi}{\partial x}$

(this was observed in the uncertainty predictions by the variance network). Thus, by also incorporating copies with vertical dispersion, this would presumably allow for an increase in precision in the wavefront derivatives in the horizontal direction. As in the case of the quadriwave lateral shearing interferometer, this can be achieved with a bi-sinusoidal form of the grating, which must be modified to also send energy to the fundamental (with some DC amplitude modulation which makes the entire grating partially transmissive). One notes that there was also already sufficient space on the sensor to achieve this.

Chapter 6

Conclusion

The structure of ultra-intense laser pulses is essential for all applications, and therefore it is crucial to be able to characterise it. The best existing diagnostic methods can be broadly divided into two categories; those that assume the separability of the field into spatial and temporal components and measure one of them in a single shot, and those that measure the full spatiotemporal field (potentially including polarisation), but require multiple shots of the laser. The former category covers the typical methods to capture the spectral phase, such as FROG or SRSI, as well as methods that capture the spatial wavefront, like the Shack Hartmann sensor or the lateral shearing interferometer. However, due to their blindness to at least one dimension, these methods cannot detect spatiotemporal couplings, which have been shown to play a critical role in the behaviour of ultra-intense laser pulses.

The existing techniques that can resolve the spatiotemporal field include FALCON, TERMITES and INSIGHT. However, their multi-shot nature is a major limitation. For instance, any shot-to-shot fluctuations in the spatiotemporal structure of the laser pulses (which were shown to be significant in Chapter 5) are invisible to these diagnostics; this is critical in laser wakefield acceleration where reproducibility and stability are current impediments to achieving widespread use [240]. Additionally, there is growing interest in the deliberate shaping of the spatiotemporal pulse structure to produce structured light with novel properties, such as the flying focus. While the spatial and temporal

profiles of the pulse are often optimized in automated feedback loops, the currently long characterisation time of the spatiotemporal structure of the pulse is prohibitive to these ideas being extended to the spatiotemporal domain. Moreover, recent interest in tailoring the polarisation of pulses is similarly hindered, as most existing methods are not sensitive to vectorial fields.

The culmination of this thesis is RAVEN, a technique for the *Real-time Acquisition of VEctorial Near-fields*. This method is capable of measuring the complete spatiotemporal vector field of individual ultra intense laser pulses, at a resolution that is sufficient to resolve the pulse at focus. Not only does this give insight into the physics of the shot-to-shot fluctuations created in the generation of these phenomena, but it also enables the easy capture of crucial data about specific facilities to be used within simulations and to inform future experiments. The fact that it can characterize a pulse in a single-shot of a laser means that it can provide feedback for the implementation of future spatiotemporal beam shaping techniques. Furthermore, it is hoped that RAVEN's sensitivity to the polarisation structure of the pulse will also help to enable the experimental realization of vectorial structured beams in ultra-intense physics.

A RAVEN is now permanently stationed at the ATLAS-3000 laser in the Center for Advanced Laser Applications, where it monitors the daily fluctuations of the laser. Furthermore, other students PhD projects (in the group of Dr. Andreas Döpp) are now developing further techniques using RAVEN as a building block; for example in the Bayesian assessment of process noise of these lasers. Colleagues have also taken the device to other facilities to perform measurements, such as at the J-KAREN-P laser at the Kansai Photon Science Institute. Future experiments will involve deploying RAVEN at multiple locations along a beam line in order to diagnose the causes of specific spatiotemporal couplings (STCs) at the ATLAS-3000 facility. Additionally, by working with mirror leakages, experiments will also be performed to monitor the STCs in situ during laser wakefield acceleration experiments. This will allow the detection of correlations between

the STCs and the characteristics of accelerated electrons. Similar correlation studies can also be performed for other uses of ultra-intense lasers, such as ion acceleration.

With the increased interest in the spatiotemporal structure of laser pulses, it is desirable to be able to manipulate this structure freely in spatiotemporal beam shaping. To achieve this, a perhaps inelegant method would be to spectrally disperse the pulse and to send each wavelength channel onto a separate deformable mirror before recombining them. With the real-time nature of RAVEN (it can reconstruct the pulse at a rate faster than the repetition rate of the laser), it could be used as a diagnostic to achieve an adaptive optics feedback loop. A very simplified version of this was discussed in Chapter 5, where a compressor grating angle was adjusted to minimize the pulse front tilt measured by RAVEN. Generalising this beam shaping to other types of spatiotemporal couplings would provide greater flexibility in pulse shaping.

It is hoped that the other work presented in this thesis will also provide benefit to the scientific community. Chapter 3 demonstrated the potential difficulties with implementing snapshot compressive imaging with coherent laser light, due to distortions from diffraction. Naturally, this depends on the relative size of the mask pixels compared to the wavelength of the light, and so for short wavelengths this work may still provide a method for high-spatial-resolution characterisation – for example, in characterising light from high-harmonic generation, as currently being explored by Mr. Elliott Denis in the group of Professor Peter Norreys. Two general methods were introduced in Chapter 4. Firstly, CoordGate, designed for the efficient implementation of convolution and deconvolution with static (fixed optical setup) spatially varying point spread functions. The other method used an over-complete phase dictionary for wavefront reconstruction in situations where the wavefront cannot be efficiently represented in the Cartesian or Zernike bases, such as the optical vortex. This technique was later used along with RAVEN to characterize optical vortices.

References

[1] Colin N Danson, Constantin Haefner, Jake Bromage, Thomas Butcher, Jean-Christophe F Chanteloup, Enam A Chowdhury, Almantas Galvanauskas, Leonida A Gizzi, Joachim Hein, David I Hillier, et al. Petawatt and exawatt class lasers worldwide. *High Power Laser Science and Engineering*, 7:e54, 2019.

[2] Xiaoming Zeng, Kainan Zhou, Yanlei Zuo, Qihua Zhu, Jingqin Su, Xiao Wang, Xiaodong Wang, Xiaojun Huang, Xuejun Jiang, Dongbin Jiang, et al. Multi-petawatt laser facility fully based on optical parametric chirped-pulse amplification. *Optics letters*, 42(10):2014–2017, 2017.

[3] Jae Hee Sung, Hwang Woon Lee, Je Yoon Yoo, Jin Woo Yoon, Chang Won Lee, Jeong Moon Yang, Yeon Joo Son, Yong Ha Jang, Seong Ku Lee, and Chang Hee Nam. 4.2 pw, 20 fs ti: sapphire laser at 0.1 hz. *Optics letters*, 42(11):2058–2061, 2017.

[4] Christophe Radier, Olivier Chalus, Mathilde Charbonneau, Shanjuhan Thambirajah, Guillaume Deschamps, Stephane David, Julien Barbe, Eric Etter, Guillaume Matras, Sandrine Ricaud, et al. 10 pw peak power femtosecond laser pulses at eli-np. *High Power Laser Science and Engineering*, 10:e21, 2022.

[5] KWD Ledingham, P McKenna, and RP Singhal. Applications for nuclear phenomena generated by ultra-intense lasers. *Science*, 300(5622):1107–1111, 2003.

[6] Hannah Ritchie and Pablo Rosado. Electricity mix. *Our World in Data*, 2020. https://ourworldindata.org/electricity-mix.

[7] Jin Woo Yoon, Yeong Gyu Kim, Il Woo Choi, Jae Hee Sung, Hwang Woon Lee, Seong Ku Lee, and Chang Hee Nam. Realization of laser intensity over 10 23 w/cm 2. *Optica*, 8(5):630–635, 2021.

[8] Claus Fröhlich and Judith Lean. Solar radiative output and its variability: evidence and mechanisms. *The Astronomy and Astrophysics Review*, 12(4):273–320, 2004.

[9] Gérard A Mourou. Ultraintense lasers and their applications. In *Atoms, Solids, and Plasmas in Super-Intense Laser Fields*, pages 1–13. Springer, 2002.

[10] Toshiki Tajima and John M Dawson. Laser electron accelerator. *Physical review letters*, 43(4):267, 1979.

[11] François Amiranoff, S Baton, D Bernard, B Cros, D Descamps, F Dorchies, F Jacquet, Victor Malka, JR Marques, G Matthieussent, et al. Observation of laser wakefield acceleration of electrons. *Physical Review Letters*, 81(5):995, 1998.

[12] Andrea Macchi, Marco Borghesi, and Matteo Passoni. Ion acceleration by superintense laser-plasma interaction. *Reviews of Modern Physics*, 85(2):751–793, 2013.

[13] James CL Chow and Harry E Ruda. Flash radiotherapy: Innovative cancer treatment. *Encyclopedia*, 3(3):808–823, 2023.

[14] Marko von der Leyen. *Study of novel electron injection mechanisms for laser-wakefield accelerators*. PhD thesis, University of Oxford, 2021.

[15] Christian Graeff, Lennart Volz, and Marco Durante. Emerging technologies for cancer therapy using accelerated particles. *Progress in particle and nuclear physics*, 131:104046, 2023.

[16] Florian Kroll, Florian-Emanuel Brack, Constantin Bernert, Stefan Bock, Elisabeth Bodenstein, Kerstin Brüchner, Thomas E Cowan, Lennart Gaus, René Gebhardt, Uwe Helbig, et al. Tumour irradiation in mice with a laser-accelerated proton beam. *Nature Physics*, 18(3):316–322, 2022.

[17] Z-H He, B Beaurepaire, JA Nees, G Gallé, SA Scott, JR Sánchez Pérez, MG Lagally, K Krushelnick, AGR Thomas, and J Faure. Capturing structural dynamics in crystalline silicon using chirped electrons from a laser wakefield accelerator. *Scientific Reports*, 6(1):36224, 2016.

[18] Andrew Webb. *Introduction to biomedical imaging*. John Wiley & Sons, 2022.

[19] Nikolay Kardjilov, Ingo Manke, André Hilger, Markus Strobl, and John Banhart. Neutron imaging in materials science. *Materials Today*, 14(6):248–256, 2011.

[20] John E Eberhardt, Stephen Rainey, RJ Stevens, Brian D Sowerby, and JR Tickner. Fast neutron radiography scanner for the detection of contraband in air cargo containers. *Applied Radiation and Isotopes*, 63(2):179–188, 2005.

[21] BE Allman, PJ McMahon, Keith Alexander Nugent, D Paganin, David L Jacobson, Muhammad Arif, and SA Werner. Phase radiography with neutrons. *nature*, 408(6809):158–159, 2000.

[22] R Kodama, PA Norreys, Kunioki Mima, AE Dangor, RG Evans, H Fujita, Yoneyoshi Kitagawa, K Krushelnick, T Miyakoshi, N Miyanaga, et al. Fast heating of ultrahigh-density plasma as a step towards laser fusion ignition. *Nature*, 412(6849):798–802, 2001.

[23] Rl Kodama, H Shiraga, K Shigemori, Y Toyama, S Fujioka, H Azechi, H Fujita, H Habara, T Hall, Y Izawa, et al. Fast heating scalable to laser fusion ignition. *Nature*, 418(6901):933–934, 2002.

[24] F Zhang, HB Cai, WM Zhou, ZS Dai, LQ Shan, H Xu, JB Chen, FJ Ge, Q Tang, WS Zhang, et al. Enhanced energy coupling for indirect-drive fast-ignition fusion targets. *Nature Physics*, 16(7):810–814, 2020.

[25] T Yamanaka, Y Kato, Y Kitagawa, R Kodama, H Takabe, H Fujita, T Kanabe, M Nakatsuka, and K Mima. Ultra-intense lasers for fast ignition in inertial confinement fusion at ile, osaka. In *Applications of High Field and Short Wavelength Sources*, page ThE20. Optica Publishing Group, 1997.

[26] H Abu-Shawareb, R Acree, P Adams, J Adams, B Addis, R Aden, P Adrian, BB Afeyan, M Aggleton, L Aghaian, et al. Achievement of target gain larger than unity in an inertial fusion experiment. *Physical Review Letters*, 132(6):065102, 2024.

[27] H Abu-Shawareb, R Acree, P Adams, J Adams, B Addis, R Aden, 2P Adrian, BB Afeyan, M Aggleton, L Aghaian, et al. Lawson criterion for ignition exceeded in an inertial fusion experiment. *Physical review letters*, 129(7):075001, 2022.

[28] T Ditmire, M Roth, PK Patel, D Callahan, G Cheriaux, P Gibbon, D Hammond, A Hannasch, LC Jarrott, G Schaumann, et al. Focused energy, a new approach towards inertial fusion energy. *Journal of Fusion Energy*, 42(2):27, 2023.

[29] Marius Schollmeier, Sven Steinke, Christian Bild, Heike Freund, Erhard Gaul, Dan Nebe, Gaurav Raj, Hartmut Ruhl, Moritz von der Linden, and Georg Korn. Marvel fusion: Ultrafast high-intensity laser-driven fusion with nanostructured targets. In *APS Division of Plasma Physics Meeting Abstracts*, volume 2022, pages BM10–009, 2022.

[30] RW Paddock, TS Li, E Kim, JJ Lee, H Martin, RT Ruskov, S Hughes, SJ Rose, CD Murphy, RHH Scott, et al. Energy gain of wetted-foam implosions with auxiliary heating for inertial fusion studies. *Plasma Physics and Controlled Fusion*, 66(2):025005, 2023.

[31] Marco Borghesi, Gianluca Sarri, CA Cecchetti, I Kourakis, D Hoarty, RM Stevenson, S James, CD Brown, P Hobbs, J Lockyear, et al. Progress in proton radiography for diagnosis of icf-relevant plasmas. *Laser and Particle Beams*, 28(2):277–284, 2010.

[32] Hideaki Takabe and Yasuhiro Kuramitsu. Recent progress of laboratory astrophysics with intense lasers. *High Power Laser Science and Engineering*, 9:e49, 2021.

[33] Mohammad Mirzaie, Calin Ioan Hojbota, Do Yeon Kim, Vishwa Bandhu Pathak, Tae Gyu Pak, Chul Min Kim, Hwang Woon Lee, Jin Woo Yoon, Seong Ku Lee, Yong Joo Rhee, et al. All-optical nonlinear compton scattering performed with a multi-petawatt laser. *Nature Photonics*, pages 1–6, 2024.

[34] C Bula, KT McDonald, EJ Prebys, C Bamber, S Boege, T Kotseroglou, AC Melissinos, DD Meyerhofer, W Ragg, DL Burke, et al. Observation of nonlinear effects in compton scattering. *Physical Review Letters*, 76(17):3116, 1996.

[35] Julian Schwinger. On Gauge Invariance and Vacuum Polarization. *Physical Review*, 82(5):664, 1951.

[36] Ben C Platt and Roland Shack. History and principles of shack-hartmann wavefront sensing, 2001.

[37] Robert K Tyson and Benjamin West Frazier. *Principles of adaptive optics*. CRC press, 2022.

[38] J. Collier, C. Hooker, S. Hawkes, C. Edwards, K Haefner, and K Braeur. Adaptive Optics for the Petawatt upgrade. *Central Laser Facility Annual Report*, pages 181–182, 2001.

[39] L Seifert, HJ Tiziani, and W Osten. Wavefront reconstruction with the adaptive shack–hartmann sensor. *Optics communications*, 245(1-6):255–269, 2005.

[40] Gustave Pariente, V. Gallet, Antonin Borot, Olivier Gobert, and Fabien Quéré. Space–time characterization of ultra-intense femtosecond laser beams. *Nature Photonics*, 10:547–553, 2016.

[41] Antoine Jeandet, Spencer W. Jolly, Antonin Borot, Benoît Bussière, Paul Dumont, Julien Gautier, Olivier Gobert, Jean-Philippe Goddet, Anthony Gonsalves, Arie Irman, Wim P. Leemans, Rodrigo Lopez-Martens, Gabriel Mennerat, Kei Nakamura, Marie Ouillé, Gustave Pariente, Moana Pittman, Thomas Püschel, Fabrice Sanson, François Sylla, Cédric Thaury, Karl Zeil, and Fabien Quéré. Survey of spatio-temporal couplings throughout high-power ultrashort lasers. *Opt. Express*, 30(3):3262–3288, Jan 2022.

[42] B Webb, MJ Guardalben, C Dorrer, S Bucht, and J Bromage. Simulation of grating compressor misalignment tolerances and mitigation strategies for chirped-pulse–amplification systems of varying bandwidths and beam sizes. *Applied optics*, 58(2):234–243, 2019.

[43] Yang Zhao, Fenxiang Wu, Cheng Wang, Jiabing Hu, Zongxin Zhang, Xingyan Liu, Xiaojun Yang, Peile Bai, Haidong Chen, Jiayi Qian, et al. Spatiotemporal aberrations due to the groove density mismatching of compression gratings in ultra-intense femtosecond lasers. *Scientific Reports*, 14(1):18231, 2024.

[44] Andrew Forbes, Michael de Oliveira, and Mark R Dennis. Structured light. *Nature Photonics*, 15(4):253–262, 2021.

[45] Chao He, Yijie Shen, and Andrew Forbes. Towards higher-dimensional structured light. *Light: Science & Applications*, 11(1):205, 2022.

[46] Andrew Forbes Justin Harrison, Darryl Naidoo and Angela Dudley. Progress in high-power and high-intensity structured light. *Advances in Physics: X*, 9(1):2327453, 2024.

[47] D Mugnai, A Ranfagni, and R Ruggeri. Observation of superluminal behaviors in wave propagation. *Physical review letters*, 84(21):4830, 2000.

[48] Dustin H Froula, David Turnbull, Andrew S Davies, Terrance J Kessler, Dan Haberberger, John P Palastro, Seung-Whan Bahk, Ildar A Begishev, Robert Boni, Sara Bucht, et al. Spatiotemporal control of laser intensity. *Nature photonics*, 12(5):262–265, 2018.

[49] D. H. Froula, J. P. Palastro, D. Turnbull, A. Davies, L. Nguyen, A. Howard, D. Ramsey, P. Franke, S.-W. Bahk, I. A. Begishev, R. Boni, J. Bromage, S. Bucht, R. K. Follett, D. Haberberger, G. W. Jenkins, J. Katz, T. J. Kessler, J. L. Shaw, and J. Vieira. Flying focus: Spatial and temporal control of intensity for laser-based applications. *Physics of Plasmas*, 26(3):032109, 2019.

[50] Aaron Liberman, Ronan Lahaye, Slava Smartsev, Sheroy Tata, Salome Benracassa, Anton Golovanov, Eitan Levine, Cedric Thaury, and Victor Malka. Use of spatiotemporal couplings and an axiparabola to control the velocity of peak intensity. *Optics Letters*, 49(4):814–817, 2024.

[51] J. P. Palastro, J L Shaw, P Franke, D Ramsey, T T Simpson, and D. H. Froula. Dephasingless Laser Wakefield Acceleration. *Physical Review Letters*, 124(13):134802, March 2020.

[52] J Vieira, JT Mendonça, and F Quéré. Optical control of the topology of laser-plasma accelerators. *Physical Review Letters*, 121(5):054801, 2018.

[53] Eduard Atonga, Killian Martineau, Ramy Aboushelbaya, Aurélien Barrau, Marko von der Leyen, Sunny Howard, Abigail James, Jordan Lee, Chunshan Lin, Heath Martin, et al. Gravitational waves from high-power twisted light. *Physical Review D*, 110(4):044023, 2024.

[54] Ramy Aboushelbaya, Kevin Glize, Alexander F Savin, Marko Mayr, Benjamin Spiers, Robin Wang, John Collier, Mattias Marklund, Raoul MGM Trines, Robert Bingham, et al. Orbital angular momentum coupling in elastic photon-photon scattering. *Physical Review Letters*, 123(11):113604, 2019.

[55] Dai Wanjun, Hu Dongxia, Zhou Wei, Zhao Junpu, Jing Feng, Yang Zeping, Zhang Kun, Jiang Xuejun, Deng Wu, Zhao Runchang, et al. Beam wavefront control of a thermal inertia laser for inertial confinement fusion application. *Applied optics*, 48(19):3691–3694, 2009.

[56] Bruno Gonzalez-Izquierdo, Ross J Gray, Martin King, Robbie Wilson, Rachel J Dance, Haydn Powell, David A MacLellan, John McCreadie, Nicholas MH Butler, Steve Hawkes, et al. Influence of laser polarization on collective electron dynamics in ultraintense laser–foil interactions. *High Power Laser Science and Engineering*, 4:e33, 2016.

[57] A Jain and DN Gupta. Improvement of electron beam quality in laser wakefield acceleration by a circularly-polarized laser pulse. *Plasma Physics and Controlled Fusion*, 63(7):075007, 2021.

[58] Mohammad Rezaei-Pandari, Mohammad Mirzaie, Calin Ioan Hojbota, Tae Gyu Pak, Sang Beom Kim, Geon Woo Lee, Reza Massudi, Ali Reza Niknam, Seong Ku Lee, Ki-Yong Kim, et al. Laser wakefield electron acceleration with polarization-dependent ionization injection. *Physical Review Applied*, 20(3):034026, 2023.

[59] Vivek Sharma, Sandeep Kumar, Niti Kant, and Vishal Thakur. Enhanced laser wakefield acceleration by a circularly polarized laser pulse in obliquely magnetized under-dense plasma. *Optical and Quantum Electronics*, 55(13):1150, 2023.

[60] Zhi-Kai Lin, Jian-Xun Liu, Ting Gao, Jin-Jian Lv, Cang-Zhen Meng, Hong-Bin Jin, Yuan Zhao, Tong-Pu Yu, and Jun Zhao. Positron generation via ultra-intense circularly polarized laser pulses colliding in near-critical-density plasmas with different thickness. *Physics of Plasmas*, 29(8), 2022.

[61] Feng Wan, Rashid Shaisultanov, Yan-Fei Li, Karen Z Hatsagortsyan, Christoph H Keitel, and Jian-Xing Li. Ultrarelativistic polarized positron jets via collision of electron and ultraintense laser beams. *Physics Letters B*, 800:135120, 2020.

[62] XF Li, P Gibbon, A Hützen, M Büscher, SM Weng, M Chen, and ZM Sheng. Polarized proton acceleration in ultraintense laser interaction with near-critical-density plasmas. *Physical Review E*, 104(1):015216, 2021.

[63] Xiaomei Zhang, Baifei Shen, Xuemei Li, Zhangying Jin, Fengchao Wang, and Meng Wen. Efficient gev ion generation by ultraintense circularly polarized laser pulse. *Physics of Plasmas*, 14(12), 2007.

[64] Wenpeng Wang. Proton acceleration driven by relativistic femtosecond laguerre–gaussian lasers. *Reviews of Modern Plasma Physics*, 8(1):1–29, 2024.

[65] Neïl Zaïm, Diego Guénot, Ludovic Chopineau, Adrien Denoeud, Olle Lundh, Henri Vincenti, Fabien Quéré, and Jérôme Faure. Interaction of ultraintense radially-polarized laser pulses with plasma mirrors. *Physical Review X*, 10(4):041064, 2020.

[66] Goëry Genty, Lauri Salmela, John M Dudley, Daniel Brunner, Alexey Kokhanovskiy, Sergei Kobtsev, and Sergei K Turitsyn. Machine learning and applications in ultrafast photonics. *Nature Photonics*, 15(2):91–101, 2021.

[67] Andreas Döpp, Christoph Eberle, Sunny Howard, Faran Irshad, Jinpu Lin, and Matthew Streeter. Data-driven science and machine learning methods in laser–plasma physics. *High Power Laser Science and Engineering*, 11:e55, 2023.

[68] Peter W Hatfield, Jim A Gaffney, Gemma J Anderson, Suzanne Ali, Luca Antonelli, Suzan Başeğmez du Pree, Jonathan Citrin, Marta Fajardo, Patrick Knapp, Brendan Kettle, et al. The data-driven future of high-energy-density physics. *Nature*, 593(7859):351–361, 2021.

[69] Miguel Miranda, Marija Kotur, Piotr Rudawski, Chen Guo, Anne Harth, Anne L'Huillier, and Cord L Arnold. Spatiotemporal characterization of ultrashort laser pulses using spatially resolved Fourier transform spectrometry. *Optics Letters*, 39(17):5142, 2014.

[70] Benjamín Alonso, Ignacio Lopez-Quintas, Warein Holgado, Rokas Drevinskas, Peter G Kazansky, Carlos Hernández-García, and Íñigo J Sola. Complete spatiotemporal and polarization characterization of ultrafast vector beams. *Communications Physics*, 3(1):151, 2020.

[71] Apostolos Zdagkas, Venkatram Nalla, Nikitas Papasimakis, and Nikolay I Zheludev. Spatio-temporal characterization of ultrashort vector pulses. *APL Photonics*, 6(11), 2021.

[72] Benjamín Alonso, Andreas Döpp, and Spencer W Jolly. Space–time characterization of ultrashort laser pulses: A perspective. *APL Photonics*, 9(7), 2024.

[73] Arnold Sommerfeld, Otto Laporte, and Peter A Moldauer. Optics: Vol. 5 of lectures on theoretical physics, 1955.

[74] Joseph W Goodman. Introduction to fourier optics. *Introduction to Fourier optics, 3rd ed., by JW Goodman. Englewood, CO: Roberts & Co. Publishers, 2005*, 1, 2005.

[75] Fabin Shen and Anbo Wang. Fast-fourier-transform based numerical integration method for the rayleigh-sommerfeld diffraction formula. *Applied optics*, 45(6):1102–1110, 2006.

[76] Soheil Mehrabkhani and Thomas Schneider. Is the rayleigh-sommerfeld diffraction always an exact reference for high speed diffraction algorithms? *Opt. Express*, 25(24):30229–30240, Nov 2017.

[77] Arvind S Marathay and John F McCalmont. Vector diffraction theory for electromagnetic waves. *Journal of the Optical Society of America A*, 18(10):2585–2593, 2001.

[78] Rafael G González-Acuña and Simon Thibault. Exact vectorial model for nonparaxial focusing of freeform wavefronts. *Optics Express*, 30(13):23656–23663, 2022.

[79] A Einstein. 7 on the quantum theory of radiation. *The Old Quantum Theory: The Commonwealth and International Library: Selected Readings in Physics*, page 167, 2016.

[80] H. Yokoyama and K. Ujihara. *Spontaneous Emission and Laser Oscillation in Microcavities*. Laser & Optical Science & Technology. Taylor & Francis, 1995.

[81] Karl F Renk. *Basics of laser physics*. Springer, 2012.

[82] Arthur L Schawlow and Charles H Townes. Infrared and optical masers. *Physical review*, 112(6):1940, 1958.

[83] Joshua Yang, Kasper Van Gasse, Daniil M Lukin, Melissa A Guidry, Geun Ho Ahn, Alexander D White, and Jelena Vučković. Titanium: sapphire-on-insulator integrated lasers and amplifiers. *Nature*, 630(8018):853–859, 2024.

[84] Peter F Moulton. Spectroscopic and laser characteristics of ti: Al 2 o 3. *JOSA B*, 3(1):125–133, 1986.

[85] Jean-Claude Diels and Wolfgang Rudolph. *Ultrashort laser pulse phenomena*. Elsevier, 2006.

[86] David H Auston, KB Eisenthal, RM Hochstrasser, CK Johnson, A Laubereau, Dvd Linde, A Seilmeier, CV Shank, and W Zinth. *Ultrashort laser pulses: Generation and applications*, volume 60. Springer Science & Business Media, 2006.

[87] Simon Hooker and Colin Webb. *Laser physics*, volume 9. Oxford University Press, 2010.

[88] P Maine, D Strickland, P Bado, M Pessot, and G Mourou. Generation of ultrahigh peak power pulses by chirped pulse amplification. *IEEE Journal of Quantum electronics*, 24(2):398–403, 1988.

[89] Ivan Vladimirovich Yakovlev. Stretchers and compressors for ultra-high power laser systems. *Quantum Electronics*, 44(5):393, 2014.

[90] Jiro Itatani, Yasuo Nabekawa, Kiminori Kondo, and Shuntaro Watanabe. Generation of 13-tw, 26-fs pulses in a ti: sapphire laser. *Optics communications*, 134(1-6):134–138, 1997.

[91] Iustin Ouatu, BT Spiers, Ramy Aboushelbaya, Qingsong Feng, MW von der Leyen, RW Paddock, Robin Timmis, C Ticos, KM Krushelnick, and PA Norreys. Ionization states for the multipetawatt laser-qed regime. *Physical Review E*, 106(1):015205, 2022.

[92] Spencer Jolly, Olivier Gobert, and Fabien Quere. Spatio-temporal characterization of ultrashort laser beams: a tutorial. *Journal of Optics*, 22, 08 2020.

[93] R. W. Gerchberg. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

[94] Chao Zuo, Jiaji Li, Jiasong Sun, Yao Fan, Jialin Zhang, Linpeng Lu, Runnan Zhang, Bowen Wang, Lei Huang, and Qian Chen. Transport of intensity equation: a tutorial. *Optics and Lasers in Engineering*, 135:106187, 2020.

[95] Yohei Nishizaki, Matias Valdivia, Ryoichi Horisaki, Katsuhisa Kitaguchi, Mamoru Saito, Jun Tanida, and Esteban Vera. Deep learning wavefront sensing. *Opt. Express*, 27(1):240–251, Jan 2019.

[96] JetLSOGNO Primot and L Sogno. Achromatic three-wave (or more) lateral shearing interferometer. *JOSA A*, 12(12):2679–2685, 1995.

[97] Jean-Christophe Chanteloup, Frédéric Druon, Marc Nantel, Anatoly Maksimchuk, and Gérard Mourou. Single-shot wave-front measurements of high-intensity ultrashort laser pulses with a three-wave interferometer. *Opt. Lett.*, 23(8):621–623, Apr 1998.

[98] Sabrina Velghe, Jérôme Primot, Nicolas Guérineau, Riad Haïdar, Sébastien Demoustier, Mathieu Cohen, and Benoît Wattellier. Advanced wave-front sensing by quadri-wave lateral shearing interferometry. *Interferometry XIII: Techniques and Analysis*, 6292:62920E, 2006.

[99] Bastien Rouzé, Georges Giakoumakis, Adrien Stolidi, Julien Jaeck, Cindy Bellanger, and Jérôme Primot. Extracting more than two orthogonal derivatives from a shack-hartmann wavefront sensor. *Optics Express*, 29(4):5193–5204, 2021.

[100] Mitsuo Takeda, Hideki Ina, and Seiji Kobayashi. Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *J. Opt. Soc. Am.*, 72(1):156–160, Jan 1982.

[101] Sabrina Velghe, Jérôme Primot, Nicolas Guérineau, Mathieu Cohen, and Benoit Wattellier. Wave-front reconstruction from multidirectional phase derivatives generated by multilateral shearing interferometers. *Opt. Lett.*, 30(3):245–247, Feb 2005.

[102] Fengzhao Dai, Jie Li, Xiangzhao Wang, and Yang Bu. Exact two-dimensional zonal wavefront reconstruction with high spatial resolution in lateral shearing interferometry. *Optics Communications*, 367:264–273, 2016.

[103] Biswajit Pathak and Bosanta R Boruah. Improved wavefront reconstruction algorithm for shack–hartmann type wavefront sensors. *Journal of Optics*, 16(5):055403, apr 2014.

[104] William H Southwell. Wave-front estimation from wave-front slope measurements. *JOsA*, 70(8):998–1006, 1980.

[105] J-C Chanteloup. Multiple-wave lateral shearing interferometry for wave-front sensing. *Applied optics*, 44(9):1559–1571, 2005.

[106] Xinkang Tian, Masahide Itoh, and Toyohiko Yatagai. Simple algorithm for large-grid phase reconstruction of lateral-shearing interferometry. *Applied optics*, 34(31):7213–7220, 1995.

[107] F Zernike. Phase contrast, a new method for the microscopic observation of transparent objects Part II. *Physica*, 9(10):974–986, 1942.

[108] Junzhong Liang, Bernhard Grimm, Stefan Goelz, and Josef F Bille. Objective measurement of wave aberrations of the human eye with the use of a hartmann–shack wave-front sensor. *JOSA A*, 11(7):1949–1957, 1994.

[109] Yulong He, Zhiwei Liu, Yu Ning, Jun Li, Xiaojun Xu, and Zongfu Jiang. Deep learning wavefront sensing method for shack-hartmann sensors with sparse sub-apertures. *Opt. Express*, 29(11):17669–17682, May 2021.

[110] Sunny Howard, Jannik Esslinger, Robin HW Wang, Peter Norreys, and Andreas Döpp. Hyperspectral compressive wavefront sensing. *High Power Laser Science and Engineering*, 11:e32, 2023.

[111] G Harbers, PJ Kunst, and GWR Leibbrandt. Analysis of lateral shearing interferograms by use of zernike polynomials. *Applied optics*, 35(31):6162–6172, 1996.

[112] Fengzhao Dai, Yazhong Zheng, Yang Bu, and Xiangzhao Wang. Modal wavefront reconstruction based on zernike polynomials for lateral shearing interferometry. *Applied Optics*, 56(1):61–68, 2016.

[113] Thomas Oksenhendler, Sebastien Coudreau, Nicolas Forget, V Crozatier, S Grabielle, R Herzog, O Gobert, and D Kaplan. Self-referenced spectral interferometry. *Applied Physics B*, 99(1):7–12, 2010.

[114] D. J. Kane and R. Trebino. Characterization of Arbitrary Femtosecond Pulses Using Frequency-Resolved Optical Gating. *Ieee Journal of Quantum Electronics*, 29(2):571–579, February 1993.

[115] A. Jeandet, A. Borot, K. Nakamura, S. W. Jolly, A. J. Gonsalves, C. Tóth, H.-S. Mao, W. P. Leemans, and F. Quéré. Spatio-temporal structure of a petawatt femtosecond laser beam. *Journal of Physics: Photonics*, 1:035001, 2019.

[116] P Hariharan and D Sen. Radial shearing interferometer. *Journal of Scientific Instruments*, 38(11):428, 1961.

[117] Antonin Borot and Fabien Quere. Spatio-spectral metrology at focus of ultrashort lasers: a phase-retrieval approach. *Optics Express*, 26(20):26444–26461, October 2018.

[118] Antoine Jeandet, Antonin Borot, Kei Nakamura, Spencer W Jolly, Anthony J Gonsalves, Csaba Tóth, Hann-Shin Mao, Wim P Leemans, and Fabien Quéré. Spatio-temporal structure of a petawatt femtosecond laser beam. *Journal of Physics: Photonics*, 1(3):035001, 2019.

[119] N. Weiße, J. Esslinger, S. Howard, F. M. Foerster, F. Haberstroh, L. Doyle, P. Norreys, J. Schreiber, S. Karsch, and A. Döpp. Measuring spatio-temporal couplings using modal spatio-spectral wavefront retrieval. *arXiv preprint arXiv:2303.01360*, 2023.

[120] Seth L Cousin, Juan M Bueno, Nicolas Forget, Dane R Austin, and J Biegert. Three-dimensional spatiotemporal pulse characterization with an acousto-optic pulse shaper and a hartmann–shack wavefront sensor. *Optics letters*, 37(15):3291–3293, 2012.

[121] Pablo Gabolde and Rick Trebino. Single-shot measurement of the full spatio-temporal field of ultrashort pulses with multi-spectral digital holography. *Opt. Express*, 14(23):11460–11467, Nov 2006.

[122] Elizabeth Grace, Tammy Ma, Zhe Guang, Rana Jafari, Jaebum Park, Jerry Clark, Gregory Kemp, James Moody, Michelle Rhodes, Yuan Ping, et al. Single-shot complete spatiotemporal measurement of terawatt laser pulses. *Journal of Optics*, 23(7):075505, 2021.

[123] Yeong Gyu Kim, Ji In Kim, Jin Woo Yoon, Jae Hee Sung, Seong Ku Lee, and Change Hee Nam. Single-shot spatiotemporal characterization of a multi-pw laser using a multispectral wavefront sensing method. *Opt. Express*, 29(13):19506–19514, Jun 2021.

[124] Haocheng Tang, Ting Men, Xianglei Liu, Yaodan Hu, Jingqin Su, Yanlei Zuo, Ping Li, Jinyang Liang, Michael C Downer, and Zhengyan Li. Single-shot compressed optical field topography. *Light: Science & Applications*, 11(1):244, 2022.

[125] Da-Chao Deng and Hui-Chun Wu. Compression and acceleration of ions by ultrashort, ultraintense azimuthally polarized light. *Physical Review E*, 109(6):065211, 2024.

[126] KP Singh. Electron acceleration by a circularly polarized laser pulse in a plasma. *Physics of plasmas*, 11(8):3992–3996, 2004.

[127] M Chen, A Pukhov, ZM Sheng, and XQ Yan. Laser mode effects on the ion acceleration during circularly polarized laser pulse interaction with foil targets. *Physics of Plasmas*, 15(11), 2008.

[128] Meng Wen, Yousef I Salamin, and Christoph H Keitel. Electron acceleration by a radially-polarized laser pulse in a plasma micro-channel. *Optics Express*, 27(2):557–566, 2019.

[129] Q. Cao, P. Zheng, and Q. Zhan. Vectorial sculpturing of spatiotemporal wavepackets. *APL Photonics*, 7:096102, 2022.

[130] Charles Varin, Stéphane Payeur, Vincent Marceau, Sylvain Fourmaux, Alexandre April, Bruno Schmidt, Pierre-Louis Fortin, Nicolas Thiré, Thomas Brabec, François Légaré, et al. Direct electron acceleration with radially polarized laser beams. *Applied Sciences*, 3(1):70–93, 2013.

[131] Pengpeng Ding, Zihan Guo, Dalong Qi, Jiali Yao, Chengzhi Jin, Ning Xu, Zhiling Zhang, Yunhua Yao, Lianzhong Deng, Zhiyong Wang, et al. Polarimetric-spectral-sensitive compressed ultrafast photography enables real-time intensity visualization of time-varying polarized ultrafast laser pulses. *Laser & Photonics Reviews*, 18(5):2300873, 2024.

[132] N. Weisse, J. Esslinger, S. Howard, F. M. Foerster, F. Haberstroh, L. Doyle, P. Norreys, J. Schreiber, S. Karsch, and A. Döpp. Measuring spatio-temporal couplings using modal spatio-spectral wavefront retrieval. *Opt. Express*, 31(12):19733–19745, Jun 2023.

[133] Bo Zhang, Josiane Zerubia, and Jean-Christophe Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread function models. *Appl. Opt.*, 46(10):1819–1829, Apr 2007.

[134] G. Pretzler, A. Kasper, and K.J. Witte. Angular chirp and tilted light pulses in cpa lasers. *Applied Physics B*, 70(1):1–9, 2000.

[135] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathematique*, 346(9-10):589–592, 2008.

[136] Timothy P Lillicrap, Adam Santoro, Luke Marris, Colin J Akerman, and Geoffrey Hinton. Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6):335–346, 2020.

[137] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 12 2014.

[138] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[139] Sergey Bartunov, Adam Santoro, Blake Richards, Luke Marris, Geoffrey E Hinton, and Timothy Lillicrap. Assessing the scalability of biologically-motivated deep learning algorithms and architectures. *Advances in neural information processing systems*, 31, 2018.

[140] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[141] Roman Pogodin, Yash Mehta, Timothy Lillicrap, and Peter E Latham. Towards biologically plausible convolutional networks. *Advances in Neural Information Processing Systems*, 34:13924–13936, 2021.

[142] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.

[143] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[144] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[145] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807, Los Alamitos, CA, USA, jul 2017. IEEE Computer Society.

[146] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. arxiv e-prints. *arXiv preprint arXiv:1512.03385*, 10, 2015.

[147] Ross Wightman, Hugo Touvron, and Hervé Jégou. Resnet strikes back: An improved training procedure in timm. *arXiv preprint arXiv:2110.00476*, 2021.

[148] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[149] Meng-Hao Guo, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R Martin, Ming-Ming Cheng, and Shi-Min Hu. Attention mechanisms in computer vision: A survey. *Computational Visual Media*, 8(3):331–368, 2022.

[150] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 294–310, Cham, 2018. Springer International Publishing.

[151] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. In *Medical Imaging with Deep Learning*, 2018.

[152] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[153] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7794–7803, 2018.

[154] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015.

[155] Antoine Jeandet, Spencer W. Jolly, Antonin Borot, Benoît Bussière, Paul Dumont, Julien Gautier, Olivier Gobert, Jean-Philippe Goddet, Anthony Gonsalves, Arie Irman, Wim P. Leemans, Rodrigo Lopez-Martens, Gabriel Mennerat, Kei Nakamura, Marie Ouillé, Gustave Pariente, Moana Pittman, Thomas Püschel, Fabrice Sanson, François Sylla, Cédric Thaury, Karl Zeil, and Fabien Quéré. Survey of spatio-temporal couplings throughout high-power ultrashort lasers. *Opt. Express*, 30(3):3262–3288, Jan 2022.

[156] Pamela Bowlan, Pablo Gabolde, Aparna Shreenath, Kristan McGresham, Rick Trebino, and Selcuk Akturk. Crossed-beam spectral interferometry: a simple, high-spectral-resolution method for completely characterizing complex ultrashort pulses in real time. *Opt. Express*, 14(24):11892–11900, Nov 2006.

[157] Miguel López-Ripa, Ínigo J Sola, and Benjamín Alonso. Bulk lateral shearing interferometry for spatiotemporal study of time-varying ultrashort optical vortices. *Photon. Res.*, 10(4):922–931, Apr 2022.

[158] Seth L. Cousin, Juan M. Bueno, Nicolas Forget, Dane R. Austin, and J. Biegert. Three-dimensional spatiotemporal pulse characterization with an acousto-optic pulse shaper and a hartmann-shack wavefront sensor. *Opt. Lett.*, 37(15):3291–3293, Aug 2012.

[159] Lejia Hu, Shuwen Hu, Wei Gong, and Ke Si. Learning-based shack-hartmann wavefront sensor for high-order aberration detection. *Opt. Express*, 27(23):33504–33517, Nov 2019.

[160] Lejia Hu, Shuwen Hu, Wei Gong, and Ke Si. Deep learning assisted shack–hartmann wavefront sensor for direct wavefront detection. *Opt. Lett.*, 45(13):3741–3744, Jul 2020.

[161] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Opt. Express*, 15(21):14013–14027, Oct 2007.

[162] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 47(10):B44–B51, Apr 2008.

[163] Peng Wang and Lihong V Wang. Single-shot reconfigurable femtosecond imaging of ultrafast optical dynamics. *Advanced Science*, 10(13):2207222, 2023.

[164] Mu Qiao, Yangyang Sun, Jiawei Ma, Ziyi Meng, Xuan Liu, and Xin Yuan. Snapshot coherence tomographic imaging. *IEEE Transactions on Computational Imaging*, 7:624–637, 2021.

[165] Xin Yuan, David J. Brady, and Aggelos K. Katsaggelos. Snapshot compressive imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 38(2):65–88, 2021.

[166] Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.

[167] Zhuoyuan Wu, Jian Zhang, and Chong Mou. Dense deep unfolding network with 3d-cnn prior for snapshot compressive imaging. *arXiv preprint arXiv:2109.06548*, 2021.

[168] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, may 2018.

[169] Lizhi Wang, Chen Sun, Ying Fu, Min H. Kim, and Hua Huang. Hyperspectral image reconstruction using a deep spatial-spectral prior. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8024–8033, 2019.

[170] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[171] Lizhi Wang, Chen Sun, Maoqing Zhang, Ying Fu, and Hua Huang. Dnu: Deep non-local unrolling for computational spectral imaging. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1658–1668, 2020.

[172] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, and Feng Wu. Dual-camera design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 54(4):848–858, Feb 2015.

[173] F. Yasuma, T. Mitsunaga, D. Iso, and S.K. Nayar. *Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum*. Nov 2008.

[174] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image processing*, 16(12):2992–3004, 2007.

[175] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Wenjun Zeng, and Feng Wu. Compressive hyperspectral imaging with complementary rgb measurements. In *2016 Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2016.

[176] Lingfei Song, Lizhi Wang, Min H. Kim, and Hua Huang. High-accuracy image formation model for coded aperture snapshot spectral imaging. *IEEE Transactions on Computational Imaging*, 8:188–200, 2022.

[177] Jingdan Liu, Miguel Marquez, Yingming Lai, Heide Ibrahim, Katherine Légaré, Philippe Lassonde, Xianglei Liu, Michel Hehn, Stéphane Mangin, Grégory Malinowski, et al. Swept coded aperture real-time femtophotography. *Nature Communications*, 15(1):1589, 2024.

[178] Chengzhi Jin, Yingming Xu, Dalong Qi, Yunhua Yao, Yuecheng Shen, Lianzhong Deng, Ruozhong Han, Zhen Pan, Jiali Yao, Yilin He, et al. Single-shot intensity-and phase-sensitive compressive sensing-based coherent modulation ultrafast imaging. *Physical Review Letters*, 132(17):173801, 2024.

[179] Xinkun Chu, Hao Zhang, Zhiyu Tian, Qing Zhang, Fang Wang, Jing Chen, and Yuanchao Geng. Detection of laser-induced optical defects based on image segmentation. *High Power Laser Science and Engineering*, 7:e66, 2019.

[180] Zhan Li, Lu Han, Xiaoping Ouyang, Pan Zhang, Yajing Guo, Dean Liu, and Jianqiang Zhu. Three-dimensional laser damage positioning by a deep-learning method. *Optics Express*, 28(7):10165–10178, 2020.

[181] Jinpu Lin, Florian Haberstroh, Stefan Karsch, and Andreas Döpp. Applications of object detection networks in high-power laser systems and experiments. *High Power Laser Science and Engineering*, 11:e7, 2023.

[182] Xu Jia, Bert De Brabandere, Tinne Tuytelaars, and Luc V Gool. Dynamic filter networks. *Advances in neural information processing systems*, 29, 2016.

[183] O. Semih Kayhan and J. C. van Gemert. On translation invariance in cnns: Convolutional layers can exploit absolute spatial location. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14262–14273, Los Alamitos, CA, USA, jun 2020. IEEE Computer Society.

[184] Md Amirul Islam, Sen Jia, and Neil D. B. Bruce. How much position information do convolutional neural networks encode? 2020.

[185] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[186] Xinlong Wang, Tao Kong, Chunhua Shen, Yuning Jiang, and Lei Li. Solo: Segmenting objects by locations. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*, pages 649–665. Springer, 2020.

[187] Hengshuang Zhao, Jiaya Jia, and Vladlen Koltun. Exploring self-attention for image recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10073–10082, 2020.

[188] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2437–2446, 2019.

[189] R. El Jurdi, C. Petitjean, P. Honeine, and F. Abdallah. Coordconv-unet: Investigating coordconv for organ segmentation. *IRBM*, 42(6):415–423, 2021.

[190] Yongpeng Dai, Tian Jin, Yongkun Song, Shilong Sun, and Chen Wu. Convolutional neural network with spatial-variant convolution kernel. *Remote Sensing*, 12(17), 2020.

[191] Arnas Uselis, Mantas Lukoševičius, and Lukas Stasytis. Localized convolutional neural networks for geospatial wind forecasting. *Energies*, 13(13):3440, 2020.

[192] Hang Su, Varun Jampani, Deqing Sun, Orazio Gallo, Erik Learned-Miller, and Jan Kautz. Pixel-adaptive convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11166–11175, 2019.

[193] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann Lecun. Spectral networks and locally connected networks on graphs. In *International Conference on Learning Representations (ICLR2014), CBLS, April 2014*, 2014.

[194] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022.

[195] Jiawei Yang, Katie Z Luo, Jiefeng Li, Congyue Deng, Leonidas Guibas, Dilip Krishnan, Kilian Q Weinberger, Yonglong Tian, and Yue Wang. Denoising vision transformers. *arXiv preprint arXiv:2401.02957*, 2024.

[196] Kyrollos Yanny, Kristina Monakhova, Richard W. Shuai, and Laura Waller. Deep learning for fast spatially varying deconvolution. *Optica*, 9(1):96–99, Jan 2022.

[197] Bogdan Toader, Jérôme Boulanger, Yury Korolev, Martin O Lenz, James Manton, Carola-Bibiane Schönlieb, and Leila Mureşan. Image reconstruction in Light-Sheet microscopy: Spatially varying deconvolution and mixed noise. *J Math Imaging Vis*, 64(9):968–992, June 2022.

[198] Tod Lauer. Deconvolution with a spatially-variant psf. *Proc SPIE*, 4847, 08 2002.

[199] Farrens, S., Ngolè Mboula, F. M., and Starck, J.-L. Space variant deconvolution of galaxy survey images. *A&A*, 601:A66, 2017.

[200] Kyrollos Yanny, Nick Antipa, William Liberti, Sam Dehaeck, Kristina Monakhova, Fanglin Linda Liu, Konlin Shen, Ren Ng, and Laura Waller. Miniscope3d: optimized single-shot miniature 3d fluorescence microscopy. *Light, Science & Applications*, 9, 2020.

[201] Vebjorn Ljosa, Katherine Sokolnicki, and Anne Carpenter. Annotated high-throughput microscopy image sets for validation. *Nature methods*, 9:637, 06 2012.

[202] Y. Zhang, Y. Zhu, E. Nichols, Q. Wang, S. Zhang, C. Smith, and S. Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11702–11710, Los Alamitos, CA, USA, jun 2019. IEEE Computer Society.

[203] Gamaleldin Elsayed, Prajit Ramachandran, Jonathon Shlens, and Simon Kornblith. Revisiting spatial invariance with low-rank local connectivity. In *International Conference on Machine Learning*, pages 2868–2879. PMLR, 2020.

[204] Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

[205] Yijie Shen, Xuejiao Wang, Zhenwei Xie, Changjun Min, Xing Fu, Qiang Liu, Mali Gong, and Xiaocong Yuan. Optical vortices 30 years on: Oam manipulation from topological charge to multiple singularities. *Light: Science & Applications*, 8(1):90, 2019.

[206] David McGloin and Kishan Dholakia. Bessel beams: diffraction in a new light. *Contemporary physics*, 46(1):15–28, 2005.

[207] Nikolaos K Efremidis, Zhigang Chen, Mordechai Segev, and Demetrios N Christodoulides. Airy beams and accelerating waves: an overview of recent advances. *Optica*, 6(5):686–701, 2019.

[208] Robert J Noll. Zernike polynomials and atmospheric turbulence. *JOsA*, 66(3):207–211, 1976.

[209] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[210] Ron Rubinstein, Alfred M Bruckstein, and Michael Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.

[211] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[212] Torben B Andersen. Efficient and robust recurrence relations for the zernike circle polynomials and their derivatives in cartesian coordinates. *Optics Express*, 26(15):18878–18896, 2018.

[213] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[214] Seung-Jean Kim, Kwangmoo Koh, Michael Lustig, Stephen Boyd, and Dimitry Gorinevsky. An interior-point method for large-scale \ℓ_1-regularized least squares. *IEEE journal of selected topics in signal processing*, 1(4):606–617, 2007.

[215] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trend in Machine learning*, 3(1):1–122, 2011.

[216] John H McLeod. The axicon: a new type of optical element. *JOSA*, 44(8):592–597, 1954.

[217] V Garcés-Chávez, David McGloin, H Melville, Wilson Sibbett, and Kishan Dholakia. Simultaneous micromanipulation in multiple planes using a self-reconstructing light beam. *Nature*, 419(6903):145–147, 2002.

[218] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[219] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

[220] Ignacio Lopez-Quintas, Warein Holgado, Rokas Drevinskas, Peter G Kazansky, Íñigo J Sola, and Benjamín Alonso. Optical vortex production mediated by azimuthal index of radial polarization. *Journal of Optics*, 22(9):095402, 2020.

[221] Miguel López-Ripa, Íñigo J Sola, and Benjamín Alonso. Bulk lateral shearing interferometry for spatiotemporal study of time-varying ultrashort optical vortices. *Photonics Research*, 10(4):922–931, 2022.

[222] Benjamín Alonso, Warein Holgado, and Íñigo J Sola. Compact in-line temporal measurement of laser pulses with amplitude swing. *Optics Express*, 28(10):15625–15640, 2020.

[223] Miguel López-Ripa, Óscar Pérez-Benito, Benjamín Alonso, Rosa Weigand, and Íñigo Sola. Few-cycle pulse retrieval using amplitude swing technique. *Optics Express*, 32(12):21149–21159, 2024.

[224] Churchill Eisenhart. Expression of the uncertainties of final results: Clear statements of the uncertainties of reported values are needed for their critical evaluation. *Science*, 160(3833):1201–1204, 1968.

[225] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[226] D.A. Nix and A.S. Weigend. Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, volume 1, pages 55–60 vol.1, 1994.

[227] Nicki Skafte, Martin Jørgensen, and Søren Hauberg. Reliable training and estimation of variance networks. *Advances in Neural Information Processing Systems*, 32, 2019.

[228] Fred C Leone, Lloyd S Nelson, and RB Nottingham. The folded normal distribution. *Technometrics*, 3(4):543–550, 1961.

[229] William H Southwell. Wave-front estimation from wave-front slope measurements. *JOsA*, 70(8):998–1006, 1980.

[230] BD Hall. On the propagation of uncertainty in complex-valued quantities. *Metrologia*, 41(3):173, 2004.

[231] Ángel F Doval, Cristina Trillo, J Carlos López-Vázquez, and José L Fernández. Propagation of the measurement uncertainty in fourier transform digital holographic interferometry. *Optical Engineering*, 55(12):121709–121709, 2016.

[232] Lucid Vision Labs. Phoenix 5.0 mp polarized model. https://thinklucid.com/product/phoenix-5-0-mp-polarized-model/, n.d. Accessed: 2025-04-21.

[233] Antoinette Jane Taylor, George Rodriguez, and Tracy Sharp Clement. Determination of n 2 by direct measurement of the optical phase. *Optics letters*, 21(22):1812–1814, 1996.

[234] Seryeyohan Cho, Jihoon Jeong, Seungjin Hwang, and Tae Jun Yu. Thermal lens effect model of ti: sapphire for use in high-power laser amplifiers. *Applied Physics Express*, 11(9):092701, 2018.

[235] WP Wang, C Jiang, H Dong, XM Lu, JF Li, RJ Xu, YJ Sun, LH Yu, Z Guo, XY Liang, et al. Hollow plasma acceleration driven by a relativistic reflected hollow laser. *Physical Review Letters*, 125(3):034801, 2020.

[236] Jorge Vieira, Raoul MGM Trines, Eduardo Paulo Alves, RA Fonseca, JT Mendonça, Robert Bingham, Peter Norreys, and LO Silva. Amplification and generation of ultra-intense twisted laser pulses via stimulated raman scattering. *Nature communications*, 7(1):10371, 2016.

[237] Mindaugas Gecevičius, Rokas Drevinskas, Martynas Beresna, and Peter G. Kazansky. Single beam optical vortex tweezers with tunable orbital angular momentum. *Applied Physics Letters*, 104(23):231110, 06 2014.

[238] Sarayut Deachapunya, Sorakrai Srisuphaphon, and Sitti Buathong. Production of orbital angular momentum states of optical vortex beams using a vortex half-wave retarder with double-pass configuration. *Scientific Reports*, 12(1):6061, 2022.

[239] Carmelo Rosales-Guzmán, Bienvenu Ndagano, and Andrew Forbes. A review of complex vector light fields and their applications. *Journal of Optics*, 20(12):123001, 2018.

[240] P Winkler, M Trunk, L Hübner, A Martinez de la Ossa, S Jalas, M Kirchen, I Agapov, SA Antipov, R Brinkmann, T Eichner, et al. Active energy compression of a laser-plasma electron beam. *Nature*, pages 1–4, 2025.

[241] Ryan Tibshirani. Coordinate descent. *Course Convex Optimization*, pages 10–725, 2022.

Appendix A

Sparse Optimization

A.1 Equivalency of the Minimization of the L0 and L1 Norm

In the reconstruction step of Compressed Sensing, one seeks the most-sparse solution; ie, the solution that minimizes the L0 norm,

$$\min_{\tilde{f}} |\tilde{f}|_0 \quad \text{subject to} \quad y = (\Phi\Psi)\tilde{f}. \tag{A.1}$$

Unfortunately, this is not computationally smooth. However; minimizing the L1 norm can achieve the same result in many situations. To see why, consider the following example for f with 2 elements,

$$(\Phi\Psi) \quad f \quad = y, \tag{A.2}$$

$$\begin{bmatrix} 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \end{bmatrix} = 1. \tag{A.3}$$

The solutions to such an equation are shown in Fig. A.1. To demonstrate the ability of the L1 norm to select the sparse solutions, one can turn to a geometrical argument.

Consider first, minimizing the L2 norm, until one finds the solution with the minimum value of the L2 norm. This is shown in figure Fig. A.2a. The L2 norm is a circle in the 2d space, and in the general case, its minimization does not result in a sparse solution. On the other hand, the L1 norm forms a diamond shape, resulting in its minimization biasing solutions on the axes, which are the sparse solutions.

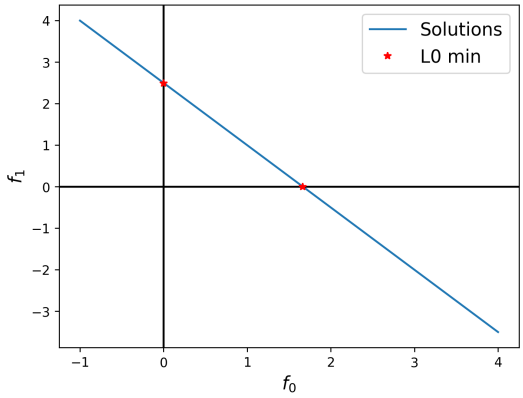
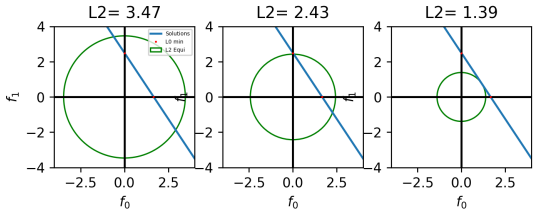
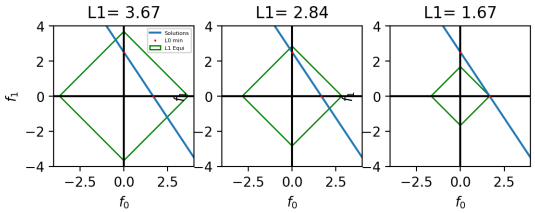


Fig. A.1 Showing the possible solutions for the system in Eq. (A.3)

Although this example is only two dimensional, the argument follows for high dimensional spaces. As the system of equations is linear, the solutions will draw a line through this higher dimensional space. The solution with the minimum L1 norm will result in sparsity.



(a) L_2 norm minimization.



(b) L_1 norm minimization.

Fig. A.2 Effect of minimizing the norms for both the L_1 and L_2. From right to left is the successive minimization, until the solution is reach with the minimum value of the norm.

A.2 LASSO Regression

The least absolute shrinkage and selection operator (LASSO) is a technique that leverages the L1 norm to find regularised solutions to least squares problems. Consider one has a measurement vector, $\vec{y}$, and a dictionary, $\mathbf{x}$. One wishes to find the sparsest possible coefficient vector c, subject to the condition that $\vec{y} = \mathbf{x}\vec{c}$. The minimization problem described by LASSO is,

$$\min_{c} \{||\vec{y} - \mathbf{x}\vec{c}||_2 + \alpha||\vec{c}||_1\} \tag{A.4}$$

where $||a||_2 = \sum_j a_j^2$, $||c||_1 = \sum_j |c_j|$, and α is a constant determining the level of sparsity.

The scikit-learn implementation solves this problem using coordinate descent. Here, the modal coefficients are optimized one at a time[241]. To optimize parameter c_i, we first define $\mathbf{x}_{-i}$ and c_{-i} as the dictionary and coefficients with index i removed, meaning the remaining residual for component i is found by $(y - \mathbf{x}_{-i}c_{-i})$. Solving the LASSO equation for c_i gives,

$$c_i = \frac{\mathbf{x}_i^T(y - \mathbf{x}_{-i}c_{-i})}{\mathbf{x}_i^T\mathbf{x}_i} - \frac{\alpha}{\mathbf{x}_i^T\mathbf{x}_i}\frac{\partial|c_i|}{\partial c_i}. \tag{A.5}$$

where $\frac{\partial|c_i|}{\partial c_i} = \text{sign}(c_i)$. This equation clearly has a term from the least squares, and a term from the L1 regularisation. To enforce sparsity, a soft threshold is used to set the parameter to zero if its value is under a threshold.

$$c_i = \text{sign}(c_i) \cdot \min\left(\left|\frac{\mathbf{x}_i^T(y - \mathbf{x}_{-i}c_{-i})}{\mathbf{x}_i^T\mathbf{x}_i}\right| - \frac{\alpha}{\mathbf{x}_i^T\mathbf{x}_i}, 0\right) \tag{A.6}$$

This process is iterated over all coefficients $i \in [1, N_c]$, and is then repeated until the final error falls below a predefined threshold, or a predefined termination number of loops is reached.

Appendix B

Energy Distribution in Airy Rings

We briefly revise the theory of focusing laser pulses contained by slit and top-hat apertures. Consider a rectangular aperture of width D, which can be defined as

$$\text{rect}\left(\frac{x}{D}\right) = \begin{cases} 1 & \text{for } |x| \leq \frac{D}{2} \\ 0 & \text{otherwise.} \end{cases}$$

The Fourier transform of this rectangular function yields a sinc function, defined as $\text{sinc}(u) = \sin(\pi u)/(\pi u)$. If v_x denotes the spatial frequency variable corresponding to x, the Fourier transform is

$$\mathcal{F}\left\{\text{rect}\left(\frac{x}{D}\right)\right\} = D \cdot \text{sinc}(v_x D).$$

The width D of the aperture inversely affects the main lobe width of the sinc function. Note that the zero-crossings of the sinc function are located at $[v_x = n/D \text{ for } n = \pm 1, \pm 2, \ldots]$. According to the Nyquist–Shannon sampling theorem, the sampling rate must be at least twice the highest frequency present in the signal. So to sample up to the n^{th} zero-crossing of the sinc function we need a sampling rate of $2n/D$ or, equivalently, a spatial resolution $\Delta x = D/(2n)$. In the case of a rectangular aperture, we get $\Delta x = D_x/(2n)$ and

$\Delta y = D_y/(2n)$. The same argumentation holds for the spectrum distribution of the laser, where $\Delta\lambda = \Delta\lambda_{pulse}/(2n)$.

In high-intensity laser applications, a top-hat beam profile is often used to create a uniform fluence profile. A top-hat forms a circular aperture with diameter D and the far-field intensity distribution is described by the Airy pattern. The aperture function in this case is

$$\text{rect}\left(\frac{r}{D}\right) = \begin{cases} 1 & \text{for } r \leq \frac{D}{2} \\ 0 & \text{otherwise,} \end{cases}$$

where r is the radial coordinate. Following the theory of Fraunhofer diffraction, the intensity distribution in the far field for a circular aperture is

$$I(\theta) = I_0 \left(\frac{2J_1\left(k\frac{D}{2}\sin(\theta)\right)}{k\frac{D}{2}\sin(\theta)} \right)^2 .$$

Here, J_1 is the Bessel function of the first kind of order one, $k = \frac{2\pi}{\lambda}$ is the wavenumber, and θ is the angle of observation. The first zero crossing of the Bessel function $J_1(x)$ at $x = 3.8317$ marks the size of the first dark ring in the diffraction pattern and thus defines the size of the Airy disk, which is equivalent to the well-known Rayleigh criterion for resolution

$$r_1 = f \cdot \sin(\theta) = 1.22 \cdot f\frac{\lambda}{D} = 1.22\lambda F,$$

where f is the focal length and $F = \frac{f}{D}$ is the f-number. Subsequent zero-crossings are located at $r_2 \simeq 7.01$, $r_3 \simeq 10.1735$, ..., up to $r_n \simeq (n + \frac{1}{4}))\pi$ for $n \gg 1$. Thus, for an Airy beam the approximately same sampling condition holds as for rectangular beams. The total energy within a circle of given size in the diffraction plane is

$$E(\theta) = E_0\left(1 - J_0^2(x) - J_1^2(x)\right)$$

Table B.1 summarizes the energy distribution within the first 10 Airy rings.

Ring Number	Energy (%)	Cumulative Energy (%)
1	83.778	83.778
2	7.215	90.993
3	2.772	93.765
4	1.467	95.232
5	0.908	96.140
6	0.618	96.758
7	0.447	97.205
8	0.339	97.544
9	0.266	97.809
10	0.214	98.023

Table B.1 Energy distribution across rings

B.0.1 Peak Power of a Gaussian Pulse

The peak power of high-intensity laser systems is often estimated using analytical approximations. In particular, it is common to approximate the temporal pulse shape as Gaussian in time, based on the full width at half maximum (FWHM), Δt, observed. We can use the relationship $\Delta t = 2\sqrt{2\ln 2}\sigma_t \simeq 2.355\sigma_t$ to relate this to parameters of a Gaussian distribution.

The power is related to the energy as $P = dE/dt$, and hence, the peak power can be defined via

$$E = \int P(t)dt = P_{peak} \int \frac{P(t)}{P_{peak}} dt,$$

which in the case of a Gaussian pulse centered at $t = 0$ becomes simply

$$E = P_{peak} \int \exp\left(-\frac{t^2}{2\sigma_t^2}\right) dt.$$

We can then evaluate the integral in terms of the FWHM δt

$$\int_{-\infty}^{+\infty} \exp\left[-\frac{t^2}{2\sigma_t^2}\right] dt = \sqrt{\frac{\pi}{4\ln 2}}\Delta t.$$

from where we then arrive at a common equation for the estimated peak power:

$$P_{max} = \sqrt{\frac{4 \ln 2}{\pi}} \frac{E}{\Delta t} \simeq 0.94 \frac{E}{\Delta t}.$$

Appendix C

Additional details of the RAVEN Technique

C.1 Unique Encoding of Wavefront and Spectrum

The combination of microlens arrays and dispersive elements is a well-known configuration in hyperspectral imaging, also known as integral field spectrography from its original use in astronomy. The main idea of these devices is to divide the field of view into sub-apertures using a microlens or lenslet array, effectively focusing light from a larger region into many smaller spatial elements. This process leverages spatial multiplexing, allowing the light from each of these smaller regions to be dispersed into its constituent wavelengths by a dispersive element such as a grating or prism. By concentrating light from a larger area into a more compact region, the available space is efficiently used for spectral dispersion, providing hyperspectral information across the entire field of view, albeit at the expense of spatial resolution. The encoding process can be represented as the matrix multiplication of a system matrix $\mathbf{S}$ and a vectorised hypercube $\mathbf{m}$ to give the vectorised measurement, $\mathbf{n}$. The system matrix includes the effects of the dispersion of the grating, and spectral integration. To illustrate the unique encoding properties of the optical setup, consider a

simple example with three colours, on a one-dimensional spatial grid with five bins, see Fig. C.1.

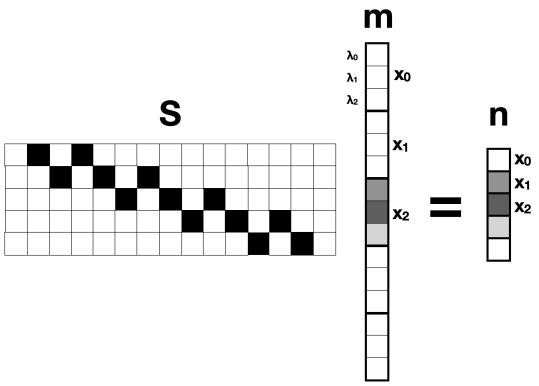


Fig. C.1 Illustration of encoding of spectral information in an integral field spectrograph.

However, when illuminated with a coherent light source such as a laser, local wavefront gradients can shift the microlens foci, i.e. the local Poynting, prior to the application of the system matrix. In this case, a single dispersion is no longer sufficient. This situation is illustrated in the following example Fig. C.2, which yields the same coded shot as the case of Fig. C.1, for a different initial measurement vector, **m**.

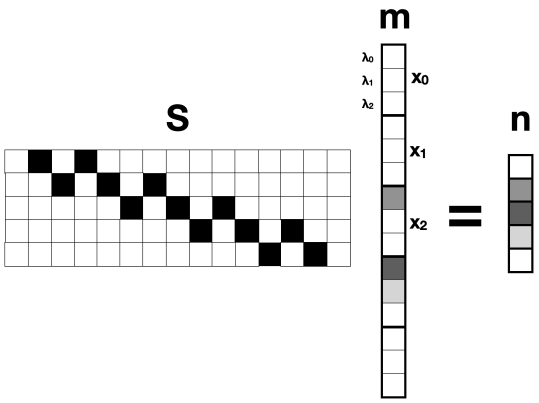


Fig. C.2 Illustration of ambiguous encoding given wavefront gradients.

In principle, there are infinitely many possibilities for input measurements that will yield the same coded shot **n**; this ambiguity must be removed to be able to perform an accurate measurement. The significance of the particular example of Fig. C.2, is that the phase shift does not contain large jumps relative to the wavelength. Thus, regularization

constraints such as a smoothness property are not enough to remove it from the solution space. However, one can remove this ambiguity by adding more views that contain different amounts of dispersion. For example, an additional view with dispersion opposite to that of the existing view can be incorporated into **S**. Fig. C.3 illustrates how the ambiguity is removed in this situation.

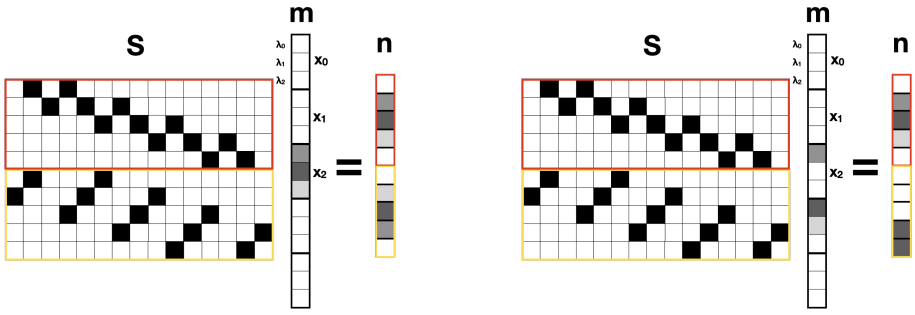


Fig. C.3 Resolved ambiguity given two different dispersed views. For each hypercube, **m**, the resulting measurement **n** is different.

In this work a diffraction grating is used to produce three copies of the microlens focii, each with a different level of dispersion, meaning that the system is actually over-determined, providing robustness against noise. Furthermore, we use a microlens array that is rotated by some angle, in order for the streaks in the ± 1 to not overlap.

C.2 Simulation of Polarization effects

In order to calculate the effects of polarization on the resulting intensities, including the effect of a birefringent medium and the projections of the field onto each of the polarization orientations of the sensor, Jones calculus is used. One starts with the vectorial field, where for simplicity the spatial and spectral coordinates are suppressed,

$$\mathbf{E} = \begin{bmatrix} \|E_x\| \cdot e^{i\phi_x} \\ \|E_y\| \cdot e^{i\phi_y} \end{bmatrix}. \tag{C.1}$$

The Jones matrix for the birefringent medium that causes a relative delay between the x and y polarizations is,

$$\mathbf{M}_b = \begin{bmatrix} e^{i\omega\tau} & 0 \\ 0 & 1 \end{bmatrix}, \tag{C.2}$$

which can be easily rotated by an angle η by applying rotation matrices around the Jones matrix ($\mathbf{R}(\eta)\mathbf{M}_b\mathbf{R}(-\eta)$, where,

$$\mathbf{R}(\eta) = \begin{bmatrix} \cos(\eta) & -\sin(\eta) \\ \sin(\eta) & \cos(\eta) \end{bmatrix} \tag{C.3}$$

Finally, the effect of a linear polarizer at angle θ is represented by,

$$\mathbf{M}_p(\theta) = \begin{bmatrix} \cos^2(\theta) & \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) & \sin^2(\theta) \end{bmatrix}. \tag{C.4}$$

The final intensity for a given birefringent medium orientation, η, and linear polariser pixel orientated at angle θ was:

$$I(\theta, \eta) = |\mathbf{M}_p(\theta) \cdot \mathbf{R}(\eta) \cdot \mathbf{M}_b \cdot \mathbf{R}(-\eta) \cdot \mathbf{E}|^2 \tag{C.5}$$

Using these equations, Mathematica was used to find the final fields.

Appendix D

Pictures from RAVEN Experiments

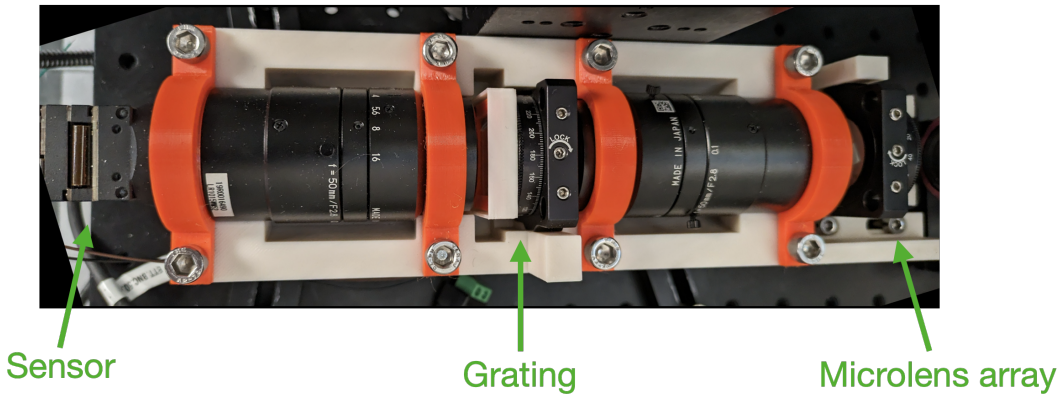


Fig. D.1 The RAVEN hardware. Note that if one wishes to resolve the fourth Stokes parameter, an additional chromatic birefringent medium should be added before the microlens array.

ATLAS experiment

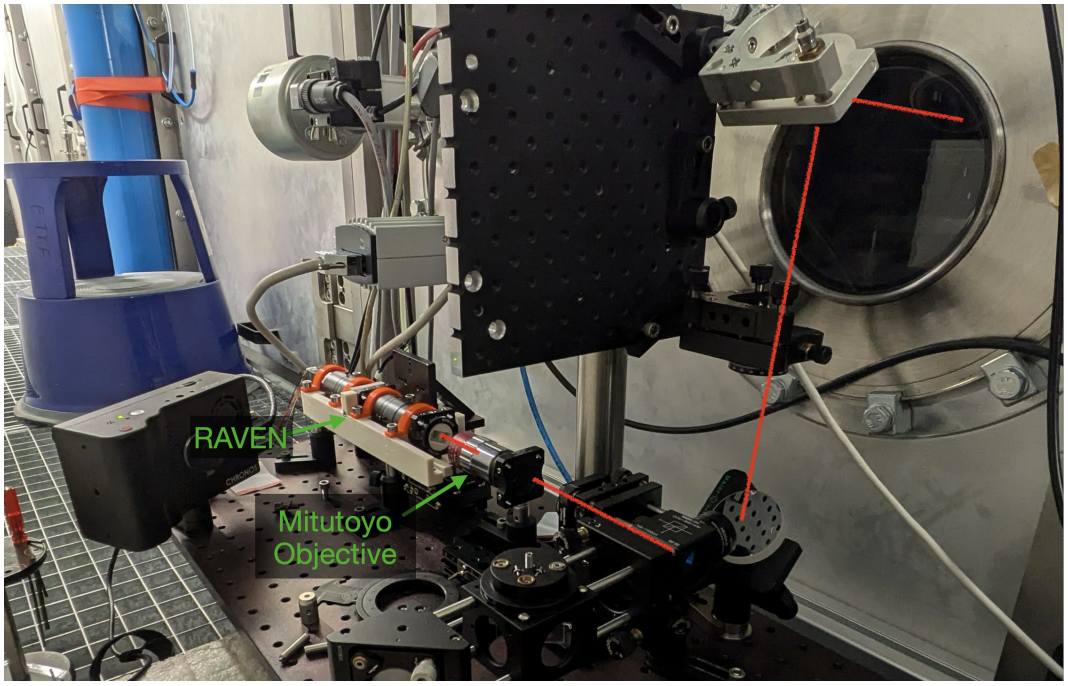


Fig. D.2 The installation of RAVEN at the Center for Advanced Laser Applications where measurements were performed on the ATLAS-3000 laser.

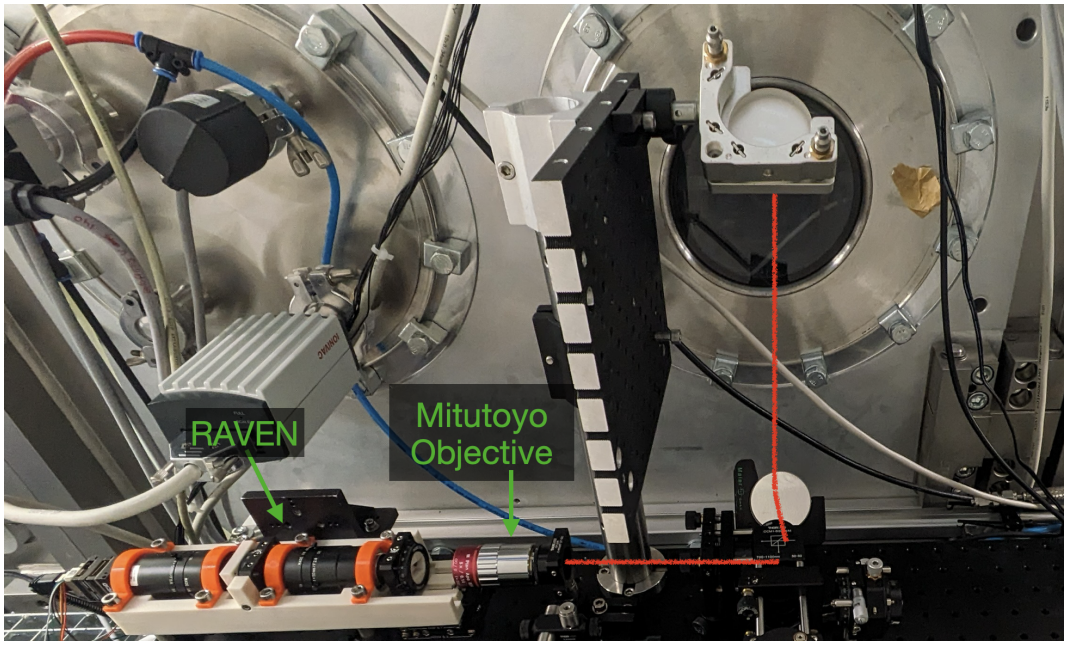


Fig. D.3 The installation of RAVEN at the Center for Advanced Laser Applications where measurements were performed on the ATLAS-3000 laser.

Vortex experiment

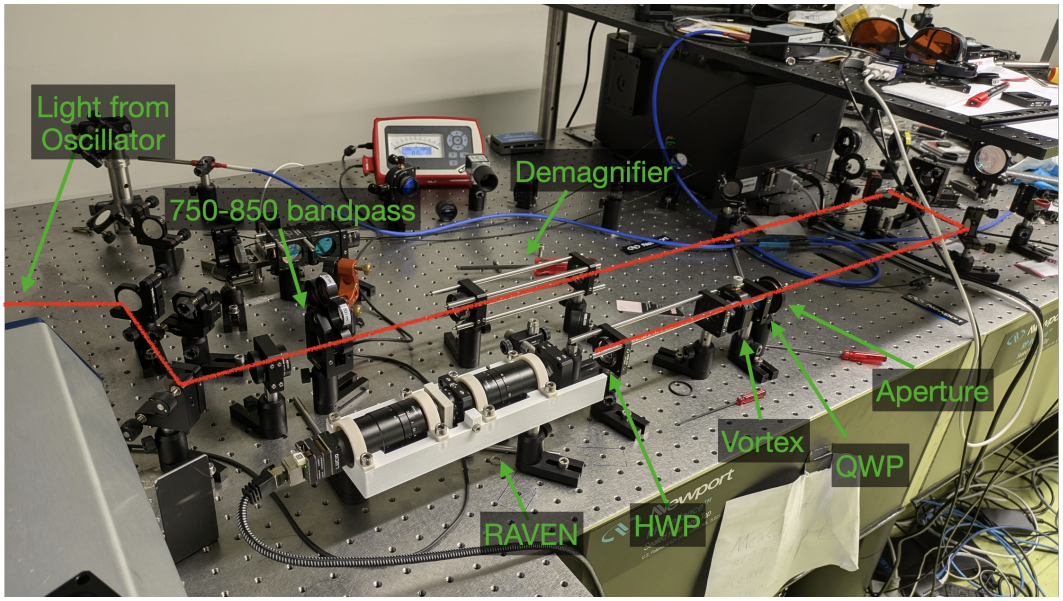


Fig. D.4 The experimental setup for the circularly polarised vortex measurement. Light from the oscillator goes through a bandpass filter to better approximate the ATLAS laser. The circularly polarised vortex is then formed by aid of a quarter waveplate (QWP) and a vortex retarder. A half waveplate is used as a birefringent medium to allow for the handedness to be detected.

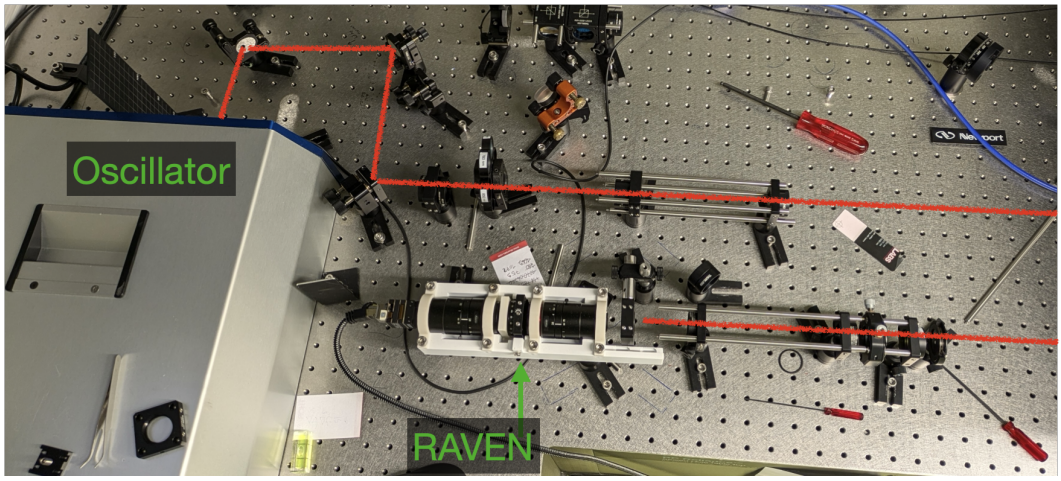


Fig. D.5 The experimental setup for the circularly polarised vortex measurement. Light from the oscillator goes through a bandpass filter to better approximate the ATLAS laser. The circularly polarised vortex is then formed by aid of a quarter waveplate (QWP) and a vortex retarder. A half waveplate is used as a birefringent medium to allow for the handedness to be detected.