Singularimetry: Phase measurement using vortices and caustics

---

Samuel A. Eastwood

BSc(Hons)

School of Physics and Astronomy, Monash University

February 27, 2015

Contents

List of Figures

Abstract

This thesis investigates singularimetry in both optical and matter wave fields. The utility of optical singularities for performing phase measurements was demonstrated using a three–beam interferometer. Three–wave interference was used to generate a uniform lattice of optical vortices, which was distorted by the presence of an object inserted into one arm of the interferometer. Using theoretical ideas from singular wave optics, a proportionality between the transverse displacement of the vortices and the phase shift in the object wave was derived and experimentally tested. Tracking the vortices permitted the phase of the object to be reconstructed. We demonstrated the method experimentally using a simple lens and a more complex object, namely the wing of a common house fly. Since the technique is implemented in real space, it is capable of reconstructing the phase locally.

By studying the extreme opposite of nodal singularities, an alternative phase retrieval technique was designed and numerically tested, which utilizes the natural intensity singularities of caustics. Using catastrophe theory, we showed that, given the formation of a fold caustic in the wave field, the functional form of the wave's phase may be expressed as a truncated Taylor series. We then outlined how all expansion coefficients in the series may be determined by quantifying unfolding of the caustic through focus, thus framing this otherwise ill–posed inverse problem into a well–posed one. The method was then successfully implemented on simulated data. Possibilities of extending the technique to higher order catastrophes are also discussed.

The relationship between caustics and vortices in matter waves was also investigated. By inducing aberrations in the magnetic lenses of a conventional transmission electron microscope, it is shown that diffraction catastrophes may be created in the electron beam. As a consequence of the duality of caustics and vortices—a central theme to the thesis—electron vortices were created in the electron beam in the vicinity of the caustics. To measure and explore the quantized anomalous Gouy shift of focused astigmatic electron waves, caustics were again generated by inducing aberrations in the lenses and the Gouy phase shift

determined by phase retrieval. Multiple theoretical descriptions of the Gouy phase anomaly are presented and discussed. It is shown that these various interpretations may be unified into a single theoretical framework.

Lastly, we investigated caustics in the context of order–parameter manifolds. We show that caustic surfaces also appear when a real or complex field is mapped to its order–parameter manifold. We exemplify these structures in the context of spin–$1/2$ fields, where the order–parameter manifold is the Bloch sphere. These generic structures are a manifestation of catastrophe theory and are stable with respect to perturbations. The corresponding field configurations are also stable and represent a new type of topological defect, which we call order–parameter catastrophe defects. Equations governing the conditions for the existence and unfolding of the defects are derived.

Disclosure

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other institution. To the best of my knowledge the thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis. Parts of this thesis are based on joint research or publications, the relative contributions of the respective authors being as follows.

The publication Eastwood et al. (2012) that Chap. 4 is based on has the following author contribution: S. A. Eastwood 70%, A. I. Bishop 15%, T. C. Petersen 5%, D. M. Paganin 5% and M. J. Morgan 5%.

Chapter 6 contains content from three separate publications. For §6.1, containing content from Petersen et al. (2013b), the author contributions are: T.C. Petersen 30%, M. Weyland 14%, D. M. Paganin 14%, T. P. Simula 14%, S. A. Eastwood 14%, M. J. Morgan 14%.

For §6.2, containing content from Petersen et al. (2013a), the author contributions are: T.C. Petersen 30%, D. M. Paganin 14%, M. Weyland 14%, T. P. Simula 14%, S. A. Eastwood 14% and M. J. Morgan 14%.

For §6.3, containing content from Petersen et al. (2014), the author contributions are: T.C. Petersen 30%, D. M. Paganin 14%, M. Weyland 14%, T. P. Simula 14%, S. A. Eastwood 14% and M. J. Morgan 14%.

Chapter 7 contains work from Eastwood et al. (2014) (included in Appendix C). The author contributions are: S. A. Eastwood 70%, D. M. Paganin 10%, T. C. Petersen 10% and M. J. Morgan 10%.

Samuel A. Eastwood

Prof. Michael J. Morgan

Prof. David M. Paganin

Dr. Timothy C. Petersen

eThesis Copyright Notices

Notice 1

Notice 2

Acknowledgements

Over the course of my PhD I have been extremely fortunate to have been the beneficiary of the best supervision I could have possibly hoped for. For this I am indebted to my three esteemed supervisors; Prof. Michael Morgan, Prof. David Paganin and Dr. Timothy Petersen. I am immensely grateful for the considerable time and effort they invested in me, for allowing and encouraging me to develop and pursue my own research ideas and and for their support and understanding during the more trying periods of my postgraduate study.

Specifically, I thank Michael for his sage advice on current research skills and practicalities. I will no doubt continue to employ his wisdom throughout the course of my future career. David, I thank for his endless optimism and enthusiasm regarding my research. His vast optics knowledge has been a tremendous resource for me since beginning my honours degree and for this I am extremely grateful. Last but not least, I wish to thank and acknowledge the huge space with grace Tim created for me. His availability for impromptu discussion was invaluable to me, as was his great conviction towards my work, even when I had lost mine. These generosities where crucial factors in helping me to complete my postgraduate studies.

I'd also like to sincerely thank Dr. Alexis Bishop whom I collaborated with on the vortex lattice interferometry and super–resolution imaging work. Not only did he design and build the experiment, but his input and expertise was pivotal in the success of the research.

I am grateful to the School of Physics & Astronomy for the many valuable resources they provided me with. I would particularly like to thank Jean Pettigrew, Julia Barmes and Robert Seefeld for the tireless work and effort they put into me and all postgraduate students in our department. I also wish to thank Theo Hughes, Jenny McCabe and Manuel Pumarol Crestar for their much appreciated assistance during my time as a third year demonstrator.

During my postgraduate studies, I have been extremely fortunate to have met and befriended some of the funniest and most entertaining people I've ever met. In no particular order, I'd like to thank Amelia Fraser–McKelvie, Jacob Crossett, Jeremy Brown, Jessica Crawshaw, Rhiannon Murrie, Katie Lee, Genevieve Buckly, Nathanael Lampe, Elliot Hutchi-

son, Martijn Jasperse, Andrew Leong, Chris Billington, Sam Fischer, Mithuna Yoganathan, Philip Starkey, Damien Przybylski, Shaun Johnstone, Lisa Bennie, Alexander Wood, Dane Kleiner, Mark Edmonds, Jack Hellerstedt and Jian Yen for all the fun, laughter and friendship.

Finally, I'd like to express my deepest thanks and gratitude to my family. To my brothers Alexander and Oliver, for their love, humour and support. To Mum and Dad in particular, I owe a great deal of thanks as I firmly believe it was their hand in nurturing my curiosity and love for learning that I chose to study physics. It was also through their emotional and financial support, that I was enabled to finish my studies. Thank you so very much.

1 Introduction

Singular optics involves the study of defects or singularities in light. However, what constitutes a singularity depends on the mathematical formalism we choose to model the behaviour of the optical wave field. Consider the situation shown in Fig. 1.1, where light passes through a perfect lens, which is focused to a point after one focal length. In geometrical optics we describe the light field as being comprised of a series of rays, which point in the direction of wave propagation. As the rays pass through the lens the changing refractive index causes a change in the angle of the ray, with the number density of the rays proportional to the intensity of the light. In this model, of a perfect lens that is illuminated by perfectly parallel incident rays, all rays must pass thought the focal point, resulting in an infinite intensity of light at the focus. Singularities, corresponding to infinite intensities in ray theory are known as caustics, with the above example constituting a point caustic.

Now consider the identical situation described in terms of a complex scalar wave equation. In this picture the focal point becomes "blurred" due to the optical analogue of the uncertainty principle and the finite wavelength of the light (Padgett, 2008). This results in a bright spot of non–zero width and finite intensity. Thus we see how the dramatic behaviour of a point caustic singularity in ray theory is softened, or smoothed, by the introduction of the wave formalism of light. Although singularities of infinite intensity do not appear in the wave theory of light, the transition from ray theory to modelling light as a complex–valued scalar wavefunction results in different types of singularities. When the real and imaginary parts of a complex scalar wavefunction vanish simultaneously, a point singularity known as a vortex arises in the phase of the wavefunction. In contrast to caustics, vortices are associated with zeros of intensity. Various other types of singularities can also occur in wave fields when

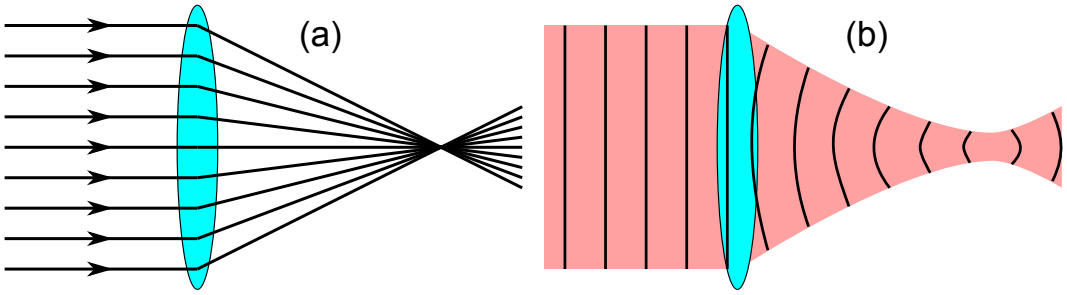

Figure 1.1: A beam of light is focused by a perfect lens. In the geometrical theory of light, light rays are refracted, causing every ray to pass through the focal point of the lens, as shown in (a). The density of rays at the focal point is infinite, giving rise to a point singularity of infinite intensity according to geometrical optics. Describing this situation in the context of the wave theory of light softens this infinitely intense focal point to a bright spot possessing a non–zero width, as shown in (b).

we extend our optical models to include the vector or particle nature of light. The way in which some singularities are "tamed" when described by another theory, resulting in other types of singularities, is referred to as the "singularity hierarchy" (Berry, 1998). By viewing optical phenomena at a scale where the quantum nature of light must be considered, vacuum fluctuations reveal that the amplitude at a vortex core does not in fact vanish (Barnett, 2008; Berry and Dennis, 2004). Vortices and caustics may be thought of as complementary, being singularities of light's darkness and brightness, serving to "bookend" the study of singular optics.

The work presented in this thesis provides a detailed analysis of the duality between caustics and vortices. However, rather than considering on the intrinsic nature of these singularities, which has already been studied extensively, this thesis concerns itself with the question: Can these singularities be exploited for their unique properties in order to measure or image an object or field? This concept of using the singularities of the field as the experimental measurement tool has been referred to as "Singularimetry" (Dennis and Götte, 2012, 2013). In the past this terminology has only been used in the context of the vortices or zeros of the wave field. However, here we use the term singularimetry as an umbrella term for any method that exploits any form of optical singularity as an experimental tool.

1.1 Vortices

Vortices are a ubiquitous aspect of the physical world. A vortex may be defined as the centre of the local rotation of a field. Examples of vortices appearing in the natural world include the swirling of water as it goes down a plug hole, at the eye of a cyclone, or the centre of our own spiral galaxy. Whilst these are all examples of dynamically rotating vortices, a screw type lattice defect in a crystal is also a vortex in the position of the atoms. Here the amount that a screw type dislocation can "rotate" is restricted to being an integer number of crystal monolayers. An optical vortex is a screw dislocation in the planes of constant phase of a complex scalar wavefunction. It shares similar properties with that of a crystal screw defect in that its circulation is also limited to quantised values. This analogue between optical vortices and crystal dislocations was used in the seminal paper by Nye and Berry (1974), which described vortices as dislocations of wave theory.

1.1.1 Definition of a vortex

Consider a time–independent complex scalar wavefunction given by

$$\Psi(\mathbf{x}) = \sqrt{I(\mathbf{x})}\exp(i\phi(\mathbf{x})), \tag{1.1}$$

where $\mathbf{x}$ is the position vector and I and ϕ are the intensity and phase of the wavefunction, respectively. Although nature demands Ψ be a continuous function, the oscillatory nature of the complex exponential means that there is no such restriction on ϕ. Discontinuities can therefore occur in ϕ as a consequence of its multivaluedness. It is this property that allows for the existence of phase vortices.

An optical vortex is a screw type dislocation in the phase of the complex–valued scalar wave field. To compute the circulation about a vortex, consider a smooth closed contour Γ, that encloses a vortex point. Along every point of Γ, by assumption, there will be non–vanishing intensity. The line integral along Γ of the phase gradient will be an integer multiple of 2π, i.e.,

$$\oint_\Gamma \mathrm{d}\phi = \oint_\Gamma \nabla\phi \cdot \mathbf{t}\,\mathrm{ds} = 2\pi n, \tag{1.2}$$

where $\mathbf{t}$ is the unit tangent vector to Γ, ds is the infinitesimal line element and the integer n is known as the topological charge. This integer represents the number of times the phase winds by 2π around the vortex core as the curve Γ is traversed; its sign specifies the

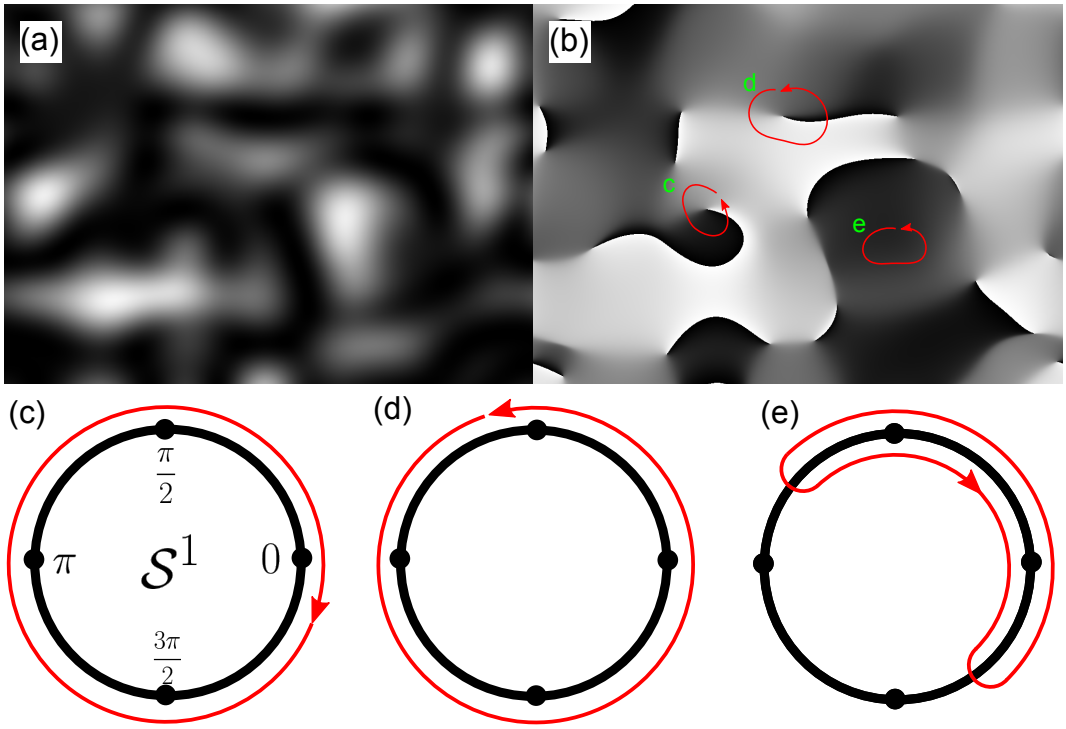

Figure 1.2: (a) The intensity of an optical speckle pattern. (b) The phase of the speckle pattern containing numerous vortices. Three contours are drawn in red in (b) and their order–parameter mappings (see §1.1.2) are given in (c)-(e). The two contours that encircle a vortex fully wrap the order–parameter space (a 1D sphere, denoted by $\mathcal{S}^1$). However, the contour in (e) only partially covers $\mathcal{S}^1$ and is thus equivalent to a single point in the order–parameter space.

rotational direction of the phase winding, with anti–clockwise conventionally denoting a positive topological charge. An example of the intensity and phase of a speckle field is shown in Fig. 1.2. Several vortices can be observed in the phase of the wave with a vortex and anti–vortex circled in red by the contours labelled ‘c’ and ‘d’, respectively. The abundance of vortices within this speckle field highlights that vortices are a generic phenomenon in such fields.

Vortices always occur at points of zero intensity of a complex scalar wave field. Consider a complex wavefunction represented as a phase vector (phasor) in the complex plane with the magnitude of the wave amplitude and the angle equal to the phase. At points where the real and imaginary parts vanish the phasor has zero length, becoming a single point at the origin. There is no way to define a phase angle of the phasor when this occurs, leading to a phase ambiguity in the complex function. This construction demands that an intensity zero is always located at a vortex point; however, the converse statement, that a vortex always occurs

at an intensity zero, is not universally true. This is similar to the phase ambiguity in the azimuthal angle at either pole of a sphere when parameterized in spherical polar coordinates.

1.1.2 Stability

An important property of a vortex is that it is a stable defect of a field, as it cannot be eliminated by any continuous perturbation. From a topological perspective, we can view the phase as a function that assigns every point in space (where the wave is defined), a value between 0 and 2π. The phase may be regarded as an order–parameter of the field, with all its possible values constituting an order–parameter space (Mermin, 1979). In the present example, the phase defines an order–parameter space, which is a circle, denoted by $\mathcal{S}^1$ (see Figs. 1.2(c)–(e)). As we traverse the closed path Γ we take the value of ϕ at each point on the path and map it to the corresponding value on $\mathcal{S}^1$. If Γ represents contour "c" in Fig. 1.2(b), which encloses a vortex, then the value of each point on Γ samples the full range of values of ϕ. Since all values of $\mathcal{S}^1$ are sampled, Γ topologically wraps the entire order–parameter space. Alternatively, the Γ' contour "e" in Fig. 1.2 does not enclose a vortex because the line integral given by Eq. (1.2) is zero, as the beginning and end points of Γ' are equal. The values along Γ' only form a subset of $\mathcal{S}^1$; the order–parameter space is therefore only partially covered by Γ, as shown in Fig. 1.2. Suppose we make a smooth and continuous transformation to Γ', such that we reduce its length but maintain a closed loop. The covering of $\mathcal{S}^1$ is continuously deformed in this process. Eventually Γ' becomes infinitesimally small corresponding to only a single point on $\mathcal{S}^1$. Therefore, the covering of the order–parameter space in Fig. 1.2(e) is topologically equivalent to a single point on $\mathcal{S}^1$. However, the values along Γ completely cover $\mathcal{S}^1$; as we smoothly deform Γ we find that it cannot be reduced to a point as there is no way to "unwrap" its covering on $\mathcal{S}^1$. This illustrates that a vortex point is topologically distinct from other non–vortical points of the field. The fact that we cannot unwrap the phase around a vortex core in $\mathcal{S}^1$ implies that the vortex point cannot be transformed away by any continuous deformations.

1.1.3 Higher order vortices

Higher order vortices with topological charges $|n| > 1$ do not share this structural stability. To illustrate this point, consider a simple wavefunction containing an nth order vortex at the

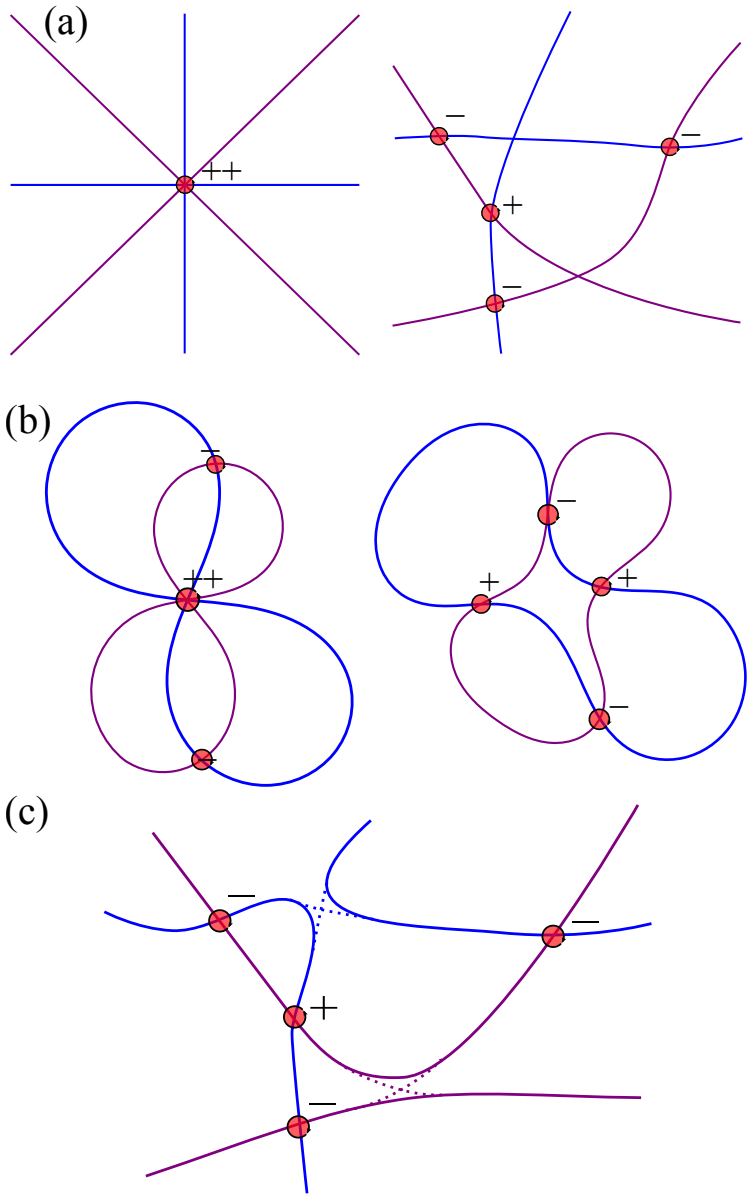

Figure 1.3: Decay of a doubly charged optical vortex. Here the purple and blue curves represent where $\mathrm{Re}[\Psi(x,y)] = 0$ and $\mathrm{Im}[\Psi(x,y)] = 0$, respectively. (a) A critical point explosion of a double charged vortex. The point where all four curves intersect is highly unstable. A small perturbation lifts this single point of intersection resulting in four points of intersection. Each intersection corresponds to a first–order vortex. Thus the second–order vortex decays into one vortex and three anti–vortices. (b) Alternative decay path of a second order vortex. The curves of vanishing real and imaginary parts form closed loops when they are perturbed in such a way that they no longer self–intersect. In this situation the second–order vortex can decay into two first–order vortices. (c) The decayed second–order vortex shown in (a), which has been perturbed further removing the intersections between the curves (marked by the dashed lines), lifts the ambiguity in the vortex sign. Figure adapted from Freund (1999) and Paganin (2006).

origin that is given by a polynomial solution of the d'Alembert wave equation:

$$\Psi(x, y, z, t) = (x \pm iy)^n e^{i(kz-\omega t)}. \tag{1.3}$$

Here k is the magnitude of the wave–vector and ω is the angular frequency of the wave field. Such polynomial solutions have been described as pure screw dislocations by Nye and Berry (1974). For simplicity let $z = t = 0$ and $n = 1$, giving $\Psi(x, y) = x \pm iy$. The lines corresponding to the vanishing of the real and imaginary parts in this case are $x = 0$ and $y = 0$. Applying smooth infinitesimal perturbations to either curves using the real numbers ϵ_1 and ϵ_2, the wavefunction becomes $\Psi(x, y) = x + \epsilon_1 + i(y + \epsilon_2)$, which shifts the vortex to the point (ϵ_1, ϵ_2). Thus a vortex of $n = 1$ is only shifted by a perturbation. If $n = 2$, the wavefunction is $\Psi(x, y) = x^2 - y^2 + i2xy$ and the curves describing where $\mathrm{Re}[\Psi(x, y)] = \mathrm{Im}[\Psi(x, y)] = 0$ are now $y = \pm x$ and $x = y = 0$. Their intersection at the origin is shown by the second–order vortex in Fig. 1.3(a). Any small perturbation to the real or imaginary parts of Ψ causes the single point of intersection between all four curves to bifurcate into two points of intersection. Because the total topological charge of a wave field is a conserved quantity, the $n = 2$ vortex will decay into two first–order vortices. The instability of higher order vortices holds for vortices of any topological charge with magnitude greater than 1. This bifurcation of a higher order vortex into multiple first–order vortices has been called a "critical point explosion" by Freund (1999).

Considering vortex points as intersections of the curves $\mathrm{Re}[\Psi(x, y)] = 0$ and $\mathrm{Im}[\Psi(x, y)] = 0$, and upon traversing these curves, Freund and Shvartsman (1994) showed that the sign of adjacent vortices situated on the same line must alternate. This characteristic is known as the sign principle. The major implication of this result is that, given a field containing multiple vortices, and knowing the sign of a single vortex, we immediately know the sign of all other vortices in the field. This is demonstrated in Fig. 1.3(b), which shows an alternative decay path of a second–order vortex. In this situation the higher order vortex is perturbed such that the self–intersections of its curves of vanishing $\mathrm{Re}[\Psi]$ and $\mathrm{Im}[\Psi]$ are removed. Instead of decaying into four first–order vortices, it splits into two first–order vortices of the same sign, which preserves the topological charge of the system. By following either curve, $\mathrm{Re}[\Psi(x, y)] = 0$ or $\mathrm{Im}[\Psi(x, y)] = 0$, it is observed that the sign of each vortex alternates at every point of intersection between the two curves, thus satisfying the sign rule.

Note the ambiguity in applying the sign rule for the case shown in Fig. 1.3(a). This is because the curves $\mathrm{Re}[\Psi(x,y)] = 0$ or $\mathrm{Im}[\Psi(x,y)] = 0$ were perturbed in such a way that their points of self–intersection were not removed. At these points of intersection, it is unclear in which direction each curve continues. An additional perturbation must be applied to the situation in Fig. 1.3(a) in order to remove the self–intersections and therefore lift the ambiguity in the sign rule. An example of the resultant curves after the application of such a perturbation is shown in Fig. 1.3(c). Here, all curves of vanishing real and imaginary parts can be traced without ambiguity, and it is now apparent how the sign rule relates to the initial situation in Fig, 1.3(a). This demonstrates how self–intersection of these curves introduces ambiguities when applying the sign rule.

1.1.4 Nodal lines, nucleation and annihilation

In two spatial dimensions, the curves of vanishing real and imaginary parts of the field trace out a network, with vortices occurring at the intersection of these curves (see Fig. 1.4). In three dimensions the lines defining $\mathrm{Re}[\Psi(x,y)] = 0$ and $\mathrm{Im}[\Psi(x,y)] = 0$ become planes, with their intersections defining nodal lines, i.e., the two–dimensional representation of a vortex. Nodal lines possess the same topological protection that vortices enjoy, so that when present in a wave field nodal lines must persist as a wave propagates. Nodal lines can therefore persist as a wave propagates to $z \to \infty$; however, they can also terminate at the edge of a boundary, such as an opaque object. Nodal lines may also form closed loops in space. A closed nodal loop is a result of the nucleation of a vortex/anti–vortex pair, which later comes together to annihilate at some point as the wave field propagates. Berry (1998) emphasized that the nodal lines of the wave field are more fundamental than vortices. We are free to choose any arbitrary plane to observe a wave field, whence the notion of a single vortex nucleation point then becomes dependant on our image plane. However, the geometry of the nodal lines is defined for all three spatial dimensions, no matter where a detector is placed; the geometry would be found to be invariant. We must therefore regard vortices as points of intersection between an image plane and the more fundamental structure described by the nodal line.

Closed nodal lines can have interesting topology. O'Holleran et al. (2009a) showed in a numerical study of an optical speckle field that two nodal loops can be threaded together.

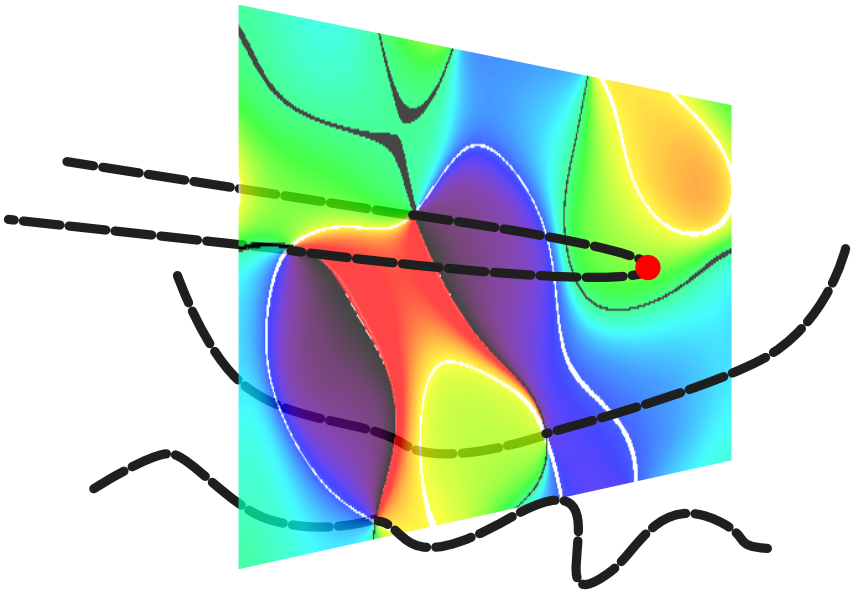

Figure 1.4: A two–dimensional image of the phase of a complex scalar wave field containing vortices. The lines where the real and imaginary parts of the wavefunction vanish in three dimensions are drawn in black and white, respectively. Vortices are located at all intersections of these curves. In three dimensions these curves are two–dimensional surfaces with their intersections tracing out the nodal lines (dashed lines). The nodal lines may extend indefinitely, or form closed loops, corresponding to the creation/annihilation of a vortex/anti–vortex pair.

They also observed the occurrence of multiple loops linking together to form long chains of nodal loops. These instances were shown to be generic features of the field, rather than special cases, and the probability of a nodal loop not being threaded exponentially decreased for larger loops. It was also noted that nodal lines in such speckle fields possess similar fractal properties to that of Brownian motion, with the size distribution of loops in optical speckle fields exhibiting scale invariance (O'Holleran et al., 2008). Similar to how a long piece of string is prone to become knotted, the authors expected nodal lines to form knots at large length scales. Surprisingly, no knotted nodal structures were observed in their simulations. Dennis et al. (2010) later showed that knotted nodal structures are still possible. Complex fields with knotted nodal lines were found as special solutions to the paraxial wave equation. This theoretical insight allowed Dennis et al. (2010) to derive optical masks using spatial light modulators, which produced knotted nodal lines in a physical wave field.

1.2 Caustics

Consider now the other extreme of optical behaviour—the singularities of infinite intensity. Caustics are singularities in the ray picture of light, in which we take a "zoomed out" perspective of light, to a scale where the wavelength becomes negligible. At this scale the finite detail in our optical field, such as inference effects, focal width or vortex dislocations, cannot be resolved. Recall our description of a point caustic given in the introduction, namely a point at which all rays of a focused beam intersect. This isolated point caustic isn't generic in the sense that point foci rarely arise in nature, since lenses are engineered optical elements specifically designed to create a point focus. Point caustics are also unstable to perturbations e.g., aberrating a lens would cause the rays to no longer pass though a common point. A more natural situation would be the focal volume of an aberrated lens, such as a drop of water or rays reflected off a warped surface. The focus of this lens would no longer be a point, but rather smeared out over some length, with the envelope of the rays forming a caustic surface as shown in Fig. 1.5(a). The caustic surface is comprised of the locus of the family of rays; caustics of this type are stable. Perturbing the lens in Fig. 1.5(a) alters the ray trajectories, but they must still maintain a point of intersection within the ray family; this causes deformation of the caustic surface, but does not destroy it.

A caustic surface in the volume of an aberrated lens is more generic than a point caustic, meaning a point caustic has zero probability of forming compared to a caustic surface. This arises in many natural instances. Typical examples include the bright shimmering pattern of caustics at the bottom of a swimming pool, or the cusp caustic formed from the reflection of light inside a coffee mug. Caustics also appear at the cosmological scale, as the space–time curvature caused by a galaxy or star that bends the light around it. These constitute cosmologically massive aberrated gravitational lenses which also focus light, yielding caustics (Ohanian, 1983). Such a large scale caustic has been observed by Elliot et al. (1977) during an occultation of a star by Mars, which caused a shadow to sweep across the Earth's surface. The atmosphere of Mars refracts star light to such a degree that it behaved as a aberrated spherical lens, giving rise to a Mars–sized caustic at the surface of the Earth. (Berry, 1981). A naturally occurring caustic formed by light refracted by a water bottle is shown in Fig. 1.5(b). It is the ubiquity of structurally stable caustics that lead Nye (1999) to used the term natural focussing in reference to the creation of such generic caustics,

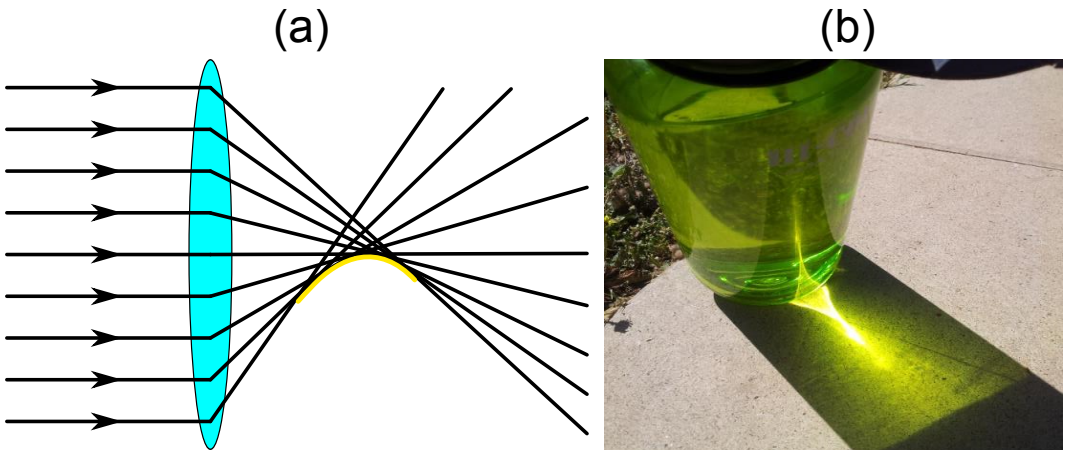

Figure 1.5: (a) A caustic envelope of a family of rays, shown in yellow, produced in the focal volume of an aberrated lens. Here the focal length of the lens increases from bottom to top. (b) The natural focusing of light by a plastic water bottle generating a cusp caustic.

as opposed to the unnatural case of a point caustic.

The profusion of caustics is also not limited to optics. Propagating ripples in the fabric of space–time, known as gravitational waves, are also subject to natural focusing. As a gravitational wave propagates, small local contractions and expansions of space–time perturb the wave, resulting in caustics. Bondi and Pirani (1989) described caustics of gravitational waves as causing all test particles to come together and collide, also forming a caustic surface of test particles. Another example is a rogue wave in the ocean, which is characterized as a wave having an amplitude much greater than the background ocean. Metzger et al. (2014) showed that rogue waves are a result of focusing water waves, analogous to focusing from an aberrated lens. Diffraction integrals were used to model this behaviour. Conversely, this link between water and optical waves has motivated the creation of optical analogues of rogue waves (Solli et al., 2007). These examples illustrate how initial weak fluctuations can result in very extreme events given enough time or propagation distance. This is the key principle behind caustic formation. Caustics also arise in the Argand–plane representation of a complex scalar wavefunction (Rothschild et al., 2012, 2014). These Argand–plane caustics were shown to occur due to the vorticity of the complex–valued scalar field changing sign. Caustics therefore arise from having at least one vortex and anti–vortex pair in a wavefunction. This further highlights the duality between vortices and caustics.

1.2.1 Introduction to catastrophe theory

Caustic surfaces are ubiquitous in naturally focused or reflected light, yet only ever appear in a small number of particular geometries. This is because caustics are a manifestation of topologies that originate from catastrophe theory, which describes the structural stability of physical systems. Catastrophe theory stems from early work in structural stability by Thom (1975, 1977, 1983), which was later developed into catastrophe theory by Arnol'd (1975). It was made popular by Zeeman (1977), who applied it to a diverse range of physical systems. Berry (1976) was the first to use catastrophe theory in optics, and elucidated the concept of "diffraction catastrophes".

Instability is a consequence of degenerate critical points in a system's potential function, which describes the energy of a system in any given state and therefore determines a system's behaviour. If a system's potential function contains a degenerate critical point, small perturbations can lead to bifurcations in the potential, which cause an abrupt qualitative change in the behaviour of the system. For example, consider the simple function $V(x) = x^3 + ax$, where $a \in \mathbb{R}$ represents the degree of perturbation to the degenerate critical point x^3. For $a \geq 0$, $V(x)$ has a point of inflection at $x = 0$. However, for $a < 0$ the critical point splits into a maximum and minimum, leading to a bifurcation at $a = 0$, hence the system is governed by a potential function $V(x)$, which exhibits an abrupt change in its behaviour/dynamics. For higher order degenerate points, additional perturbation terms are needed to classify the ways in which a critical point may bifurcate. The set of points that describe the occurrence of bifurcations then define a bifurcation set of the potential function. Depending on the number of different bifurcations that can occur in a given potential function, the bifurcation set can be a curve, surface or even a hyper–surface. Catastrophe theory is the study of degenerate critical points and their bifurcations, leading to the classification of all types of critical points with their corresponding bifurcation sets. It is the geometry of the bifurcation sets that we observe as caustic surfaces in optics. Moreover, catastrophe theory shows that the bifurcation sets are stable to perturbations, which is also the origin of the structural stability of caustic surfaces and perhaps also of their generality in nature. Catastrophe theory is a central theme of this thesis and a detailed account is presented in Chap. 3.

1.2.2 Catastrophe optics

We have presented caustics as singularities arising in the geometrical theory of light, being surfaces of infinite intensity due to the infinite density of rays along the locus of a family of rays. We do not of course observe an infinite intensity. To better understand the physics of caustics, we need to account for the non–zero wavelength of light, as well as wave superposition. Caustics formed by coherent wave fields are known as diffraction catastrophes (Berry and Upstill, 1980; Berry and Howls, 2006).

For coherent fields, we can no longer regard intensity peaks of a diffraction catastrophe as the envelope of a family of rays. Instead the caustic surface results from focusing of optical intensity at these points, the caustic surface defining where the wavefront is collapsing most rapidly. The caustic surface becomes decorated by diffraction detail as a result of interference. Ultimately, adjacent regions of bright intensity are accompanied by lines of perfect destructive interference associated with dislocations in the wavefront. An intuitive reason for this duality is that as optical energy is focused, creating a caustic surface, there must be regions where optical energy density is decreasing. This creates regions of local intensity minima, which ultimately become intensity zeros after sufficient propagation. Because of this, caustics typically are found to coexist with amplitude zeros; consequently, the phase structure surrounding caustic surfaces are generally littered with vortex singularities. In essence, vortices form a "singular skeleton" that stabilizes and fills the voids of a caustic network body. This demonstrates how both singularities of amplitude zeros and of infinite intensity are inextricably linked (Angelsky et al., 2004).

All caustics may be classified as being part of a small number of particular geometries. There are five different geometries for catastrophes that we observe in full, as there are only four bifurcation sets with a dimensionality[1] of less than 3 in catastrophe theory. The evocative nomenclature for all five are: fold, cusp, swallow tail, elliptic umbilic and hyperbolic umbilic catastrophes. Figure 1.6 shows examples of the intensity and phase for some of these catastrophes. Almost all caustic surfaces will exhibit one of these five structures. Diffraction catastrophes are no less abundant than caustic surfaces for coherent wave fields. They have also been created and studied in electron waves (Petersen et al., 2013b) and non–linear

[1]The dimensionality of the caustic refer to the number of variables required to fully define the caustic surface, not the number of spatial dimensions in which it is embedded. This is known as the codimension of the caustic as explained in Chap. 3.

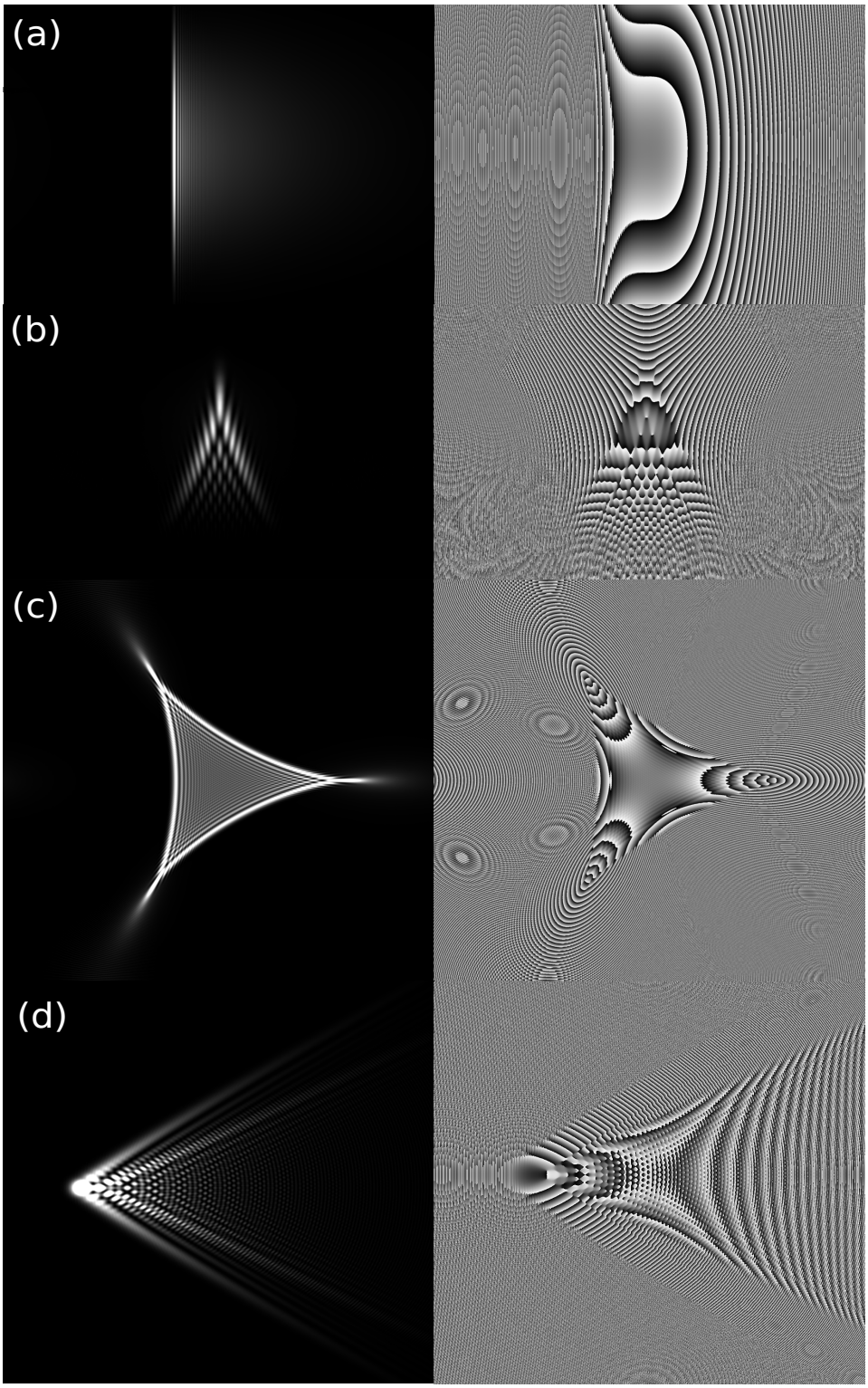

Figure 1.6: Intensity (left column) and phase (right column) of the various diffraction catastrophes: (a) fold, (b) cusp, (c) elliptic umbilic and (d) hyperbolic umbilic.

fields, such as Bose–Einstein condensates (Simula et al., 2013). The phase structure of these diffraction catastrophes once again highlights the complementarity between caustics and phase dislocations. A lattice of vortices can be observed in the phase of the cusp, elliptic umbilic and hyperbolic umbilic catastrophes. These vortices are produced by the interference between three wavefronts within the caustic surface (Pearcey, 1946; Berry, 1976). The generation of optical vortices via three–wave interference is presented in detail in Chaps. 2 and 4. This shows an important qualitative distinction between the interference of three and four waves, which we will revisit in §2.2. The phase of the fold catastrophe doesn't exhibit any vortices as its variation is only along one dimension; the phase singularity of the fold is instead an edge dislocation at the fold's locus. The fold catastrophe is also a diffraction free "Airy beam" (Berry and Balazs, 1979) and these edge dislocations are related to the Gouy phase (Pang et al., 2011). The relationship between the Gouy phase and caustics formed by electron beams is discussed further in §6.2.

1.3 Overview of the Thesis

The remainder of the thesis is structured as follows. In Chap. 2 we review the literature related to vortices, concentrating specifically on the methods of vortex generation in various contexts of wave physics. Detailed theory and calculation on the formation of optical vortex lattices by the interference arising from three–wave superposition then follows. This section is included to serve as a theoretical background to the content of Chap. 4, which exploits optical vortices to perform phase measurement. Lastly, a review of singularimetry techniques using optical vortex lattices for phase measurement and super–resolution is given.

Chapter 3 gives an introduction to the mathematical theory behind caustics; namely, catastrophe theory. The foundational concepts of catastrophe theory are presented independently of its relationship to optics, with the focus being on the structural stability of families of functions. We choose to present it in this way as the content of Chap. 5 and Chap. 7 relies heavily on these mathematical principles. Catastrophe theory in the context of diffraction physics is then presented, highlighting how the concepts of structural stability and diffraction physics can be related to one another. These ideas underpin the analysis presented in Chap. 6.

Chapter 4 presents a novel singularimetry technique that uses vortices. A three–arm interferometer is used to generate a lattice of optical vortices. A holography–type experiment

is conducted with an object in one of the three interferometer arms perturbing the vortex lattice. We describe the theory of how the phase shift in the wavefront from phase and phase–amplitude objects may be locally retrieved from the positions of each vortex, using an algebraic equation. Experimental results of the measurement of a simple spherical lens and the wing of a common house fly are also presented.

In Chap. 5 we outline a singularimetry technique using caustics. This technique represents a complementary method to the work presented in Chap. 4, highlighting once again the duality between caustics and vortices emphasised throughout the thesis. In this chapter we consider the ill–posed inverse problem of retrieving the phase of a complex scalar wavefunction, given that a fold caustic is formed at its aberrated focus. We outline a method for transforming this ill–posed inverse problem into a well–posed inverse problem, which possesses a solution. Using key theorems of catastrophe theory, discussed in Chap. 3, it is shown that the phase function of the wavefront may be represented as a truncated Taylor series. By measuring the variation of the caustic near the focus, all the coefficients of this Taylor series may be determined. The prospect of extending this technique to higher order caustics is also discussed.

Chapter 6 revisits the complementarity between caustics and phase defects, namely vortices and the Gouy phase anomaly. This chapter consists of two parts. The first presents experimental results of the generation of vortices in an electron beam. Aberrations were introduced to the lenses of a transmission electron microscope (TEM), which induces diffraction catastrophes within the microscope column. Iterative phase retrieval was performed on the diffraction catastrophe, which showed that electron vortices had indeed formed. The second part of this chapter studies the relationship between caustics and the Gouy phase anomaly in matter waves. By deliberately inducing astigmatism into the lenses of a TEM, the Gouy phase shift is directly measured across the two line foci caused by the aberrated lens. Various theoretical interpretations of the Gouy phase anomaly are also discussed. It is shown that these interpretations may be unified into a single theoretical framework.

In Chapter 7 we investigate caustics in abstract order–parameter spaces. Here we show that when a field is mapped to its associated order–parameter manifold, singularities of the mapping manifest as caustics. Due to the topological stability of caustics, the corresponding parts of the field must therefore also be topologically stable. These order–parameter caustics

represent a new form of topological defect that may occur in fields. This chapter studies these structures in the context of a spin–1/2 system, for which the order–parameter space corresponds to the Bloch sphere. Facilitated by catastrophe theory, equations governing the conditions under which these order–parameter caustics form are derived and we demonstrate how this phenomenon represents a gradient mapping.

Finally, in Chapter 8 we summarise the key findings of the thesis and discuss future directions related to the major themes of the thesis.

2 Generation and Applications of Vortices

The work in this thesis was motivated by the recent experimental demonstration of vortical electron beams. To contextualize this research we will give a brief history of vortex production in wave fields. Experimental methods of vortex generation share much in common across a wide variety of physical systems that exhibit wave singularities. For electron vortices, our simulations of vortex production in the electron microscope show some subtle features, which are described in the current chapter. These insights lead to an exploration of vortex production using three–wave interference. A detailed description of the mechanisms of vortex production is also described. The resulting formalism serves as a theoretical background to the work in Chap. 4.

2.1 Generation of Vortices

2.1.1 Optical phase singularities

The earliest occurrence of phase vortices in waves can be attributed to Whewell (1833, 1836), and relate to observations of the dynamics of ocean tides. The periodic nature of the tides mean that they can be modelled by the real part of an oscillating complex function. The phase contours of these functions are known as co–tidal lines and correspond to points of high tide. In studying co–tidal line contour maps, Whewell postulated the existence of points of zero amplitude in the complex harmonic oscillation of the tide; points where the co–tidal lines would circulate giving vorticity to the tidal phase. Such features are known as amphidromic points and their discovery ushered in the birth of phase singularities in wave theory (Berry, 2000). However, examples of phase singularities in optics were not explicitly considered

until the 1950's. The first to appreciate the existence of optical vortices was Sommerfeld (1954); however, his work was based on earlier lectures in 1950. The first experimental observation of optical vortices was in the context of total reflection of light at an air–glass interface. Wolter (1950) noted that intensity zeros coincided with optical phase singularities, which developed above the interface near the point of reflection.[1] Soon after, Braunbek (1951) noted that optical phase dislocations could result from the interference of just three plane waves. However, Pearcey (1946) pre–empted these findings in 1946, within the context of generic radio wave interference patterns. Three–wave interference will be discussed in detail in Chap. 4.

In the early years of singular optics, vortices were considered as an emergent phenomenon observed in certain optical setups. Vortices were later observed in fringe patterns produced by wavefront division by Vaughan and Willetts (1979). The helical wavefronts in this instance were produced by the interference of two waves. Vaughan and Willetts (1979) showed that, by slightly displacing the superposed waves relative to each other and laterally inverting a single wave, intensity variations transverse to the fringes occurred. This additional variation in intensity in a diagonal direction to the fringes was sufficient to promote vortex generation. Refined optical experiments, utilizing three–wave interference to produce uniform vortex lattices in a highly controlled manner, have subsequently been investigated (Masajada and Dubik, 2001; Dreischuh et al., 2002; O'Holleran et al., 2006; Vyas and Senthilkumaran, 2007; Kurzynowski and Borwińska, 2007; Masajada et al., 2007). Vortex formation utilizing interference of three–waves will be discussed in detail in §2.2.

Singularities produced from the superposition of three complex functions also occur in other optical settings. Phase singularities, called coherence vortices (Schouten et al., 2003; Gbur and Visser, 2003), may exist in the complex–valued spectral degree of coherence (Born and Wolf, 1999), which describes the correlation between two spatial coordinates of a partially coherent wavefunction. Similar to the way vortices nucleate from three–wave interference, Marasinghe et al. (2011) showed that coherence vortices also result from Mie scattering of partially coherent light from three dielectric spheres (Bohren and Huffman, 1983). These authors go on to indicate that coherence vortices would result from a partially coherent wave scattering from a random phase screen, which suggests that coherence vortices

[1] See Wolter (2009) for an English translation of this paper.

are as common a phenomenon as optical vortices. Coherence vortices are associated with pairs of points in a wavefunction that are completely uncorrelated, since the amplitude of the coherence function must be zero at this pair of points.[2] Physically, the consequence of zero correlation means that the visibility of the interference between two uncorrelated points is zero. The existence of a coherence vortex can be established by calculating the circulation of the phase in the complex degree of coherence, using a generalization of Eq. (1.2), whereby the angular frequency and one set of spatial coordinates are fixed on the contour. The topological charge of the coherence vortex is equal to the number of oscillations that the measured visibility undergoes, as this contour is traversed (Marasinghe et al., 2010). Coherence vortices possess a higher dimensionality than phase singularities of a coherent wave field. Since the spectral degree of coherence is defined for a pair of points, coherence vortices define a 4–dimensional nodal hyper–surface embedded in 6 dimensions, rather than a 1–dimensional nodal line embedded in 3 dimensions. So for a given angular frequency, a one–dimensional (1D) partially coherent field therefore defines a two–dimensional (2D) coherence function, which has sufficient dimensionality for vortices to form. This lead Simula and Paganin (2011) to investigate the counter intuitive concept of coherence vortices in one spatial dimension. The scattering of a 1D partially coherent (i.e., mixed state) wavefunction from a step potential was considered. The incident, reflected and transmitted waves generated a uniform vortex/anti–vortex lattice of coherence vortices in an almost identical manner to the superposition of three coherent plane waves.

Beyond using wave interference to generate vortical structures, both amplitude and phase masks have been developed for the purpose of producing isolated vortices. For example, vortices can be generated holographically. In polar coordinates the polynomial vortex given in Eq. (1.3) has the form $\exp(\pm ni\theta)$. For an off–axis holographic setup, a single vortex beam interferes with a tilted reference beam, given by $\exp[i(k_x x + k_z z)]$; at the plane $z = 0$, the corresponding intensity pattern is:

$$\begin{aligned} I(x,y) &= |\exp(\pm in\theta) + \exp(ik_x x)|^2 \\ &= 2 + \exp[i(\pm n\theta - k_x x)] + \exp[-i(\pm n\theta - k_x x)] \\ &= 2 + 2\cos[k_x x \mp n\tan^{-1}(y/x)]. \end{aligned} \tag{2.1}$$

[2]Note that this does not mean that two separate points of the coherence function vanish. This terminology refers to the fact that the coherence function is a function of pairs of points of a wave field.

When x is small, the magnitude of the inverse tangent in the argument of Eq. (2.1) abruptly changes at $y = 0$. Therefore, near the origin, the fringe pattern undergoes a phase change of $n\pi$ in the y direction. This yields a fork/edge dislocation in the fringe pattern, with the integer n being the degree of the dislocation, which is illustrated in Fig. 2.1(a). This highlights how optical vortices are analogous to crystal–like dislocations, as emphasized by Nye and Berry (1974). Alternatively, an amplitude mask with a transmission function identical to Eq. (2.1) would yield three beams in the diffraction pattern, corresponding to each term in the second line of Eq. (2.1) (He et al., 1995); namely, a central beam and two vortex beams with topological charge ±1 are contained in the first diffraction order. Bazhenov et al. (1990, 1992) were the first to perform experiments using this holographic approach to vortex production. First and second order vortices were also produced in this way by fabricating forked and double–forked diffraction gratings. Due to the difficulty of making a mask with the exact functional form specified by Eq. (2.1), computer generated diffraction gratings consisting of a binary approximation were used to generate vortices. Caustics and diffraction catastrophes have also been generated using diffraction gratings (Lee, 1983a,b).

Spiral Fresnel zone plates are another holographic method for producing vortices. Conventional zone plates consist of a series of concentric circles arranged in such a way that light is diffracted to a point. Heckenberg et al. (1992) showed that if the rotational symmetry of the zone plate is broken, then the diffracted light could be focused into a helical beam. It was shown that the required boundary between zones of the binary mask is given by:

$$\pm\, n\theta = \pi/2 + Ar^2, \tag{2.2}$$

where A is a real constant and n determines the topological charge of the phase singularity. Zone plates based on Eq. (2.2) are known as "spiral zone plates", since their angular variation is proportional to r^2, giving rise to a spiral shape. Figure 2.1(b) shows an example of a spiral zone plate.

Beams with helical wavefronts using phase masks have also been used to produce vortices (Beijersbergen et al., 1994). Such masks are known as spiral phase plates and consist of a phase object having a thickness T, proportional to the polar coordinate, i.e., $T(r, \theta) \propto \theta$. An example is shown in Fig. 2.1(c). Under the projection approximation the phase shift imparted to a wave is proportional to the thickness of the object (Paganin, 2006). If the thickness of the spiral phase mask is tailored such that $T(r, \theta = 2\pi) - T(r, \theta = 0)$ corresponds

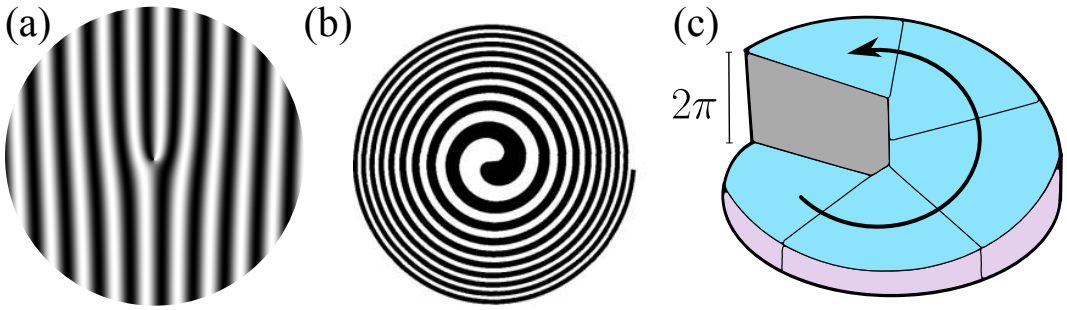

Figure 2.1: (a) Fork dislocation in a diffraction grating, which produces vortices in the far–field diffraction pattern. (b) The geometry of a $n = 1$ spiral zone plate used by Heckenberg et al. (1992). (c) Example of a spiral phase plate. The projected phase variation is given by $\exp(\pm i\theta)$, with the step corresponding to a phase shift of 2π.

to a phase shift of 2π, then under coherent plane wave illumination the transmitted beam will incur a phase shift of θ, identical to that of a beam with a helical wavefront. The projection approximation assumes that the wave receives the entire spiral phase shift at a single plane. However, because a phase object does not alter the intensity of the wave, the intensity cannot be zero anywhere in the wave at the exit surface of the phase plate. Hence a small amount of propagation of the transmitted wave is required for a vortex to form. The difficulty of fabricating a smooth and precise optical element meant that an approximate spiral phase plate consisting of a spiral staircase–like structure, could only be made. In the particular experiment of Beijersbergen et al. (1994) the phase plate consisted of 72 steps.

2.1.2 X–ray vortices

Peele et al. (2002) generated a first order vortex in the near field of a 9 keV X–ray beam, using a 16 step phase mask. Spiral zone plates have subsequently been used to produce vortices in the soft X–ray regime (Sakdinawat and Liu, 2007). These methods are useful for producing phase singularities in a controlled manner. However, intricate vortex–producing optical elements are not necessary for making X–ray vortices. Vortices naturally develop in propagating wave fields that contain moderate continuous phase modulations. The ease with which X–ray vortices may be spontaneously generated is highlighted by the work of Pavlov et al. (2011). In this work gold nanoparticles with ~1 μm diameter were deposited onto a substrate. When two adjacent nanoparticles were situated close to each other, they acted like a highly aberrated concave nanolens. Following the discussion in §1.2.1, caustics ultimately

arise from natural focusing by aberrated lenses and give rise to a singular skeleton of vortices. Propagation induced hard X–ray vortices were observed in the scattered field as a result of the natural focusing of the gold particle nanolens, in an analogous manner to the situation shown in Fig. 1.5(a). Spontaneous vortex nucleation has also been discussed by Kitchen et al. (2004), in the context of X–ray speckle fields. Using simulated data mimicking X–ray scattering from lung tissue, it was observed that phase singularities littered the resulting diffraction pattern (as discussed in §1.1.4). This example of propagation induced X–ray vortices highlights the ubiquity of vortices in optical and X–ray wave fields. It also suggests that X–ray phase singularities were likely produced as a result of random scattering of X–rays, long before they were created by spiral phase plates.

2.1.3 Electron vortices

Vortices in an electron beam was have been studied theoretically by Allen et al. (2001b) and Bliokh et al. (2007). Allen et al. (2001b) considered the effect electron vortices would have on conventional imaging methods in an electron microscope, whilst Bliokh et al. (2007) showed that wave–packet solutions to the Schrödinger equation could contain phase singularities. Vortex generation was latter demonstrated experimentally by Uchida and Tonomura (2010), who exploited a spiral phase–like plate. Since the de Broglie wavelengths of electrons are far smaller than those of visible light or X–rays, it is extremely difficult to specifically manufacture spiral phase plates for vortex production, although this was achieved by Shiloh et al. (2014). Instead, Uchida and Tonomura (2010) used regions where spontaneously stacked layers of graphite met in such a way as to form a very crude spiral phase–like plate, consisting of only three or four phase jumps. The thickness of the graphite layers loosely correlated to a phase shift of 2π. It is of interest to note that phase singularities in electron waves had been discussed many years ago by Dirac (1931), who showed that solutions to the Schrödinger equation possess a non–integrable phase given by Eq. (1.2) (Hirschfelder et al., 1974).

Until very recently, interference of planar electron waves had not been explored to create vortices in an electron microscope[3]. However, electron holography experiments have

[3]Electron vortices have now been produced by the deliberate formation of diffraction catastrophes. It can be shown that vortices associated with the hyperbolic umbilic catastrophe are the result of multiple wave interference (see Fig. 1.6(h)). This work will be discussed in detail in Chap. 6.

now been performed using two biprisms. The two reference waves plus the object wave results in three–wave interference in the electron microscope. The advantage of having two reference waves is the increased control over the fringe spacing and width of the interference pattern (Harada et al., 2004); the intensity of the fringes become modulated by the electromagnetic field, allowing for the direct visualization of the equipotential lines (Hirayama et al., 1995). In this case both biprisms were aligned parallel to each other, so that the second biprism produced an additional reference wave that was collinear with the other waves. It can be shown that degenerate wave–vectors do not produce vortices in the interference pattern. The work by Harada et al. (2005) did investigate the effect of introducing a relative rotation between the biprisms. However, their experimental setup did not produce vortices. Instead the focus of their work was the additional degree of control over the fringe pattern, which is desirable for electron holography experiments. Nevertheless it is possible to produce electron vortices using electron microscopes with relative rotation between two biprisms, as was recently demonstrated by Niermann et al. (2014) and Dwyer et al. (2015).

2.1.3.1 Simulations of propagation based vortex nucleation

Motivated by the recent publications reporting electron vortices, alternative methods of creating vortices in the electron microscope were investigated. Given the difficulty and precision required to engineer vortical masks for electrons, we sought to identify other methods of vortex production that didn't require phase–shifting optical elements. Multiple wave interference was first considered, but was not pursued since it required two biprisms to be inserted into the electron microscope. Instead we explored spontaneous generation of electron vortices, using computer simulations to identify if electron vortices could be generated in the electron microscope. Our simulations were based on amorphous carbon, which was chosen as a random phase object. This material is already used in electron microscopy for calibration purposes and/or as a specimen substrate. Showing that vortices could be generated from such a commonly used material would demonstrate the ubiquity of vortices in electron beams, and obviate the need for microscopic optical masks.

For the simulations, a thin ($\sim$ 50 nm) spatially random amorphous carbon sample was illuminated by 200 keV electrons. The exit wave was calculated assuming the projection approximation (Cowley, 1995). The wave field was numerically propagated using either

the operator or convolution formalism based on the Fresnel propagator, depending on the defocus (Paganin, 2006). Vortices were observed nucleating at a defocus of approximately ±200 μm from the sample. By extending the characteristic length and changing the interaction constant of the sample, the plane at which vortices nucleated could only be adjusted by an insignificant amount of ≈ 20 μm. However, our simulations did show that electron vortices could be generated in this way. Unfortunately the defocus was on the order of a few hundred microns, which lies far outside the typical propagation distances for the objective lens of a conventional electron microscope. The result indicates that despite vortices forming upon propagation from a spatially random scatterer, the interaction between the vortices and the material was too weak to produce large enough phase modulation for vortices to nucleate at a defocus less than hundreds of microns.

The simulated sample was then changed to model a cluster of randomly distributed latex spheres. This new object was also a homogeneous material and provided much larger phase variation of the transmitted wave (Eastwood et al., 2011). Because of its geometry, the spheres also behaved as aberrated lenses, which naturally focus the beam, emulating the experiment by Pavlov et al. (2011). Our simulated sample consisted of 30 latex spheres, each having 15 nm radii and a 5 nm root–mean–squared variation in their projected thickness; the spheres were distributed randomly over an area of 170 nm × 170 nm. Vortices formed in the scattered wave at approximately ±50 μm, half the defocus required for the amorphous carbon test case. An image of the simulated data is shown in Fig. 2.2(a), with the results of the numerically propagated intensity and phase of the scattered wave field at −100 μm defocus given in Figs. 2.2(b) and (c), respectively. Despite this improvement, the defocus values were still far outside the typical defocus range of a conventional transmission electron microscope, which is generally only capable of approximately ±5 micron defocus at most. However, microscopes fitted with Lorentz lenses can achieve much larger defocus lengths and therefore may be capable of creating electron vortices by propagation under the conditions of the simulation (De Graef, 2003)

The propagation distance required to form electron vortices could be decreased if the sample's interaction coefficient was increased. However, our initial aim was to show that vortex formation in the electron microscope is possible with very commonly used materials. Since artificially tailored experimental samples are of limited utility, we abandoned this line

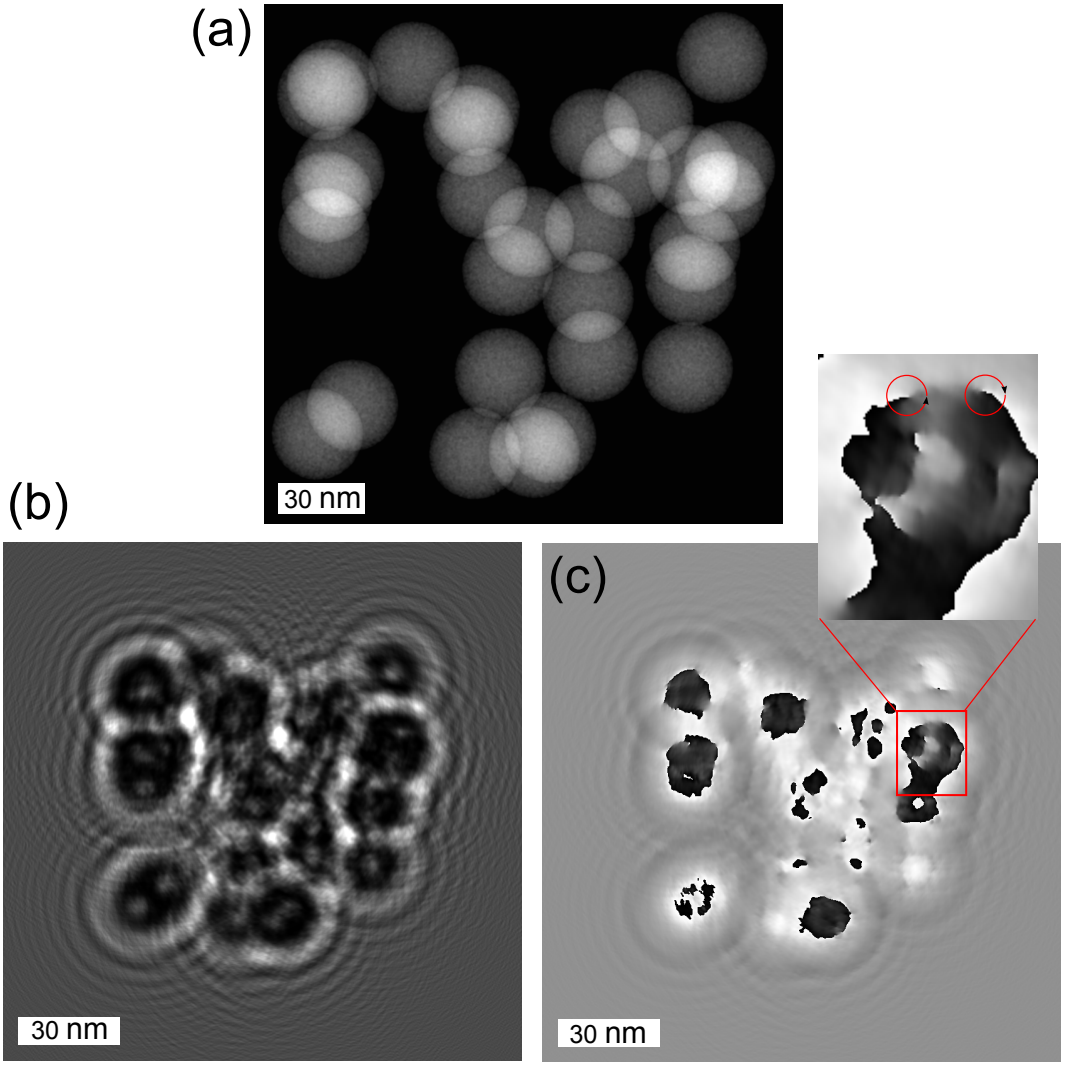

Figure 2.2: (a) Image of the simulated sample, which consisted of randomly placed 30 nm diameter latex spheres, with a 5 nm root–mean–squared variation in their projected thickness. Images (b) and (c) show the intensity and phase of the transmitted wave field at −100 μm defocus, using 200 keV electrons. Vortices are observed in the phase of the wave, where a region of the phase containing several vortices is magnified in the inset. A vortex and anti–vortex are encircled in red.

of research. Nevertheless this preliminary work stimulated our exploration of diffraction catastrophes in the electron microscope, which in turn led to the generation of electron vortices. This work is presented in Chap. 6.

Our observations contradict comments made by Uchida and Tonomura (2010) who, upon observing multiple vortices in their interference patterns, suggested that phase singularities are a common phenomenon for randomly scattered electron beams. It was their contention

that the stacked graphite flakes, which mimicked a crude phase plate, were the major determinant in generating vortices, rather than the nanoscale variations in the thickness of the material. However, it is difficult to make a direct comparison between our work and that of Uchida and Tonomura (2010) because the graphite sample has a slightly higher mean inner potential than carbon, and the defocus at which they observed phase singularities was not reported. The high degree of Fresnel fringing in the experimental images of Uchida and Tonomura (2010) suggests that these images are considerably out of focus (Petersen, 2014). If this is the case, then the formation of electron vortices may be due to their spontaneous formation from propagation, in addition to being a direct consequence of the sample acting like a spiral phase plate.

2.2 Three–Wave Interference

In §2.1.1 it was stated that a lattice of vortices and anti–vortices may be generated in the interference pattern of three or more waves. Here we consider the case of three waves interfering, which serves as the theoretical underpinning of Chap. 4. In our treatment we calculate the amplitude of an interference pattern using the summation of phasors for each wave, which is similar to the approach of Masajada and Dubik (2001) and Paganin (2006). Although we consider vortex formation for three–wave interference, vortices can also be formed from the superposition of more than three waves (O'Holleran et al., 2006).

Consider a wavefunction consisting of the superposition of three monochromatic complex scalar waves:

$$\psi(\mathbf{x}, t) = \sum_{j=0}^{n=2} \rho_j \exp\{i[\mathbf{k}^{(j)} \cdot \mathbf{x} + \phi_j(\mathbf{x})] - i\omega t\}, \tag{2.3}$$

where ρ_j is the amplitude, $\mathbf{x}_j$ is the position vector, $\mathbf{k}^{(j)}$ is the wave–vector and ϕ_j represents the spatial envelope of the jth wave. Note that the specific spatial dependence on ϕ_j is suppressed for clarity. We also omit the harmonic time dependence and factor out the first wave; furthermore we consider the interference pattern of Eq. (2.3) at the plane $z = 0$. Thus:

$$\begin{aligned}\psi(\mathbf{x}_\perp, z = 0) = \exp[i\mathbf{k}_\perp^{(0)} \cdot \mathbf{x}_\perp + i\phi_0]\{\rho_0 + \rho_1 \exp[i(\mathbf{k}_\perp^{(1)} - \mathbf{k}_\perp^{(0)}) \cdot \mathbf{x}_\perp + i(\phi_1 - \phi_0)] \\ + \rho_2 \exp[i(\mathbf{k}_\perp^{(3)} - \mathbf{k}_\perp^{(0)}) \cdot \mathbf{x}_\perp + i(\phi_2 - \phi_0)]\}.\end{aligned} \tag{2.4}$$

By making the following substitutions:

$$\mathbf{k}_\perp^A = \mathbf{k}_\perp^{(1)} - \mathbf{k}_\perp^{(0)}, \tag{2.5}$$

$$\mathbf{k}_\perp^B = \mathbf{k}_\perp^{(2)} - \mathbf{k}_\perp^{(0)}, \tag{2.6}$$

$$\phi^A = \phi_1 - \phi_0, \tag{2.7}$$

$$\phi^B = \phi_2 - \phi_0, \tag{2.8}$$

we arrive at:

$$\psi(\mathbf{x}_\perp) = \exp[i\mathbf{k}_\perp^{(0)} \cdot \mathbf{x}_\perp + i\phi_0]\{\rho_0 + \rho_1 \exp[i\mathbf{k}_\perp^A \cdot \mathbf{x}_\perp + i\phi^A] + \rho_2 \exp[i\mathbf{k}_\perp^B \cdot \mathbf{x}_\perp + i\phi^B]\}. \tag{2.9}$$

A necessary condition for the formation of vortices is that the wavefunction vanishes. For Eq. (2.9) this occurs when the terms within the braces sum to zero. Each of the three terms may be represented by a phasor in the Argand plane, with an angle relative to the real axis equal to the wave's phase and with length equal to the amplitude of the wave. Three phasors only sum to zero when they form a triangle, as shown in Fig. 2.3. For an intensity zero to occur, no single phasor magnitude can be too large, otherwise a triangle cannot be formed. For a vortex to form, a necessary condition on the amplitudes of each wave is that $\rho_i \leq \rho_j + \rho_k$, where $i, j, k = 0, 1, 2$. Assuming that intensity zeros do occur in our wavefunction, we now wish to locate their positions in the plane $z = 0$. With reference to Fig. 2.3(a), let α and β represent the value of the second and third wave's argument at the position of zero intensity, which we denote by $(\tilde{x}, \tilde{y})$. Thus:

$$\alpha = k_x^A \tilde{x} + k_y^A \tilde{y} + \phi^A + 2\pi n \tag{2.10}$$

$$\beta = k_x^B \tilde{x} + k_y^B \tilde{y} + \phi^B + 2\pi m. \tag{2.11}$$

Here we have expanded the dot products between the position and wave–vectors, and n and m are integers. Because α and β are the arguments of the oscillator complex exponentials in Eq. (2.9), we can add integer multiples of 2π to each of them. The coordinates of the vortices are found by solving Eqs. (2.10) and (2.11) simultaneously, giving

$$\tilde{x}(n, m) = \frac{k_y^B(\alpha - \phi^A - 2\pi n) + k_y^A(-\beta + \phi^B + 2\pi m)}{k_x^A k_y^B - k_y^A k_x^B}, \tag{2.12}$$

$$\tilde{y}(n, m) = \frac{k_x^B(-\alpha + \phi^A + 2\pi n) + k_x^A(\beta - \phi^B - 2\pi m)}{k_x^A k_y^B - k_y^A k_x^B}. \tag{2.13}$$

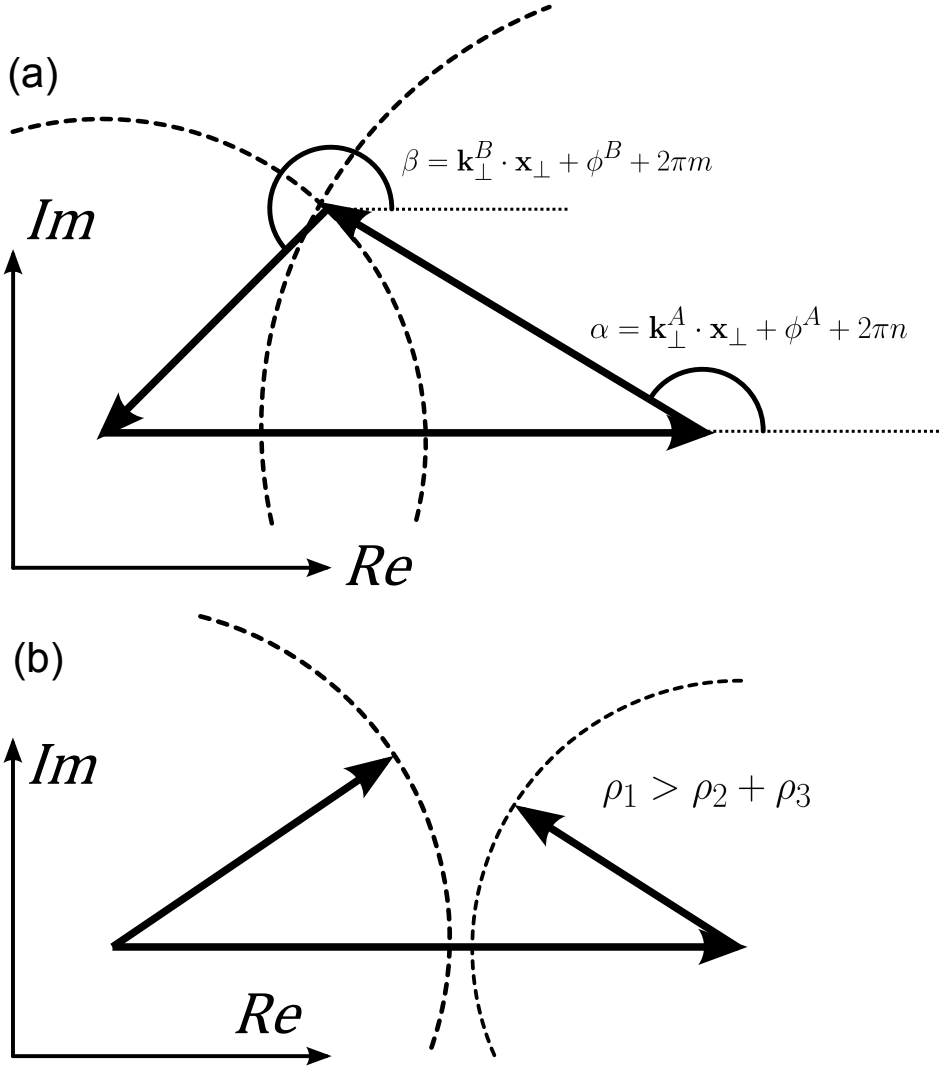

Figure 2.3: (a) Arrangement of phasors that sum to zero. The angles for which this occurs are denoted by α and β for the second and third phasors, respectively. (b) Vortices cannot form if $\rho_0 > \rho_1 + \rho_2$. The dashed arcs indicate possible values of the arguments of each wave. These arcs do not intersect it the length of the first phasor is too large, whence the phasors cannot sum to zero under these conditions. Note that for clarity the phasors have been offset from the origin of the complex plane.

Equations (2.12) and (2.13) define the Cartesian coordinates for each vortex at the (n, m)th lattice site. The geometry of this vortex lattice is determined by a combination of both the relative magnitude and angles of the wave–vectors of all three waves. The relative phase factors ϕ^A and ϕ^B lead to an irregular lattice. If they were equal to zero, corresponding to the interference of three plane waves all with zero phase at the origin, then Eq. (2.3) would

exhibit a uniform lattice of vortices.[4]

Although we have claimed that Eqs. (2.12) and (2.13) represent the coordinates of the vortices, this point is yet to be established, since the presence of an intensity zero is a necessary but not sufficient condition to establish the existence of a vortex. To determine if a vortex lies at the intensity zeros we must study the behaviour of the wavefunction in the vicinity of these points. To this end we consider points close to each vortex location and substitute $(\tilde{x} + \delta\tilde{x}, \tilde{y} + \delta\tilde{y})$, with $|\delta\tilde{x}|$ and $|\delta\tilde{y}| << 1$, into Eq. (2.9), and expand to first order in $\delta\tilde{x}$ and $\delta\tilde{y}$. Assuming the spatial modulation of each wave is sufficiently smooth, so that $\phi_j(\tilde{x} + \delta\tilde{x}, \tilde{y} + \delta\tilde{y}) \approx \phi_j(\tilde{x}, \tilde{y})$, and using $\psi(\tilde{x}, \tilde{y}) = 0$, after factoring the resulting equation into real and imaginary parts we find:

$$
\begin{aligned}
&\psi(\tilde{x} + \delta\tilde{x}, \tilde{y} + \delta\tilde{y}) \approx \\
&\qquad \rho_0 \exp[i(\mathbf{k}^{(0)} \cdot \mathbf{x} + \phi_0)]\Bigg\{\Bigg[\left(k_x^{(1)}\delta\tilde{x} + k_y^{(1)}\delta\tilde{y}\right)\frac{\rho_1}{\rho_0}\mathrm{Im}\left(\exp[i(\mathbf{k}^A \cdot \tilde{\mathbf{x}} + \phi^A)]\right) \\
&\qquad + \left(k_x^{(2)}\delta\tilde{x} + k_y^{(2)}\delta\tilde{y}\right)\frac{\rho_2}{\rho_0}\mathrm{Im}\left(\exp[i(\mathbf{k}^B \cdot \tilde{\mathbf{x}} + \phi^B)]\right)\Bigg] \\
&\qquad + i\Bigg[\left(k_x^{(0)}\delta\tilde{x} + k_y^{(0)}\delta\tilde{y}\right) + \left(k_x^{(1)}\delta\tilde{x} + k_y^{(1)}\delta\tilde{y}\right)\frac{\rho_1}{\rho_0}\mathrm{Re}\left(\exp[i(\mathbf{k}^A \cdot \tilde{\mathbf{x}} + \phi^A)]\right) \\
&\qquad + \left(k_x^{(2)}\delta\tilde{x} + k_y^{(2)}\delta\tilde{y}\right)\frac{\rho_2}{\rho_0}\mathrm{Re}\left(\exp[i(\mathbf{k}^B \cdot \tilde{\mathbf{x}} + \phi^B)]\right)\Bigg]\Bigg\}.
\end{aligned}
\tag{2.14}
$$

In the coordinates $(\delta\tilde{x}, \delta\tilde{y})$, whose origin coincides with $(\tilde{x}, \tilde{y})$, this equation is of the form $[(a\delta\tilde{x} + b\delta\tilde{y}) + i(c\delta\tilde{x} + d\delta\tilde{y})]\exp(i\chi)$, where a, b, c, b and χ are real. If we view Eq. (2.14) as a mapping of the local coordinates of a vortex to the complex plane, i.e. $\psi : (\delta\tilde{x}, \delta\tilde{y}) \mapsto \mathbb{C}$, then it can be shown that a vortex exists at the point $(\tilde{x}, \tilde{y})$ if the Jacobian determinant of this mapping is non–vanishing.[5] Additionally, the sign of the topological charge of the vortex is given by the sign of the Jacobian determinant (Rothschild et al., 2012, 2014). It follows that a vortex is present at $(\tilde{x}, \tilde{y})$ when $ad - bc > 0$, whilst an anti–vortex is found when $ad - bc < 0$. Alternatively, Eq. (1.2) could be calculated about this point to test for the presence of a vortex and determine its topological charge. We have only considered the formation of vortices at the intensity zeros when the phasors form an upward triangle; a downward phasor triangle is also formed when the phase angles of the second and third waves equal α and β, respectively. In this case vortices of the opposite topological charge are formed and their positions may be

[4] In Chap. 4 we show how the vortices' dependence on the wave envelopes may be exploited in imaging.

[5] The situation where this Jacobian determinant vanishes is also interesting and is explored further in Chap. 7.

derived in the same manner as above.

2.2.1 Nodal line topology of multiple wave interference

We have shown that a vortex lattice results from three–wave interference at the plane $z = 0$. O'Holleran et al. (2006) studied the topology of the nodal structure of a vortex lattice produced by three, four and five plane wave interference. For three–wave interference, the nodal lines propagate rectilinearly and are all parallel, hence the vortices do not move relative to each other. This rigidity of the vortex lattice is reflected in the geometry of the phasors. There is only one phase angle for each wave that can give rise to the triangle phasor geometry. As the wave propagates, all phasors rotate simultaneously, which preserves the shape of the phasor triangle at each vortex point. However, for the case of four wave interference, the additional wave introduces a degeneracy in the phase angles required for the phasors to sum to zero. Assuming equal amplitudes, complete destructive interference of four waves occurs when the four phasors form a rhombus. At each vortex point, the angles between the adjacent edges of the rhombus undergo cyclic evolution as the wave propagates. Consequently, the condition on the phase angles of each wave changes as a function of propagation, causing the vortex lattice to contract and expand as it propagates. All vortices annihilate once the lattice has contracted such that vortices of opposite topological charge are brought together. This plane of annihilation occurs where the relative angles of each phasor are either 0 or π as this means the phasors are collinear. Therefore the nodal lines do not exhibit parallel propagation, but instead intersect each when their phasors are collinear. With small changes to the relative magnitudes of each wave, the crossed nodal lines can be perturbed, such that the nodal lines form a series of closed loops.

A qualitatively different scenario happens when the amplitudes of each wave are such that the phasors can only form a trapezium. Here, for any relative phase angle the four phasors cannot be simultaneously collinear, and vortex creation and annihilation cannot occur. Similar to the nodal line topology of three–wave interference, the nodal lines extend to infinity and do not interact with adjacent nodal lines. The additional wave induces a perturbation to these lines, causing the vortices to propagate in non–straight trajectories. The nodal line topology quickly becomes too intricate to describe for interference with five plane waves, as noted by O'Holleran et al. (2006). Vortex positions are dictated by the phasors

forming a five–sided polygon with a large degree of degeneracy in the phase angles required for vortex formation.

The nodal line structure of the interference between three spherical waves generated from a 3–pinhole interferometer was studied by Ruben and Paganin (2007). It was shown that the interference of three spherical waves only generated a finite number of first order vortices in contrast to plane waves, which generate an infinitely extended vortex lattice. The nodal lines in this case expand radially outwards as the wave propagates. There are two qualitatively different classes of nodal line present for three spherical wave interference. The nodal lines either originate at the source plane or generate curved hairpin nodal lines after some critical propagation distance as a result of vortex/anti–vortex creation events. The nodal lines in the far–field become asymptotically straight, since the spherical waves are approximately planar in this limit. As more waves are interfered, the field begins to resemble that of a far–field speckle pattern. The resulting complicated field exhibits knotted nodal loops (Leach et al., 2004, 2005) and linked nodal loops (O'Holleran et al., 2009a; Padgett et al., 2011). An exotic example arises in the interference of two counter–propagating Bessel beams. Here the nodal lines become interwoven forming a braided topology (Dennis, 2003).

2.2.2 Quasi–periodic optical lattices

In multiple wave interference, where the polar angles of all wave–vectors, relative to the optic axis, are equal, the resulting wave becomes diffraction free. Additionally, when the number of waves is $2n + 5$, where $n = 0, 1, 2...$, the intensity and vortex pattern of the resulting interference patterns is quasi–periodic (Becker et al., 2011; Chen et al., 2011). The term quasi–periodic relates to the lattice having a regular and ordered structure but no translational symmetry. By phase engineering it is possible to produce quasi–periodic optical lattices that have spiral nodal line structures (Xavier et al., 2011), which upon propagation give vortices (Xavier et al., 2012). Quasi–periodic optical lattices have been used to study quasi–crystals (Shechtman et al., 1984). Guidoni et al. (1997) localized single atoms of Caesium within a 3D quasi–periodic optical lattice to physically simulate a quasi–crystal. Freedman et al. (2006, 2007) generated quasi–crystal photonic lattices within a non–linear medium to study wave and transport phenomenon quasi–periodic order systems.

2.3 Singularimetry

A number of experiments exploiting vortex lattices generated by three–beam interference have been performed using specially made three–beam interferometers; this technique is refereed to as Optical Vortex Interferometry (OVI). Masajada (2004) outlined a simple method showing how an optical vortex lattice may be used for singularimetry. One of the three beams of the OVI was reflected off a mirror, which was free to rotate about two of its axes. The wave–vectors of the other two beams were arranged so that when the mirror is rotated about its x axis, the vortex density of the lattice changed. When it was rotated about the y axis, the relative angle between the vortices altered. Equations (2.12) and (2.13) show that the positions of the vortices within the lattice are related to the relative phase difference between each wave. Tilting the reflecting mirror induces a linear phase shift in one of the beams, which in turn causes dilation and changes to the geometry of the vortex lattice. Using equations similar to Eqs. (2.12) and (2.13), Masajada (2004) showed that by selecting just three vortices, which need not be adjacent to one another, measurement of the change in the separation and relatives angles between each vortex was sufficient to determine the magnitude and orientation of the mirror. Popiołek-Masajada et al. (2007) went on to improve the precision of this technique by developing a method based on measuring the trapezoidal cell that is formed by four neighbouring vortices of the same topological charge. Changes in this cell were averaged over the entire lattice and a statistical probability distribution used to determine the tilt of the mirrors.

Polarization singularities have also been used to perform a similar type of singularimetry. A modified type of OVI, described by Kurzynowski et al. (2006), used a modified Mach–Zehnder interferometer, where one arm was fitted with a birefringent Wollaston prism (Brosseau, 1998). This prism served to rotate the polarization of one of the beams, effectively generating a third beam of altered polarization. This interferometer produced a lattice of polarization singularities known as C points (Nye and Hajnal, 1987). These singularities can be thought of as the polarization equivalent of optical vortices. For a complex vector field $\mathbf{E}$, a C point is defined as a point where the real and imaginary parts of the complex scalar $\mathbf{E} \cdot \mathbf{E}$ vanish simultaneously. Using interferometry with two Wollaston prisms, Borwińska et al. (2007) and Woźniak and Banach (2009) used the lattice of C points to measure the prism's linear birefringence parameters, i.e., the azimuthal angle of its

first eigenvector and the phase difference between the eigenwaves. Using the Jones vector description of polarized light, it was shown that the relative displacement of each sub–lattice of polarization singularities of alternate sign is a function of the phase difference between each Jones vector component introduced by the birefringent prism. The phase difference is determined by measuring the shift in the sub–lattice of one particular topological charge from the other sub–lattice. A change in the prism's first eigenvector caused a linear shift in both sub–lattices in the same direction. Measurement of this shift allowed the azimuthal angle to be determined.

The use of the OVI is limited in its scope as the technique only measures the simple global parameters of an optical element in the interferometer. Masajada et al. (2002) considered using vortices within the lattice as phase markers, which could potentially be exploited to measure non–uniform phase variations in one arm of the OVI. The first serious attempt at this is due to Masajada and Popiołek-Masajada (2005), but the proposed technique had shortcomings. Frączek and Mroczka (2008) reported a more practical phase measurement technique using an optical vortex lattice. In this method, vortices are localized within an optical vortex lattice, which has been perturbed by a non–uniform phase variation in one of the reference arms. The relative phase between each vortex within the lattice is then plotted as a function of the transverse Cartesian coordinates. A plane is fitted to the vortex coordinates, which represents the positions where each vortex would lie if the lattice was unperturbed. The distance between the positions of the vortices relative to the fitted plane is taken as the phase measurement of the reference arm. The fact that the plane is fitted to all vortices within the lattice means that this is a global technique, as the phase solution of any point still depends on the entire interferogram of the OVI. In Chap. 4 we present a technique that utilises an optical vortex lattice to retrieve a non–uniform phase variation locally. This technique is also capable of handling intensity variations in the beam and does not require a reference lattice.

2.3.1 Vortices for super–resolution measurements

Due to the unique properties of vortices, singularimetry may be used for super–resolution measurements. Super–resolution refers to imaging or measurement of a feature that is smaller than that dictated by the Rayleigh resolution criterion due to diffraction. Tychinsky (2008)

proposed that this resolution criterion should be framed in terms of phase singularities, as they define a singular feature of a wave field, which is smaller that the wavelength. This means that phase singularities are associated with an optical phenomenon known as superoscillations (Berry, 1994b,a; Dennis et al., 2008). This term refers to the ability of a band–limited signal to oscillate faster than its maximum Fourier component. This apparent paradoxical behaviour is resolved by noting that the signal becomes exponentially smaller the faster the oscillations, similar to an evanescent wave. However, unlike evanescent waves, superoscillations may persist in the far–field (Berry and Popescu, 2006). Because the superoscillatory parts of the field are oscillating faster than their maximum Fourier component (dictated by the wavelength), these features of interest have an effective Rayleigh criterion smaller than that of the rest of the field, thereby making them capable of achieving super–resolution.

Spektor et al. (2008) demonstrated the high sensitivity of phase singularities to sub–wavelength objects. An experiment using a 488 nm wavelength beam containing an edge dislocation was performed. The line phase singularity was scanned over a 1D phase step on a glass substrate with a phase height 500 nm. An integration window was defined over the initial region of the phase singularity, where the intensity is very low. As the singular beam is scanned across the phase step, the line singularity is perturbed, and made visible by observing the surrounding intensity near the singularity. Plots of the integrated intensity as a function of the position of the phase step showed that the singular beam is capable of a 20 nm lateral resolution. This was subsequently reduced to 10 nm using 532 nm wavelength light (Spektor et al., 2010). A similar experiment was also performed by Hemo et al. (2011), who used vortices rather than line singularities, to image 50 nm latex spheres.

Vortices are also sensitive to sub–wavelength objects because the wave field becomes superoscillatory near a vortex core. The phase varies by 2π radians around a vortex, with this variation occurring faster towards the centre of the vortex; thus there there will be some point near the vortex where the phase variation becomes greater than the highest spatial frequency of the wave field. This is accompanied by the intensity vanishing at the vortex core. Brunet et al. (2010) used phase–singularities to perform quantitative measurements with singular acoustic waves. An array of transducers was used to generate a helical acoustic wavefront in a rectangular tank of water. The helical beam was diffracted by a slightly off–axis circular

aperture, five times smaller than the wavelength of the acoustic wavefront. The authors showed that by measuring the shift[6] of the acoustic vortex both its diameter and the position of the aperture could be accurately imaged.

Doubly–charged optical vortices have also been utilized for the purposes of super–resolution. An experiment by Masajada et al. (2009) used a second–order vortex to laterally resolve a micro–step with a precision smaller than the radiation wavelength. However, singularimetry utilising second–order vortices is quite different to that using first–order vortices, since a small perturbation to a second–order vortex causes it to decay into two first order vortices (see §1.1.3, and the discussion related to Fig. 1.3(b)). When a singular beam is scanned towards the phase step, the two vortices exhibit asymmetric trajectories, since their positions within the beam are perturbed differently by the step, due to their opposite topological charges. As the beam continues to scan across the specimen, Masajada et al. (2009) showed that the point where the separation between each vortex was minimized corresponded to when the centre of the beam was aligned with the micro step. Working with a numerical aperture of ~0.07, the 300 nm high phase step was localized to within ±60 nm, giving an effective numerical aperture of 0.95. The authors claimed that this method would be capable of achieving a higher resolution of 25 nm.

Mari et al. (2012) also utilised second–order vortices, in a method they dubbed "optical vortex coronagraphy" (OVC), which is used for direct detection of extrasolar planets (Swartzlander, 2001). Light emitted from two incoherent point sources, which are not resolvable according to the Rayleigh criterion, are split into an on– and off–axis beam. The on–axis beam passes through a second–order spiral phase plate, which introduces a second–order optical vortex to the beam, whilst the off–axis beam avoids the spiral phase plate. This setup results in an asymmetric intensity "doughnut" in the image plane. Mari et al. (2012) found that the distance between the optic axis and the intensity minimum of the image was equal to the angular separation between the two incoherent point sources. By localizing the position of this minimum, the sources could be resolved to $0.1\lambda/D$, where D is the diameter of the aperture of the OVC. Whilst this technique could be performed using a first–order vortex, the second-order vortex has proven to give the method better sensitivity (Swartzlander, 2009).

[6] A property of vortices is that upon propagation they diffract in the direction perpendicular to the intensity gradient (Rozas et al., 1997).

Elementary Catastrophe Theory

3

The singular duals of vortices, namely, caustics and diffraction catastrophes, underpin the work of this thesis. The principal mathematical tools describing these optical singularities are defined by catastrophe theory, to which this chapter serves as an introduction. The mathematical concepts presented here are essential to understanding the contents of Chaps 5 and 7, since the key physical insights of these chapters exploit the results of catastrophe theory. In addition, we shall also show under what conditions catastrophes arise in diffraction theory and how they are related to the ray picture of light. This discussion serves to underpin the content of Chap. 6, which relies on diffraction physics producing diffraction catastrophes.

3.1 Structural Stability and Critical Points

It is difficult to precisely define what constitutes catastrophe theory. It brings together various results and principles from a number of different areas in mathematics, such as topology, singularity theory and differential geometry. At its core, it is the study of abrupt or discontinuous changes of dynamical physical systems, which arise from smooth continuous changes of associated control parameters. This process is described in the context of structural stability of functions and families of functions.

Consider an elastic material under tensile stress. As we smoothly[1] increase the applied tension, the length of the material smoothly increases. However, at some point, a further increase in the tension results in a catastrophic event; namely, the material snaps—it has abruptly changed the way it responds to the applied force. There are many examples of this behaviour in physical systems: the rolling of a ship as it tilts to one side, the collapsing of a

[1] By smoothly we refer to a constant rate of change in the applied tension.

bridge as more weight is applied, or the broken back of a camel after adding the proverbial last straw.

This type of behaviour is quantified by the mathematical concept of structural stability, with an abrupt change in a dynamical system leading to a structural instability. A dynamical system is governed by the set of n first–order differential equations:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = F_i(x(t)_1, \ldots, x(t)_n; u_1, \ldots, u_m), \tag{3.1}$$

where $i = 1, \ldots, n$, x_i are the local coordinates that define a n–dimensional phase space, and u_j is a set of real control parameters. For clarity, the explicit dependence of x_i on t is suppressed for the remainder of this chapter. The flow lines defined by Eq. (3.1) in the phase space give the phase portrait of this system. The system governed by the set of equations (3.1) is said to be structurally stable if, given a small perturbation to F_i, the change in the phase portrait is topologically equivalent (Stewart, 1982). General catastrophe theory is the study of the way in which the phase portrait undergoes a topologically distinct change due to a perturbation, leading to an instability.

Elementary catastrophe theory simplifies this analysis by dealing with a subset of the equations. Because it is strictly a local theory, we can assume the first order linear approximation by taking $F_i = \partial G/\partial x_i$, hence

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{\partial}{\partial x_i} G(x_1, \ldots, x_n; u_1, \ldots, u_m). \tag{3.2}$$

The problem considers gradients of a single smooth real function G rather than n functions given by F_i. This reduces the analysis to the local stability of points of the single function G. Specifically, one considers points where $\partial G/\partial x_i = 0$, since the critical points are the only topologically significant features of a function.

The notion of structural stability can now be defined locally. With reference to the Eq. (3.4), a smooth function f is structurally stable at a point p, when at some point p' within some local neighbourhood of p, there is a corresponding function g, such that f is equivalent to g, i.e., $f \sim g$, where the symbol $\sim$ denotes equivalence. The notion of functional equivalence is central to catastrophe theory. If two functions are deemed equivalent, then there is a one–to–one smooth and reversible mapping between them, i.e., a diffeomorphism. If f is locally equivalent to g at the origin, then

$$f(x) = g(h(x)) + \gamma, \tag{3.3}$$

where $h : \mathbb{R}^n \to \mathbb{R}^n$ is a local diffeomorphism and γ is a constant. The function h acts to rotate and dilate the coordinates, whilst γ is a simple shift of f , which does not change the critical points of f.

With functional equivalence defined, local structural stability may be summarised as follows: given two diffeomorphisms h_1 and h_2, which act to perturb p and $f(p)$, respectively, the following diagram commutes (Lu, 1976).[2]

$$\begin{array}{ccc} p & \xrightarrow{\quad f \quad} & f(p) \\ h_1 \downarrow & & h_2 \downarrow \\ p' & \xrightarrow{\quad g \quad} & g(p') \end{array} \tag{3.4}$$

Since structural stability is now characterised by the critical points of a single function, the natural question to ask is: What type of critical points are stable and which are not? The answer comes from a powerful result in singularity theory known as the Morse lemma (Poston and Stewart, 1996).

Lemma 1 (Morse Lemma)**.** *Let x_0 be a non–degenerate critical point of a smooth function $f : \mathbb{R}^n \mapsto \mathbb{R}$, for which the determinant of the Hessian matrix of f is non–zero at x_0. Then f is structurally stable at x_0 and there exists a local coordinate system $(y_1, \ldots, y_n)$ in a neighbourhood X of x_0, with $y_i(x) = 0$ for all i, such that*

$$f = f(x_0) - y_1^2 - \cdots - y_l^2 + y_{l+1}^2 + \cdots + y_n^2 \tag{3.5}$$

for all $x_0 \in X$.

Proofs of the Morse lemma may be found in Milnor (1963) and Poston and Stewart (1996). The Morse lemma states that a function is structurally stable if it contains no degenerate critical points and is thus locally equivalent to the Morse function $\pm x_1^2 \pm \cdots \pm x_n^2$. A Morse function is a function that contains no degenerate critical points and is therefore structurally stable at all points.

[2]Commuting in this context means that p may be transformed into $g(p')$ by the sequential application of f and h_2, or h_1 and g, to p.

3.2 Corank

The corollary of the Morse lemma is that an instability arises from the existence of degenerate critical points in the functions governing the dynamics of a physical system, G in Eq. (3.2). In two dimensions, the Hessian determinant of a function f, is given by:

$$\det[\mathbf{H}(f)] = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 . \tag{3.6}$$

A critical point is degenerate when $\det(\mathbf{H}) = 0$, which can be satisfied under two conditions. Equation (3.6) may vanish when each second order partial derivative vanishes simultaneously ($f_{xx} = f_{yy} = f_{xy} = 0$); f is then evidently degenerate in both the x– and y–directions (Saunders, 1980). The case where $f_{xx}f_{yy} = f_{xy}^2$ is not so clear. To proceed we must introduce another result from singularity theory, called the splitting lemma.

Lemma 2 (Splitting Lemma)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function with a degenerate critical point at x_0. Then in the neighbourhood of x_0, f is equivalent to a function of the form*

$$g(x_1, \ldots, x_r) \pm x_{r+1}^2 \pm \cdots \pm x_n^2 \tag{3.7}$$

where $g : \mathbb{R}^r \to \mathbb{R}$ is some smooth function.

The splitting lemma states that if all second order partial derivatives $\partial^2 f/\partial x_i \partial x_j$ do not vanish simultaneously, there exists a diffeomorphism, such that f can be "split" into non–essential variables that do not contribute to the instability and "essential" variables that do. The dimensionality of g, denoted by r, is called the corank of f. The corank is the number of directions in which the critical point of f is degenerate.

3.3 Unfoldings

So far we have only considered the local structural stability of the critical points of the function G, as defined by Eq. (3.2). Suppose now that the set of m constant control parameters u_i, of G varies continuously, so that G now represents an m–parameter family of functions of n variables. A family of functions is said to be structurally stable if, under a small perturbation of the whole family, the topology of the family is unchanged. A family of functions may contain structurally unstable functions or points, while the family itself may

still enjoy structural stability. These results can be generalised to allow G to represent an m–parameter family of functions (Poston and Stewart, 1996).

Since our treatment of Eq. (3.2) is local, we are free to express G as a truncated Taylor series expansion about the origin.[3] The order to which G may be truncated depends on "how local" a neighbourhood about the origin we wish to consider.[4] Therefore, a generic m–parameter family of functions G, containing a polynomial with a degenerate critical point s, is given by

$$G(x, y; u_1, \ldots u_m) = s(x, y) + \sum_i \sum_j u_{ij} x^i y^j, \tag{3.8}$$

where $i + j = m$ and for simplicity we have restricted G to only two variables (x, y). We say that G is an unfolding of the singularity s. The monomials act as perturbing terms to s; the parameters u_i control the level of perturbation. The unfolding of s is said to be versal if G represents a structurally stable family of functions. The terms in the summation are labelled as the unfolding terms, with their corresponding coefficients called control parameters. The m–dimensional Euclidean space coordinatized by the smoothly varying control parameters is the control space. Given any versal unfolding of a singularity, each point of the control space represents a single function belonging to a structurally stable family. The set of points in control space that correspond to unstable functions of that family define the bifurcation set. The bifurcation set partitions the control space in such a way that each region only contains points associated with equivalent functions.

3.4 Codimension

Recall that the number of equations required to define a manifold depends upon the dimensionality of the space within which it is embedded. A curve is fully specified by one equation in two dimensions. However, in three dimensions a single linear equation represents a plane; two equations being needed to describe a curve in three dimensions. Given an n'–dimensional object in an n–dimensional space, we require $m = n - n'$ equations to describe the manifold. The quantity $\dim(m)$ is known as the codimension of the manifold. The codimension is a useful quantity as it does not change when the dimensionality of the manifold is decreased.

[3] Taylor series truncated at the kth order are known as k–jets.

[4] This point is justified in Chap. 5.

A curve in three dimensions has codimension 2; however, a curve appears as a single point in a two–dimensional slice, which is still of codimension 2.

Equivalently, for a family of functions $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, describing the unfolding of a non–Morse function (see Eq. (3.8)), dim(m) is the codimension of the unfolding of F. Evidently the dimension of the control space is equal to the codimension of the unfolding of a non–Morse function. This non–Morse function is known as the germ of the set of functions in a family, which can be defined as the function obtained by setting all control parameters to zero. The germ is the function corresponding to the origin of the control space.

The codimension of the singularity of the germ is the minimum number of unfolding terms needed to fully describe all possible bifurcations of structurally unstable functions within the family. This is known as the universal unfolding of the germ. As each control parameter defines a set of coordinates in control space, the universal unfolding parameters define a set of linearly independent coordinates (see Eq. (3.8)). The codimension may be thought of as a measure of complexity of the germ's critical point. The higher the order of the degeneracy, the more unfolding terms are needed to describe all possible bifurcations. The derivation of the universal unfolding terms is described in Chap. 5.

3.5 Thom's Classification

Given our discussion of the structural stability of a family of functions, the purpose of catastrophe theory, may be stated as follows: Catastrophe theory is the study of the bifurcations in families of functions formed by the versal unfoldings of degenerate singularities. However, there are infinitely many ways we may unfold or parameterize singularities. Yet due to a powerful result by Thom (1975, 1977), any type of versal unfolding of a singularity may be classified by one of the so called elementary catastrophes.

Theorem 3 (Thom's Classification Theorem)**.** *Let $F : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ parameterize a family of smooth functions $f : \mathbb{R}^n \to \mathbb{R}$ of finite codimension. Suppose $k \leq 4$, then at any point $(x_i, u_j) \in \mathbb{R}^n \times \mathbb{R}^k$, $F(x_1, \ldots, x_n; u_1, \ldots, u_k)$ is equivalent to one of the universal unfoldings listed in Table 3.1.*

This classification means that for any family of smooth functions, there is a local diffeomorphism that transforms it to one of Thom's elementary catastrophes. These functions

Name	Canonical form	Corank	Codimension
Non–critical	x	–	0
Morse	$\pm x_1^2 \pm x_2^2 \pm ... \pm x_n^2$	–	0
Fold	$x^3 + u_1 x + M$	1	1
Cusp	$\pm x^4 + u_1 x^2 + u_2 x + M$	1	2
Swallowtail	$x^5 + u_1 x^3 + u_2 x^2 + u_3 x + M$	1	3
Butterfly	$\pm x^6 + u_1 x^4 + u_2 x^3 + u_3 x^2 + u_4 x + M$	1	4
Hyperbolic Umbilic	$x^3 + y^3 + u_1 xy + u_2 x + u_3 y + M$	2	3
Elliptic Umbilic	$x^3 - xy^2 + u_1(x^2 + y^2) + u_2 x + u_3 y + M$	2	3
Parabolic Umbilic	$x^4 + xy^2 + u_1 x^2 + u_2 y^2 + u_3 x + u_4 y + M$	2	4

Table 3.1: List of Thom's classification of elementary catastrophes. M represents the Morse function $\pm x_{r+1}^2 \pm x_{r+2}^2 \pm ... \pm x_n^2$ and r is the corank of f.

are known as the canonical form of each catastrophe and represent the universal unfolding of the associated singularity. An unfolding by a versal unfolding parameter is not linearly independent of a universal unfolding term.

Although Table 3.1 only gives the canonical forms of a family of functions with less than or equal to 4 parameters, the classification can be generalised to higher orders of k, making the list infinite (Arnol'd, 1975). To better describe the notion of bifurcation structure, we now detail the rich geometry of a select few elementary catastrophes. Specifically the fold, cusp and elliptic umbilic catastrophes are described in increasing order of complexity. These demonstrations of the working techniques of catastrophe theory will aid in the interpretation of more abstract applications later in this thesis; such as caustic wavefront determination in Chap. 5, and order parameter catastrophe defect classification in Chap. 7.

3.5.1 Fold catastrophe

Here we derive the bifurcation set of the fold catastrophe, whose canonical form is given by

$$f(x;u) = x^3 + ux. \tag{3.9}$$

Suppose that the germ x^3 represents the potential function governing a physical system and the term ux acts as a perturbation to the germ of the system. The family f describes the dynamics of the system. Equilibrium points of our system correspond to extrema in the

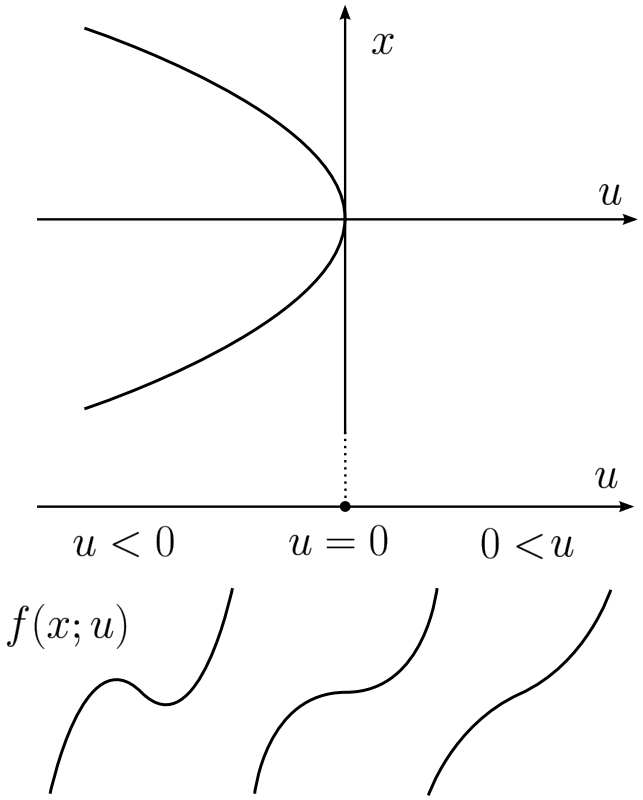

Figure 3.1: The equilibrium set, denoted by E, of a fold is parabolic in x. Projecting the critical point of E onto the control space shows that the bifurcation set is given by the point $u = 0$ in the 1D control space. The function either side of the germ shows a different number of critical points.

system's potential function. Hence we look for points of f, such that $\nabla f = 0$, where ∇ is the gradient operator with respect to the state variables only. This set of points defines the equilibrium manifold E of the fold catastrophe, given by $3x^2 + u = 0$. The instability of the system stems from all degenerate critical points of f, so we are only interested in a subset of E corresponding to a set of points where $\det[H(f)] = 0$ (see Eq. (3.6)); this is called the singularity set, denoted by S. The singularity set is the set of positions of the degenerate critical points in state space. For the fold we find $6x = 0$. Using both E and S, we eliminate the state variable of f to find the position of the unstable functions of f in control space, giving the bifurcation set B, of f. This is equivalent to the projection of the critical points of E onto the control space coordinate. Evidently B for the fold catastrophe constitutes a single point, $u = 0$. For the control parameter $u \geq 0$, f contains a single non–stationary point of inflection. When the control parameter is smoothly decreased below $u = 0$, the germ x^3

bifurcates into two critical Morse points. The state of the corresponding physical system abruptly changes its configuration when the equilibrium conditions, determined by f, change at this point. This situation is shown schematically in Fig. 3.1.

3.5.2 Cusp catastrophe

The canonical form of the cusp catastrophe is given by

$$f(x; u, v) = \pm x^4 + ux^2 + vx. \tag{3.10}$$

There exists two types of cusp catastrophe, corresponding to the alternate sign in front of the leading term in Eq. (3.10). We shall just consider the positive case. The equilibrium surface and the singularity sets of the cusp are:

$$E \qquad 4x^3 + 2ux + v = 0, \tag{3.11}$$

$$S \qquad 12x^2 + 2u = 0. \tag{3.12}$$

The singularity surface of the cusp is two–dimensional and is the cross section of all critical points of $f(x; u, v)$. Near the origin, the singularity set becomes many–to–one and the manifold can no longer be described by an explicit function, as can be seen in Fig. 3.2(a). The "folds" of the surface give the critical points of E, which are degenerate critical points of the family f. Their projection onto the control space defined by (u, v) is found by solving Eqs. (3.11) and (3.12) simultaneously. The bifurcation set of the cusp is found to be

$$8u^3 + 27v^2 = 0. \tag{3.13}$$

The geometry of B is shown in Fig. 3.2(b). The bifurcation set of the cusp partitions the control space into two regions, as it did for the fold catastrophe. Recall how the codimension of a germ is a measure of the "complexity" of its singularity. Although only a single bifurcation occurs at B, for the cusp catastrophes, the germ x^4 unfolds into three critical points, compared to two for the case of the fold catastrophe involving x^3. To fully describe all possible unfoldings of the cusp catastrophe its codimension must be of a higher dimensionality than that of the fold catastrophe. Functions not within the region bounded by B in control space only have a single critical point and a non–critical point of inflection. However, the point of inflection becomes structurally unstable at points in B which bifurcate, ensuring that $f(x; u, v)$ has three critical points within the curve determined by Eq. (3.13).

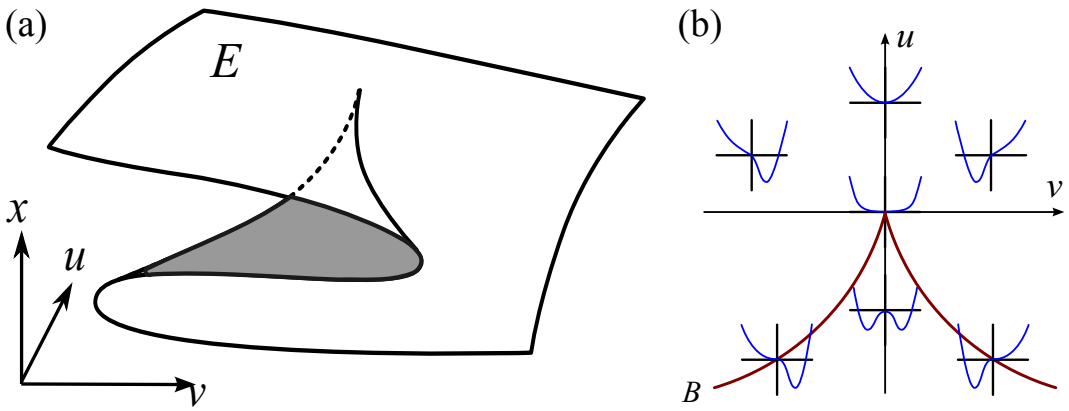

Figure 3.2: (a) The equilibrium manifold E of the cusp catastrophe. The critical points of this manifold form the bifurcation set B. The geometry of B is found by projecting the fold lines of E onto the plane defined by the control coordinates. (b) Functions inside the boundary of the cusp exhibit three critical points, while functions corresponding to points outside the cusp only contain a single minimum. Functions corresponding to points on the bifurcation set exhibit an unstable point of inflection.

3.5.3 Elliptic umbilic catastrophe

The elliptic umbilic catastrophe is one of two possible catastrophe germs with a degenerate critical point in two–dimensions. These so called umbilic points have a corank of 2 because the Gaussian curvature vanishes in all directions in the tangent space at these points. If this weren't the case then the splitting lemma could be used to eliminate the directions along which the curvature does not vanish.

The origin of the germs in Table 3.1 of corank 1 are straightforward to derive, since near any critical point in one dimension the Taylor expansion of the function has the form $ax^n + b$. In two–dimensions, the Taylor expansion about any degenerate third order critical point has the form

$$t(x, y) = (u_1 x + v_1 y)(u_2 x + v_2 y)(u_3 x + v_3 y), \tag{3.14}$$

where we have neglected the trivial zero order term in the expansion. For the moment we assume that the coefficients are real and no two ratios u_i/u_j are equal. Equation (3.14) may be simplified by applying the linear transformation $(x, y) \mapsto (u_1 x' + u_1 y', u_2 x' + u_2 y')$ to t giving

$$\begin{aligned} t(x, y) &= xy[u_3(u_1 + u_2)x + v_3(v_1 + v_2)y] \\ &\sim x^2 y + xy^2. \end{aligned} \tag{3.15}$$

After an additional transformation $(x, y) \mapsto (x+y, x-y)$ we get the elliptic umbilic catastrophe germ $x^3 - xy^2$ (Saunders, 1980). Note that the polynomial $x^2y + xy^2$ is equally valid as a germ to classify the elliptic umbilic; the only difference being that the form of the unfolding would consist of a different polynomial. The former is simply used for convenience. The germs of the other umbilic catastrophes are found when we relax the restriction on the coefficients in Eq. (3.14) to allow for complex values, corresponding to the hyperbolic umbilic catastrophe, For non–unique ratios, we observe the parabolic umbilic catastrophe.

The canonical form of the elliptic umbilic catastrophe is given by the family of functions

$$f(x, y; u, v, w) = \frac{1}{3}x^3 - xy^2 + w(x^2 + y^2) + ux + vy. \tag{3.16}$$

Either x^2 or y^2 may be used as the quadratic unfolding term in the w–direction. Since there is no distinction between either direction both are typically included, which makes the bifurcation set symmetric about that unfolding axis. It also serves to make S and B symmetric about the w–axis. The factor of 1/3 is used to ensure that the bifurcation set is symmetric. The equilibrium surface is given by the following set of elliptic equations:

$$0 = x^2 - y^2 + 2wx + u, \tag{3.17}$$

$$0 = -2xy + 2wy + v. \tag{3.18}$$

The S set is found from setting the Hessian determinant to zero, giving

$$x^2 + y^2 = w^2. \tag{3.19}$$

Equation (3.19) represents a double cone centred on the w–axis in a (x, y, w) coordinate system. In these coordinates we eliminate the state variables by parameterising Eq. (3.19) in terms of the polar angle θ using $x = \cos\theta, y = \sin\theta$. The B set is found by substituting this parameterization into Eqs. (3.17) and (3.18) and then making use of trigonometric double angle formulas, giving

$$u = -w^2(\cos 2\theta + 2\cos\theta) \tag{3.20}$$

$$v = w^2(\sin 2\theta - 2\sin\theta) \tag{3.21}$$

$$w = w. \tag{3.22}$$

The geometry of B is shown in Fig. 3.3. The three cusp points occur at the angles where the partial derivatives of u and v vanish. The cusp points follow a parabolic curve with increasing

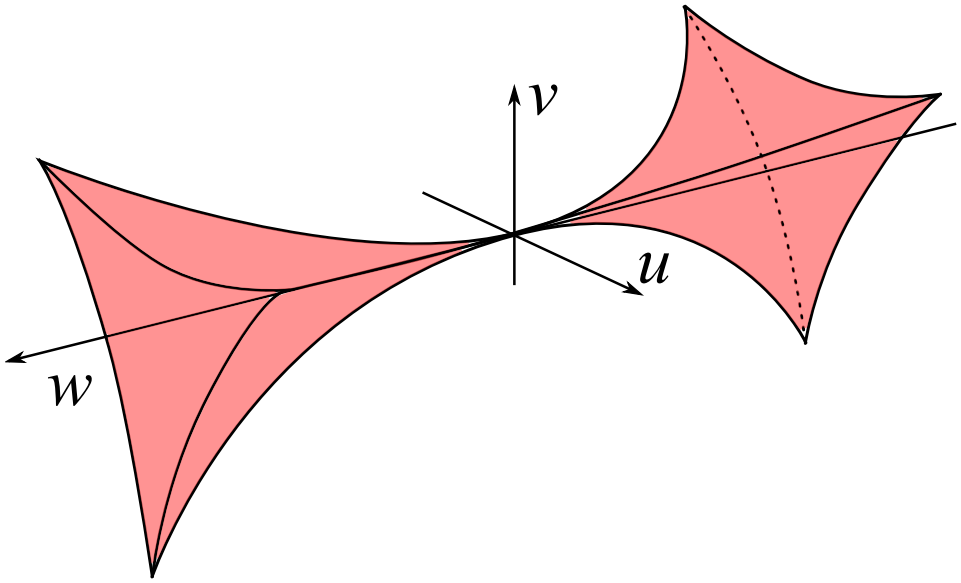

Figure 3.3: The bifurcation set of the elliptic umbilic catastrophe.

w. Note that in Eqs. (3.20) there is still an implicit dependence on the state variables due to the presence of θ. Thus θ is treated as a parameter when describing the set B for the elliptic umbilic catastrophe.

3.6 Diffraction Catastrophes

The utility of catastrophe theory lies in its applicability to a diverse range of physical systems. The formation of caustic surfaces in optics is an important example of how catastrophe theory manifests itself. To develop the connection between optics and catastrophe theory, we start with the observation that optical caustics are equivalent to the bifurcation sets of the elementary catastrophes; consequently they may be classified according to Thom's theorem. Because we are limited to three–spatial dimensions, it follows that caustics exist in the control space associated with a 3–parameter family of functions. The family of functions to which the caustic belongs is then classified according to Table. 3.1.

To show this explicitly, consider the paraxial propagation of the spatial part of a monochromatic complex scalar wavefunction ψ, with uniform intensity and phase modulation denoted by f. Let the boundary be located at $(x, y, z = 0)$ and the wave–vector k, be aligned with the z–axis. The wave field at the point (X, Y, Z) in the half space $z > 0$ is given by the convolution form of the Fresnel diffraction integral (Winthrop and Worthington,

1966):

$$\begin{aligned}\psi(X,Y,Z) &= -\frac{ik}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\{ik\left[Z-f(x,y)\right]\}\exp\left\{\frac{ik}{2Z}\left[(x-X)^2+(y-Y)^2\right]\right\}\mathrm{d}x\,\mathrm{d}y \\ &= -\frac{ik}{2\pi}\exp\left[ik\left(Z+\frac{1}{2Z}\left(X^2+Y^2\right)\right)\right] \\ &\qquad\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left\{-ik\left[f(x,y)-\frac{1}{2Z}\left(x^2+y^2\right)+\frac{X}{Z}x+\frac{Y}{Z}y\right]\right\}\mathrm{d}x\,\mathrm{d}y.\end{aligned} \tag{3.23}$$

The exponential multiplying the double integral describes an expanding spherical envelope of the propagating wavefront. Following Nye (1999), it is convenient to account for this additional parabolic term in the exponential by adding a similar reference parabola to f, i.e.,

$$f(x,y) = g(x,y) + \frac{1}{2Z_0}\left(x^2+y^2\right), \tag{3.24}$$

where Z_0 is the centre of curvature of the reference parabola (see Fig. 3.4). The phase of the exponential inside the integral in Eq. (3.23) now becomes

$$\phi(x,y;X,Y,Z) = g(x,y) + \frac{1}{2}\left(\frac{1}{Z_0}-\frac{1}{Z}\right)\left(x^2+y^2\right) + \frac{X}{Z}x + \frac{Y}{Z}y. \tag{3.25}$$

We make the following transformation to new coordinates (ξ,η,ζ) defined by:

$$\xi = \frac{X}{Z}, \tag{3.26}$$

$$\eta = \frac{Y}{Z}, \tag{3.27}$$

$$\zeta = \frac{1}{2}\left(\frac{1}{Z_0}-\frac{1}{Z}\right). \tag{3.28}$$

Substituting Eqs. (3.25), (3.26), (3.27) and (3.28) into Eq. (3.23), we obtain

$$\psi(\xi,\eta,\zeta) = -\frac{ik}{2\pi}\exp\left[ikp(\xi,\eta,\zeta)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp[-ik\phi(x,y;\xi,\eta,\zeta)]\mathrm{d}x\,\mathrm{d}y, \tag{3.29}$$

where p is a real function and

$$\phi(x,y;\xi,\eta,\zeta) = g(x,y) + \zeta\left(x^2+y^2\right) + \xi x + \eta y. \tag{3.30}$$

Equation (3.30) represents a 3–parameter family of 2–dimensional functions, where the wavefront function g, represents the germ. We identify the variables (x,y) as the state

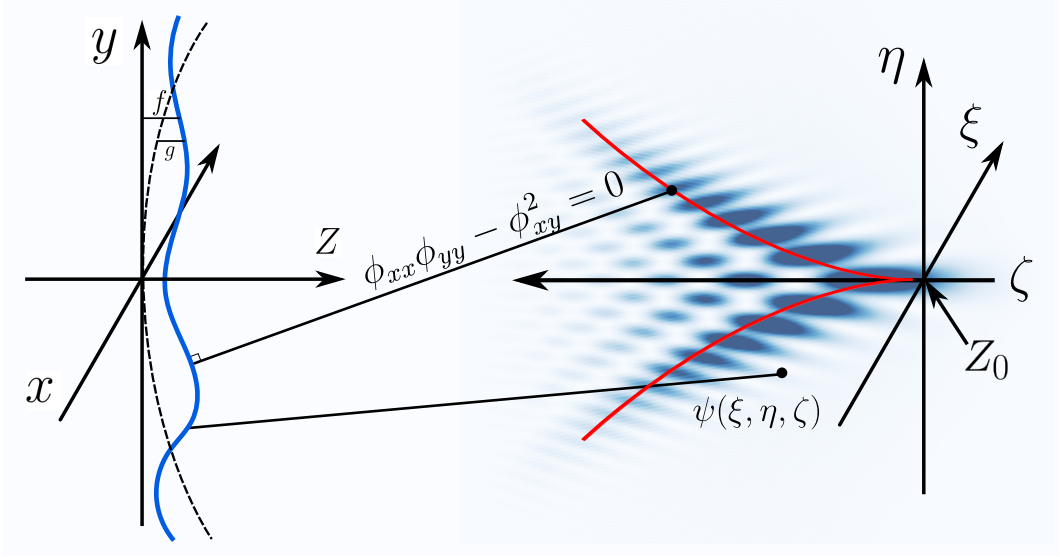

Figure 3.4: An initial wavefront $f(x, y)$, shown in blue, is mapped to the control space, forming a cusp diffraction catastrophe.. The bifurcation set of the cusp, given by Eq. (3.13), is outlined in red and marks the centre of curvature of the wavefront as it propagates near Z_0. All rays that satisfy the catastrophe condition, given by Eq. (3.36), map to the red curve. The bifurcation set is littered with intricate detail due to the wave nature of the field, which is described by Eq. (3.29).

variables and the variables (ξ, η, ζ) as the control parameters. We can view Eq. (3.23) or Eq. (3.29) as a form of projection mapping, which takes an initial wavefront in the state space (x, y) to the control space (ξ, η, ζ). Equation (3.29) is the main result of this section. If ϕ is equivalent to the canonical form of an elementary catastrophe in Table. 3.1, then the diffracted wavefront ψ will form a diffraction catastrophe. The local centre of curvature of the propagating wavefront has an equivalent geometry to the bifurcation set of the catastrophe.

Diffraction catastrophes constitute a specific example of Thom's theorem, which always consists of a two–dimensional state space mapped to a three–dimensional control space. The propagation of a wavefront via the catastrophe integral in Eq. (3.29) is a versal unfolding of the germ specified by the initial wavefront, with the unfolding terms being $x^2 + y^2$, x and y.[5]

3.6.1 Caustics are singularities of gradient maps

Here we consider how the diffraction catastrophes reduce to caustics as we transition from waves to rays. By moving to a ray–based description we assume $k \to \infty$, which causes the

[5]Coincidently this unfolding is a universal unfolding for a specific representation of the elliptic and hyperbolic umbilic catastrophes.

integrand in the diffraction integral in Eq. (3.29) to become highly oscillatory. Consequently, only significant contributions to the value of ψ will come from points where the phase is stationary, namely the critical points of Eq. (3.30). This asymptotic approximation to the diffraction integral is determined using the method of stationary phase (Walther, 1995; Born and Wolf, 1999).

To solve Eq. (3.29) in the ray approximation we first find where Eq. (3.30) is stationary, i.e.,

$$\nabla_{\perp}\phi(x, y; \xi, \eta, \zeta) = 0. \tag{3.31}$$

Equation (3.31) is often referred to as the ray condition, since it defines vectors normal to the planes of constant phase, which specify the propagation direction of each ray. Given some observation point P in control space, the number of rays contributing to the intensity at P is found by substituting Eq. (3.30) into the ray condition:

$$\frac{\partial g}{\partial x} + 2\zeta x + \xi = 0, \tag{3.32}$$

$$\frac{\partial g}{\partial y} + 2\zeta y + \eta = 0, \tag{3.33}$$

which results in

$$\xi = -\frac{\partial g}{\partial x} - 2\zeta x, \tag{3.34}$$

$$\eta = -\frac{\partial g}{\partial y} - 2\zeta y. \tag{3.35}$$

These two equations describe the ray path originating in state space and propagating to an observation plane at a constant ζ in control space. This is a form of gradient mapping of the initial wavefront g. However, Eqs. (3.34) and (3.35) only describe how a ray maps to control space, but does not provide any information as to whether a ray belongs to a family whose envelope forms a caustic. The caustic envelope is comprised of points to which multiple rays contribute. Hence a caustic appears in the mapping of rays when Eqs. (3.34) and (3.35) become singular (Berry and Upstill, 1980). Points where this occurs correspond to where the Jacobian determinant of the gradient mapping vanishes. Since Eqs. (3.34) and (3.35) define a mapping $(x, y) \mapsto (\xi, \eta)$, the vanishing of the Jacobian determinant is given by:

$$\begin{vmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} \end{vmatrix} = \begin{vmatrix} -g_{xx} - 2\zeta & g_{xy} \\ g_{xy} & -g_{yy} - 2\zeta \end{vmatrix} = 0, \tag{3.36}$$

where the subscripts indicate partial differentiation. This is known at the caustic condition, since rays that satisfy this condition form a family of rays that contribute to a caustic. By using the form of the ray condition in Eq. (3.31), the caustic condition may be represented by a vanishing Hessian determinant of ϕ:

$$\phi_{xx}\phi_{yy} - \phi_{xy}^2 = 0. \tag{3.37}$$

This caustic condition illustrates the connection between catastrophe theory and the formation of caustic and diffraction catastrophes. It shows that optical catastrophes originate from unstable degenerate critical points in the phase of the diffraction integral.

Although the form of the phase in Eq. (3.30) was derived from wave–theory, an identical result can also be derived from a ray theory formulation of propagation. In this context ϕ represents a distance function between points in the state and control space. Equation (3.30) may be derived by extremising the action of the path along which a ray travels between the two spaces (Berry and Upstill, 1980; Nye, 1999). This is equivalent to Fermat's principle in optics, with ϕ constituting a Fermat–type potential function.

4 Singularimetry using an Optical Vortex Lattice

Interferometry is commonly used for wavefront determination in visible–light optics (Heflinger et al., 1966), cold atom optics (Zhan et al., 2007), electron optics (Tonomura, 1999), and X-ray microscopy (Bonse and Hart, 1965). A two-beam interferogram is formed by interference between a wave that has passed through a specimen and a reference wave that does not interact with the specimen. Fourier-transform methods are commonly used to reconstruct the intensity and phase of the wave from an interferogram (Takeda et al., 1982). Although Fourier transform methods are simple to implement, the spatial resolution of the reconstruction is limited by how isolated the side–lobes of the spectrum are from the central maximum. These methods also involve inverting a subsection of the Fourier transform of the interferogram. Since a specific region of the desired real-space phase map cannot be localized in Fourier space, the phase is reconstructed non–locally. For example, poor fringe visibility in a subsection of the interferogram can introduce systematic errors in the recovered wave over the entire field of view. Further, Gibbs–type "ringing" can occur when processing subsections of interferograms that exhibit strong scattering or absorption, leading to poor contrast resolution.

The technique presented in this chapter uses interference of three plane waves to generate a uniform lattice of optical vortices, which are used to recover the phase of the input wave field. Perturbations to the input wave field, due to an object, distort the vortex lattice. The resulting transverse displacements of the vortices are then used to determine the phase at the "exit surface" of the object. The "exit surface" is the plane where the exit wave resides once it has propagated through the object. We show that the phase is proportional to the Cartesian components of the transverse displacements of each vortex in the lattice, thereby permitting

the phase to be retrieved locally without introducing "ringing" artefacts. Further, the method is also applicable to absorbing objects or reference waves with different intensities. This work has been published in Eastwood et al. (2012), a copy of which may be found in Appendix C.

4.1 Relation between Vortex Position and the Phase of the Wave Field

Consider again the superposition of three coherent plane scalar waves of unit intensity, given by Eq. (2.3). Suppose now that one of these plane waves traverses a phase–amplitude object, under the projection approximation. The wavefunction at the exit surface ($z = 0$) is then given by the superposition of the object wave and two reference waves:

$$\begin{aligned}\Psi(x, y, z = 0) &= \psi_{obj}(x, y) + \psi^{(1)}(x, y) + \psi^{(2)}(x, y) \\ &= A(x, y)e^{i\mathbf{k}_\perp^{(0)}\cdot\mathbf{x}_\perp + i\phi(x,y)} + \rho_1 e^{i\mathbf{k}_\perp^{(1)}\cdot\mathbf{x}_\perp} + \rho_2 e^{i\mathbf{k}_\perp^{(2)}\cdot\mathbf{x}_\perp} \\ &= e^{i\mathbf{k}_\perp^{(0)}\cdot\mathbf{x}_\perp}\left\{A(x, y)e^{i\phi(x,y)} + \rho_1 e^{i\mathbf{k}_\perp^{A}\cdot\mathbf{x}_\perp} + \rho_2 e^{i\mathbf{k}_\perp^{B}\cdot\mathbf{x}_\perp}\right\}, \end{aligned} \tag{4.1}$$

where $A(x, y)$ and $\phi(x, y)$ represent the real–valued amplitude and phase changes imparted to the object wave by the phase–amplitude object. The remaining terms have been defined in Eq. (2.3). The trivial harmonic time dependence $\exp(-i\omega t)$, where ω is the angular frequency and t is time, is suppressed throughout. The additional phase and amplitude modulation perturbs the otherwise uniform lattice of vortices. Here we show that both the phase and amplitude of the object's complex transmission function may be measured from the displacement of the vortices.

4.1.1 Non–absorbing object

We first consider the case of a wave traversing a phase object, assuming that the three waves have equal intensity. Consider the phasor diagram representing the superposition of each wave at an intensity zero. The object wave phasor undergoes an Argand–plane rotation by an angle equal to the phase of the object. Each position of a vortex in the perturbed lattice is denoted by $(\tilde{x}, \tilde{y})$. Because the amplitude of each wave is identical, the phasors must form an equilateral triangle, since the field vanishes at a vortex core. At each vortex position the value of the phase of the two reference arms, labelled α and β, will be $2\pi/3$ and $4\pi/3$, plus

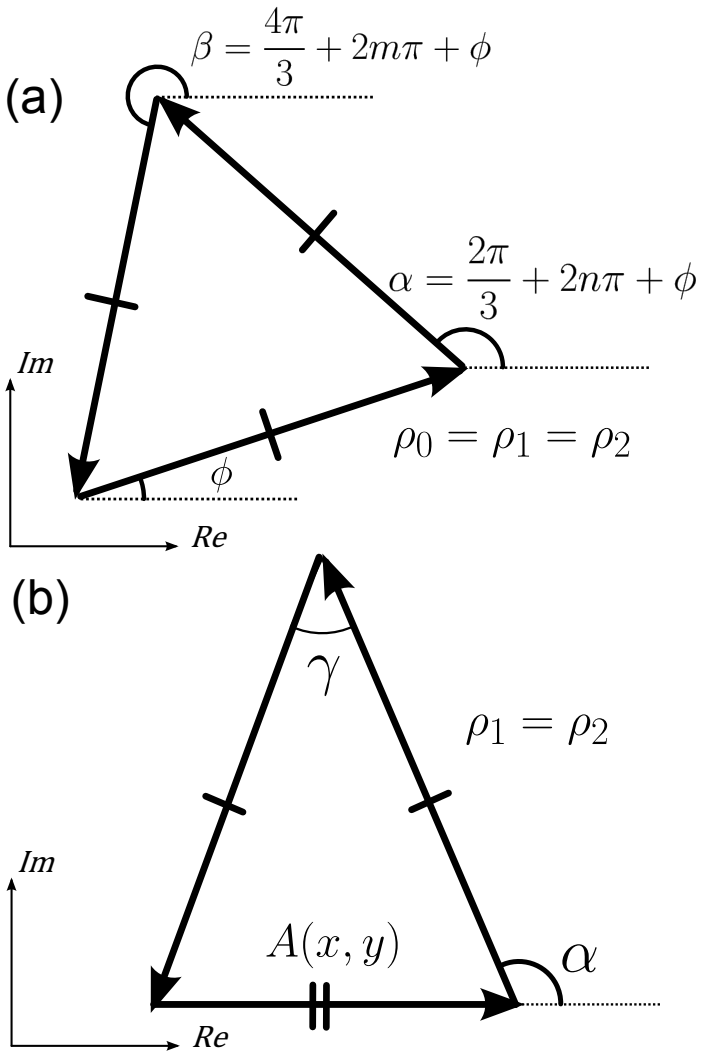

Figure 4.1: (a) The vector addition of three plane wave phasors. The length and angle of each phasor corresponds to the amplitude and phase, respectively. At a vortex core these phasors must form an equilateral triangle. The phasor geometry is rotated when the object wave undergoes a spatially dependent phase shift due to traversing the phase object. (b) Phasor arrangement when the object wave undergoes amplitude modulation.

the additional angle of rotation, thus:

$$(k_x^A \tilde{x} + k_y^A \tilde{y}) = \frac{2\pi}{3} + 2\pi n + \phi(\tilde{x}, \tilde{y}), \tag{4.2}$$

$$(k_x^B \tilde{x} + k_y^B \tilde{y}) = \frac{4\pi}{3} + 2\pi m + \phi(\tilde{x}, \tilde{y}), \tag{4.3}$$

where n and m are integers. Since the relative angles between all phasors remain constant at each vortex core, we are able to substitute Eq. (4.2) into Eq. (4.3), giving:

$$2\arg(\psi^{(2)}) = \arg(\psi^{(1)}). \tag{4.4}$$

Solving for the phase of the object we find:

$$\phi(\tilde{x},\tilde{y}) = (2k_x^A - k_x^B)\tilde{x} + (2k_y^A - k_y^B)\tilde{y} + 2\pi l, \tag{4.5}$$

where l is an integer. Equation (4.5) shows that the phase shift in the exit wave field depends directly on the coordinates of each vortex, up to an additive constant $2\pi l$, making the phase modulo 2π. Therefore, the phase of the object wave may be retrieved from the coordinates of each vortex in the distorted lattice along with knowledge of the transverse wave–vectors of each reference wave.

A second triangle is formed by complex conjugation of all three phasors, corresponding to the conjugate phasors of the two reference arms in Eq. (4.1). This dual arrangement corresponds to an anti–vortex having the opposite topological charge. A modified form of Eq. (4.5) is easily derived, which allows us to measure the phase from vortices of the opposite sign. Hence the phase of the object is related to the position of each anti–vortex by:

$$\phi(\tilde{x}',\tilde{y}') = (2k_x^B - k_x^A)\tilde{x}' + (2k_y^B - k_y^A)\tilde{y}' + 2\pi p, \tag{4.6}$$

where p is an integer, and $\tilde{x}'$ and $\tilde{y}'$ denote the Cartesian coordinates of each anti–vortex. Equation (4.6) differs from Eq. (4.5) in that the wave–vector of each reference wave is interchanged.

4.1.2 Absorbing object

Here we generalize the analysis in §4.1.1 to the case of an absorbing object. An absorbing object causes the length of the object phasor to reduce, as shown in Fig. 4.1(b). Equations (4.2) and (4.3) no longer hold in this instance, as the relation $2\arg(\psi^{(2)}) = \arg(\psi^{(1)})$ no longer applies at the positions of each vortex. To find the relationship between the object's phase and vortex positions, we note that all the exterior angles of the phasors must sum to 2π; due to the symmetry of the phasor arrangement, $\alpha + \beta = 2\pi$ is valid at each vortex position. By factoring out the first term in braces in Eq. (4.1), and substituting the arguments of each reference wave into α and β we find:

$$k_x^A\tilde{x} + k_y^A\tilde{y} + k_x^B\tilde{x} + k_y^B\tilde{y} - 2\phi(\tilde{x},\tilde{y}) = 2\pi(1 - n - m). \tag{4.7}$$

Solving for the phase gives:

$$\phi(\tilde{x},\tilde{y}) = \frac{1}{2}(k_x^A + k_x^B)\tilde{x} + \frac{1}{2}(k_y^A + k_y^B)\tilde{y} + \pi s, \tag{4.8}$$

where s is an integer. Thus the phase is directly proportional to the vortex locations, modulo π. Equations (4.5) and (4.8) are the two main results of this chapter, and form the basis of our method of phase determination.

To determine the phase we proceed as follows. First we localize all vortices in the lattice, resulting from a three–beam interference pattern, and determine their $(\tilde{x}, \tilde{y})$ coordinates by locating intensity zeros. Using either Eq. (4.5) or Eq. (4.8), we algebraically calculate ϕ for each vortex point $(\tilde{x},\tilde{y})$. The use of either equation is dependent on the object being imaged, i.e. for absorbing objects Eq. (4.8) is used, while for non–absorbing objects Eq. (4.5) is used. If Eq. (4.5) is utilized, an additional step is required in which we separate vortices by the sign of their topological charge. The wrapped phase is then recovered modulo 2π. For absorbing specimens, the phase is wrapped modulo π, as dictated by Eq. (4.8). Once the phase is computed, interpolation between each vortex is performed to recover the phase over all (x, y) points in the image array.

Note that Eq. (4.8) is applicable for strongly absorbing objects. The magnitude of the phasor corresponding to the object beam decreases with increasing absorption; unless the object is completely absorbing, the phasors will still maintain their equilateral triangular geometry, which is required for vortex formation. However, strongly absorbing objects will decrease the contrast resolution of the interference pattern. Moreover, low contrast can make it difficult to accurately localize a vortex, which requires the detection of minima in the intensity pattern. However, there are phase shifting methods that use an extra reference arm to form a fringe pattern, which can be exploited to increase the intensity gradient near a vortex core. Phase shifting the reference arm to generate a fringe pattern causes shifts in the intensity maxima, whilst the intensity minima due to vortices remain fixed at the same locations. By taking a series of interferograms, each with the fourth interferometer arm phase shifted, the absolute value of the difference between successive interferograms is calculated. These differences are summed, which increases the local intensity gradient near the vortex core. This compensates for the lack of contrast in the three-beam interferogram of a highly absorbing object (Popiołek-Masajada and Frączek, 2011). Such an approach could be used to increase the precision with which vortices can be localized in those cases where the absorption is too large to accurately locate each vortex.

The phasor geometry used to derive Eq. (4.5) only considered a single sign of the

topological charge. Therefore only the coordinates of one particular type of vortex are used to determine the phase with Eq. (4.5). This is not the case for Eq. (4.8), where two arrangements of phasors sum to zero, corresponding to positive and negative topological charges. This is due to the symmetry in the coefficients of $\tilde{x}$ and $\tilde{y}$ in Eq. (4.8), which is not present in Eq. (4.5). Thus, in determining the phase, Eq. (4.8) allows us to utilize either vortex sign in the interference pattern. However, there is a trade off; namely, the third term in Eq. (4.8) implies that the phase can only be retrieved modulo π rather than modulo 2π. However, this is only a disadvantage for objects that have large phase gradients, in which case aliasing may occur and phase unwrapping is difficult. In most situations the use of Eq. (4.5) is still preferred.

4.1.2.1 Measurement of the attenuation of an object

Using the geometry of the phasor diagram in Fig. 4.1(b), the value of the object's attenuation coefficient may also be determined from the vortex displacements. Using the law of cosines, the length of the object phasor can be related to the angle γ by

$$A^2(x, y) = 2 - 2\cos(\gamma), \tag{4.9}$$

where we have normalised the amplitudes of the reference waves to unity. Using the phasor symmetry $\gamma = 2\alpha - \pi$ and substituting in Eq. (4.8) for the phase dependence in α, we arrive at:

$$A(x, y) = \sqrt{2 + 2\cos\left(\mathbf{k}_\perp^A \cdot \tilde{\mathbf{x}}_\perp + \mathbf{k}_\perp^B \cdot \tilde{\mathbf{x}}_\perp\right)}. \tag{4.10}$$

Thus, assuming a single–material object, the attenuation of the object wave may be measured directly from the location of each vortex having either topological sign. If the object's coefficient of attenuation μ is known, Eq. (4.10) could then be used to determine the projected thickness $T(x, y)$ of the object via the Beer–Lambert law, i.e., $A(x, y) = \exp[-\mu T(x, y)]$. Equation (4.10) also shows that amplitude retrieval is independent of the effect of the object's phase on the locations of the vortices. This is possible as the displacement of a vortex, due to an amplitude gradient, may be directionally uncoupled from its displacement due to a phase gradient.

4.1.3 Dynamics of the vortex lattice

In §2.3.1 we showed that a vortex perturbed by a phase gradient moves in a direction perpendicular to one perturbed by an amplitude gradient. However, this is not the case for our vortex lattice, which is formed after the wave has interacted with the object. In this situation the vortices are not directly perturbed by the object itself, which leads to a different relationship between vortex displacement, due to amplitude and phase gradients at the object's exit surface.

The displacement vector of a vortex due to a phase gradient, denoted by $\mathbf{\Delta}_\phi$, is found by differentiating Eqs. (2.12) and (2.13) with respect to the phase, ϕ. After setting $\phi_A = \phi_B = -\phi$, we find

$$\begin{aligned}\mathbf{\Delta}_\phi &= \left(\frac{\mathrm{d}\tilde{x}}{\mathrm{d}\phi}, \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\phi}\right) \\ &= \frac{1}{k_x^A k_y^B - k_y^A k_x^B}\left(k_y^B - k_y^A, k_x^A - k_x^B\right). \end{aligned} \tag{4.11}$$

An absorbing object changes the length A, of the phasor of the object wave (see Fig. 4.1). The dependence of Eqs. (2.12) and (2.13) on A is implicitly contained in their dependence on α and β. The relationship between A and α is given by the cosine rule and the identity $\gamma = 2\alpha - \pi$, i.e.,

$$\alpha = \frac{1}{2}\arccos\left(1 - \frac{A^2}{2}\right). \tag{4.12}$$

The equations governing the positions of the vortices are expressed solely in terms of α by making use of $\alpha + \beta = 2\pi$. Upon substituting Eq. (4.12) into Eqs. (2.12) and (2.13), the position of each vortex is differentiated with respect to A to give the displacement of the vortex due to an amplitude gradient, $\mathbf{\Delta}_\mathbf{A}$:

$$\begin{aligned}\mathbf{\Delta}_A &= \left(\frac{\mathrm{d}\tilde{x}}{\mathrm{d}A}, \frac{\mathrm{d}\tilde{y}}{\mathrm{d}A}\right) \\ &= \frac{(4 - A^2)^{-1}}{k_x^A k_y^B - k_y^A k_x^B}\left(k_y^B + k_y^A, -k_x^A - k_x^B\right). \end{aligned} \tag{4.13}$$

Equations (4.11) and (4.13) show that the dynamics of the vortices depends on the relative difference of the in–plane projections of the wave–vectors. Interestingly, however, because of the additional factor $(4 - A^2)^{-1}$ in Eq. (4.13), the shift in a vortex's position, due to a phase gradient, is always greater than the shift due to an amplitude gradient. This additional factor arises because vortices do not form when the magnitude of the object wave phasor is

greater or equal to the sum of the magnitudes of the phasors of the two reference waves (i.e., when $|\psi_{obj}| \geq |\psi^{(1)}| + |\psi^{(2)}|$). The conditions under which $\mathbf{\Delta}_A$ and $\mathbf{\Delta}_\phi$ are linearly independent is found by taking the scalar product:

$$\mathbf{\Delta}_\phi \cdot \Delta_A = \frac{\left|\mathbf{k}_\perp^B\right|^2 - \left|\mathbf{k}_\perp^A\right|^2}{(4 - A^2)(k_x^A k_y^B - k_y^A k_x^B)^2}. \tag{4.14}$$

From Eq. (4.14) we make the following observations. The displacements of a vortex due to an amplitude or phase change are perpendicular to one another when each of the two reference arms have equal angles with respect to the optics axis. The displacement vectors are collinear when $|\mathbf{k}_\perp^A|^2 = |\mathbf{k}_\perp^B|^2$. It follows that the shifts due to phase are independent of the relative azimuthal angle of each wave and depend only on the wave's tilt, whilst the opposite is true for the vortex displacement due to a phase gradient Dennis and Götte (2012).

4.2 Experimental Implementation

A schematic and photograph of the experimental setup is shown in Fig. 4.2. The beam from a linearly-polarized Helium-Neon laser (Thorlabs 5 mW) is spatially filtered by focusing it through a 4.51 mm focal length aspheric lens and then passing it through a 20 μm pinhole. The filtered beam is re–collimated using a 100 mm focal length plano-convex lens and an iris, which is adjusted so that the beam is truncated at its first minimum. A neutral density filter located before the first focusing lens is used to attenuate the beam. The filtered beam is then passed through a polarizing beamsplitter cube, oriented to transmit the majority of the beam power; this ensures a pure polarization state for the resulting beam. The beam is then passed into a three–beam interferometer constructed from a pair of Mach-Zehnder interferometers that share a common arm. The path lengths through each arm of the interferometer are matched to maximise the coherence of the interference; 50:50 beamsplitter cubes are used to split and combine beams in each interferometer.

The common central arm of the combined interferometer (object beam) is used as the source of illumination for the object. The object is located close to the first beamsplitter cube in the arm and is imaged onto a CCD camera via a system of two plano-convex lenses arranged to image the object plane onto the camera with a magnification of two. The lenses are arranged so that the first lens is located one focal length away from the object ($f = 75$ mm) with the second lens ($f = 150$ mm) spaced 225 mm away from the first lens

and 150 mm away from the camera. The monochrome CCD camera (Prosilica GE1650) has 1600×1200 pixels and 12 bits per pixel.

The pixel size is 7.4 μm × 7.4 μm. The two remaining interferometer arms (reference beams) each contain a neutral density filter to match the intensity of the beam through each arm to that of the object beam. A $\lambda/2$ and a $\lambda/4$ wave plate allows the phase of the beam exiting each arm to be varied in discrete steps. The phase is stepped by either $\lambda/2$ or $\lambda/4$ (depending on which wave plate is used) by changing the alignment of the wave plate so that the fast optical axis is either parallel or perpendicular to the polarization axis of the beam.

To form a vortex lattice, two reference beams are adjusted in angle relative to the object beam, such that the wave–vectors of the three beams are non-coplanar. The angles of the beams of the reference arm can be adjusted by tilting the final beamsplitter in the top arm of the interferometer and the final mirror in the bottom arm. Different geometries of the lattice are generated depending on the mutual angle of the three beams, with a hexagonal lattice being produced when the angle between any two beams is 120 degrees. The period of the maxima in the three–beam interferogram was adjusted to be approximately 20 pixels. A phase shift of either reference beam causes a shift in the vortex in the direction of that beam's transverse wave–vector.

Our technique was experimentally tested on two objects: a Thorlabs spherical lens (part number LA1464) made from N–BK7 optical glass with a diameter of 25.4 mm, and the wing of a common house fly (Musca Domestica). For the spherical lens, in addition to the three–beam pattern, two–beam interference patterns of $\psi_{obj} + \psi^A$ and $\psi_{obj} + \psi^B$ were individually acquired by blocking each of the reference arms in turn. Both reference arms were then blocked to record an image of the Gaussian illumination. Wave–vectors of each arm were measured by locating the peak maximum in the power spectrum of the three–beam interference pattern. The interferograms were smoothed by convolution with a Gaussian filter of approximately 60 μm full width at half maximum (FWHM) to reduce noise and then flat–field corrected using a specimen–free illumination image. This was performed to increase the signal in the outer edges of the interferograms.

To determine the accuracy with which vortices can be localized we computed the normalized autocorrelation of a non-deformed section of our vortex lattice. Since the lattice is uniform, perfect localization gives complete autocorrelation between vortex points (i.e.,

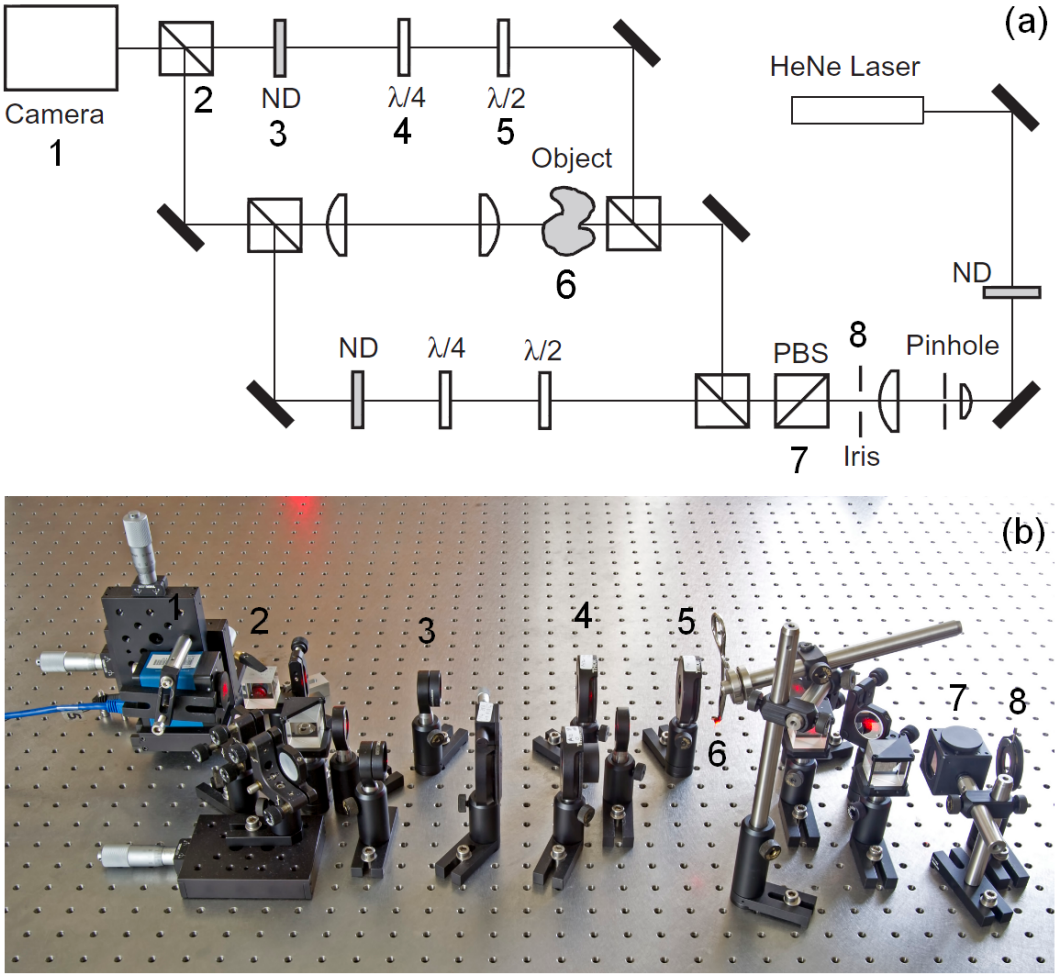

Figure 4.2: (a) Schematic and (b) Photograph of the experimental setup of the three–beam vortex interferometer. The optical elements are numbered similarly for (a) and (b). A He–Ne laser (λ = 632 nm) is spatially filtered by the pinhole and a single polarization direction is selected as it passes through the polarizing beam splitter. Two beamsplitters are used to create the three arms of the interferometer. The central beam is transmitted through an object after which two lenses are used to focus the exit surface of the object into the camera. Each reference arm contains $\lambda/2$ and $\lambda/4$ wave plates, which are used for phase stepping. A neutral density filter (ND) ensure that the two reference waves have the same intensity as the object wave.

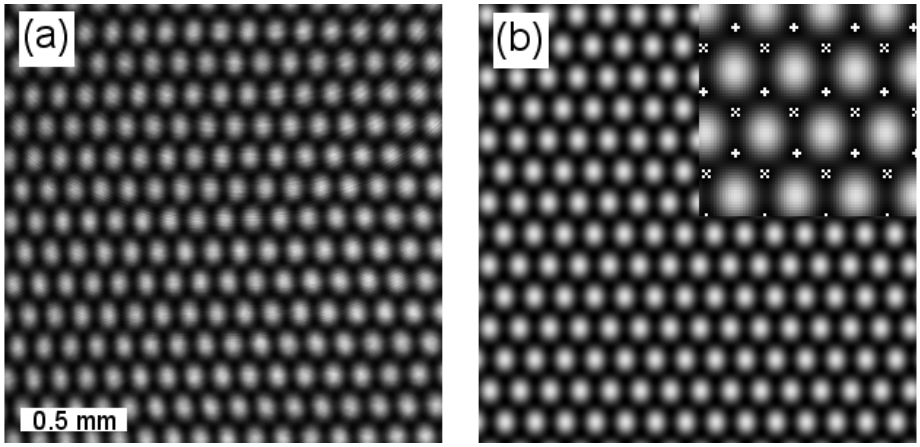

Figure 4.3: Comparison of the experimental three–beam interference pattern with a numerical simulation. (a) Experimental three–beam interference pattern. (b) Simulated three–beam interference pattern. The plus signs and crosses in the top right inset indicate the locations of vortices and anti–vortices, respectively.

the autocorrelation value is unity). However, small errors in localization decrease the degree of autocorrelation for the vortex lattice, giving some width to the central maximum of the autocorrelation function. Using the half width of the maximum as a measure of the precision of vortex localization, it was found that the technique has a precision of 1.5 pixels. Based on Eq. (4.5) this translates to a measured phase accuracy of ±0.2 radians.

4.3 Results

4.3.1 Phase measurement of a lens

A section of the three–beam interferogram of the lens along with a simulated interferogram is shown in Fig. 4.3. Excellent agreement is found between theory and experiment. To locate the vortices in the image, it is only necessary to locate the intensity zeros in the interference pattern. A zero in the intensity is a necessary but not sufficient condition for a vortex; a three–beam interference pattern of plane waves of equal intensity will always have a vortex at each intensity zero (Paganin, 2006). In this context non–vortical zeros are not stable with respect to perturbation, whereas vortical zeros are stable. The vortices were separated depending on the sign of their topological charge using the methods outlined in Frączek et al. (2005). The phase of the lens was calculated using Eq. (4.5) for vortices of positive topological charge. Two–dimensional linear interpolation was used between each

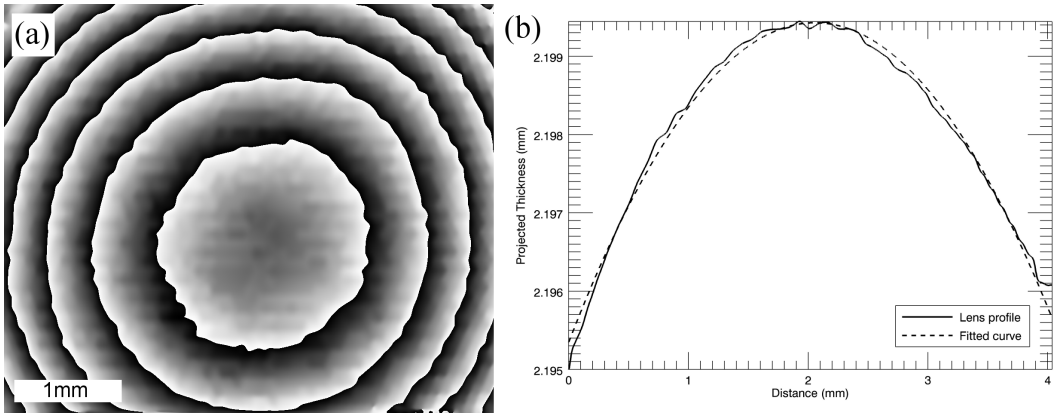

Figure 4.4: Experimental results for phase reconstruction of a spherical lens using the three–beam vortex interferometer. (a) The unwrapped reconstructed phase of the 1 m focal length lens. The greyscale in this image ranges from 0 (black) to 2π (white). (b) Unwrapped phase profile of the lens. The solid curve corresponds to the experimental data whilst the dashed curve is the fitted function. The radius of curvature of the fitted profile is 519±1 mm.

vortex point to map the phase over all Cartesian pixel coordinates within the field of view. Interpolation was performed on the real and imaginary parts of the wavefunction separately. This allowed the complete wrapped phase (modulo 2π) to be recovered using

$$\phi(x, y) = \tan^{-1}\left\{\frac{\sin[\phi(\tilde{x}, \tilde{y})]}{\cos[\phi(\tilde{x}, \tilde{y})]}\right\}. \tag{4.15}$$

The reconstruction of the lens in Fig. 4.4(a) shows a series of concentric circles where the phase has been wrapped by 2π, as would be expected for a spherical lens since its projection is rotationally symmetric.

To demonstrate that our technique is quantitative, the wrapped lens phase was unwrapped using the method described in Volkov and Zhu (2003). The gradient of the wrapped phase suffers abrupt changes of 2π per pixel at each point where the phase undergoes wrapping. To unwrap the phase, the first order partial derivatives of the wrapped phase were calculated and then a threshold was used to removed the peaks caused by wrapping. The threshold pixels are then set to a value equal to the average of the 8 adjacent pixels, giving a smooth gradient that is equivalent to the gradient of the unwrapped phase. Volkov and Zhu (2003) showed that the unwrapped phase can be recovered by making use of FFT's, i.e.,

$$\phi(x, y) = \mathrm{Re}\left\{\frac{1}{2\pi i}\mathcal{F}^{-1}\left[\frac{\mathcal{F}(\partial_x\phi)u + \mathcal{F}(\partial_y\phi)v}{u^2 + v^2}\right]\right\}, \tag{4.16}$$

where $\mathcal{F}$ denotes a Fourier transform with the conjugate variables (u, v).

Once the lens data had been unwrapped, the projected thickness was recovered by multiplying the unwrapped phase by the factor $-\delta k$, where δ is the difference in the refractive index of the lens from unity and k is the wavenumber. A circular arc was fitted to the profile centred on the recovered projected thickness in order to measure the radius of curvature of the lens. The profile of the lens and the fitted curve are shown in Fig. 4.4(b). The fitted curve returned a value of 519±1 mm, while the quoted radius of the curvature of the lens was 515.5 mm. This shows that our method is able to quantitatively reconstruct the object function.

4.3.2 Phase measurement of the wing of a fly

The method was also applied to a three–beam interferogram of the wing of a common house fly. As this sample contained higher spatial–frequency information than the lens, the sampling of the phase, and consequently, the number of vortices needed for reconstruction had to be increased. This was achieved by using the $\lambda/2$ and $\lambda/4$ wave plates described in §4.2 to perform phase shifting in each reference arm of the interferometer. Three–beam interferograms were recorded for the fly's wing, with 11 phase shifts of $\pi/4$ between images used to reconstruct the phase. Each interferogram was smoothed by convolution with a Gaussian filter of approximately 60 μm FWHM and the transverse wave–vectors and vortex points were found using the same method as for the spherical lens. Vortices were not separated by sign for the fly's wing as Eq. (4.8) is applicable to both vortices and anti–vortices. For each of the 11 phase–stepped interferograms, the phase was recovered using Eq (4.8). Before the phase of each image could be combined for interpolation, the relative offset between the phase of each image needed to be determined. Two–dimensional cross–correlation of a small portion of the vortex lattice was used to determine the offset between each phase step. The coordinates of the maximum in each cross–correlation were substituted into Eq. (4.8), which gave the value of the phase difference between each vortex lattice. This phase difference was added to each reconstruction and linear interpolation applied to every vortex point in all 11 phase stepped lattices. This resulted in a single phase image which incorporated information from all 11 recorded vortex lattices. The recovered wrapped and unwrapped phase reconstruction of the fly's wing is shown in Figs. 4.5(a) and (b), respectively. For comparison, the phase determined by the Takeda method (Takeda et al.,

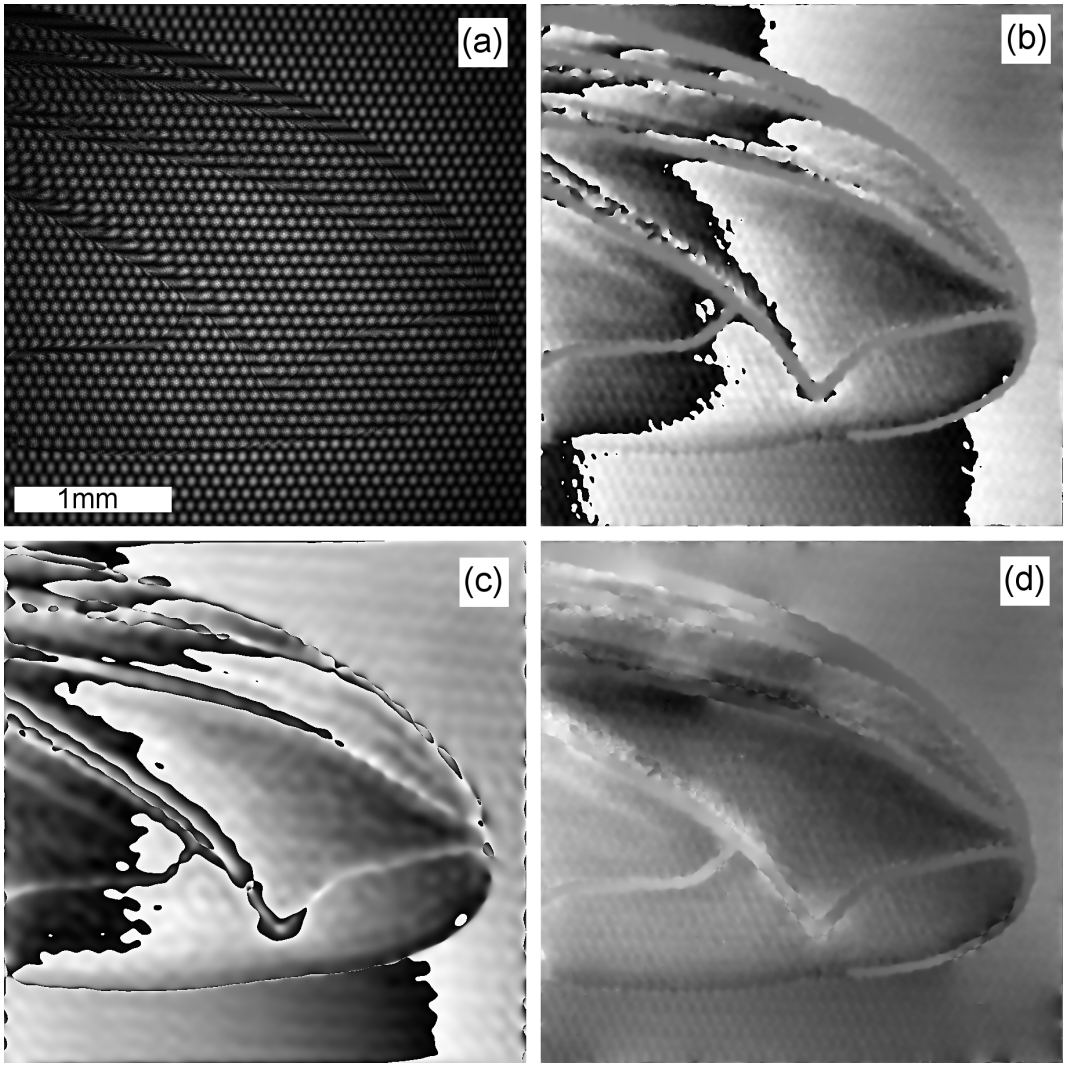

Figure 4.5: Experimental results for the wing of a fly. (a) Experimental three-beam interferogram of the fly's wing. (b) Recovered phase of the fly wing using Eq. (4.8). (c) Recovered phase using the Takeda method (Takeda et al., 1982). (d) The unwrapped phase in (b). The greyscale in (b) and (c) is wrapped to the range $[-\pi/2, \pi/2]$ from black to white.

1982). This method works by firstly Fourier transforming an image of a fringe pattern, then isolating one side–band from the rest of the spectrum using a mask. The side–band is then shifted to the centre of the image and then the inverse Fourier transform is taken, resulting in the complex function describing any perturbation to the fringe pattern. The phase of the function gives the phase of an object placed in the arm of the interferometer. Using a mask of 25×25 pixels, we applied the Takeda method to a two–beam interference pattern of the wing of a fly. With the measured phase shown in Fig. 4.5(c). The two–beam interference

pattern required for input into the Takeda method was obtained by blocking one of the reference arms of the three–beam interferometer. This was also carried out on a subsection of the spectrum of the three–beam interference pattern, producing a result consistent with the phase obtained for the two–beam case. This is expected as a three–beam interference pattern implicitly contains all the information in the two–beam interference pattern. The phase recovered from the vortex interferometer was unwrapped using the method of Volkov and Zhu (2003) and is shown in Fig. 4.5(d).

Since Eq. (4.8) has been used to recover the phase of the fly wing, it is wrapped from $-\pi/2$ to $\pi/2$. The phase recovered from the Takeda method has also been wrapped between these values, which facilitates comparison of the two results. Both images in Figs. 4.5 (a) and (b) show good quantitative agreement with each other. Minor differences are observed around the frame of the wing, where the absorption of the light is strongest. However, the phase measured by the vortex lattice technique is mostly unaffected by the absence of information in these regions. The local nature of our technique allows the phase reconstruction to simply ignore these strongly absorbing regions, which are then interpolated. For Fourier–based methods these regions result in larger phase gradients. This is observed at the bottom edge of the wing where the phase rapidly increases, which is not observed in Fig. 4.5(b). Ringing artifacts in Fig. 4.5(c) are also observed, corresponding to spatial frequencies not present in the fly's wing. Our vortex lattice method is not susceptible to this type of artifact as the technique uses a local algebraic solution to recover the object's phase.

4.4 Comments on Vortex Interferometry

A surprising aspect of the technique described in this chapter is that vortices, which correspond to points at which the phase becomes undefined, are used to determine the phase shifts of an object. In many forms of phase retrieval vortices are often undesirable, since their presence can be detrimental to accurate phase retrieval (Allen et al., 2001a).

The separation of vortex sub-lattices was realised using Eq. (4.5), which depends on a particular geometry for the three phasors; it also applies to a fixed topological sign of the vortex sub–lattice. The ability to separate vortices based on their sign is a consequence of the vortex sign rule (Freund and Shvartsman, 1994). A vortex occurs at the intersection of lines where the real and imaginary parts of the wavefunction vanish; the sign rule states that

each time a line intersects with another, the sign of the topological charge alternates. If we assign a particular change to an arbitrary vortex, the sign rule allows us to determine the sign of every vortex present in the lattice. Hence we may separate the vortices in the lattice based on their topological charge. To reconstruct the phase without a sign ambiguity, using Eq. (4.5), the sign of k_x and k_y for each reference wave and the topological charge of each vortex must be determined. This is because the sign of the vortex determines the form of the equation used to calculate the phase, and the sign of k_x and k_y determines the value of the coefficients of $\tilde{x}$ and $\tilde{y}$. These steps can be bypassed by using *a priori* knowledge of the object, e.g., an object with refractive index greater than unity results in a positive sign in the phase shift, as a consequence of the retardation of the wave as it passes through the medium. If it is necessary to determine the sign of the refractive index as well as the sign of k_x and k_y for each reference arm and the topological charge of the vortex sub–lattice, they can be experimentally measured using the method outlined in Frączek et al. (2005, 2006).

The spatial resolution of the technique is determined by the spacing of the vortices in the lattice, which increases with the angle that each wave–vector makes with the z–axis; this allows the spatial resolution to be tuned. However, increasing the vortex density too much results in aliasing, which leads to inaccurate vortex localization. Changing the relative angles between the transverse wave–vectors can also be used to increase sampling, as this affects the geometry of the lattice, e.g., a hexagonal lattice geometry occurs when the angle between each wave–vector is 120°, whilst a rectangular geometry occurs when two wave–vectors are orthogonal. This could also be used as an alternative to phase stepping, obviating the need for wave plates. In this case the relative angles between wave–vectors can be changed for each image. The changing lattice would result in a shift of the vortices and therefore sample more of the object.

An advantage of using vortex interferometry for phase retrieval is its robustness in the presence of noise. Because the method uses a real space calculation, the power spectrum of the noise has little or no effect on the reconstruction. In this case, noise only affects the vortex localization itself, since small variations due to noise may either lead to false detection of a vortex or inaccurate localization of the vortices. These effects are only significant if the vortex lattice spacing is small and the intensity “blobs” in Fig. 4.3 are under–sampled.

4.5 Extension of Technique to Super–resolution Imaging

In an attempt to perform super–resolution imaging an additional experiment was carried out using a slightly modified three–beam interferometer. This investigation was motivated by the superoscillatory nature of vortices, as discussed in §2.3.1. We conjectured that by increasing the angles that the three wave–vectors make with the optical axis, so that the spacing between the vortices was less than one wavelength, we could perform super–resolution measurements. Furthermore the presence of the two sub–lattices of vortices and anti–vortices means that they are not subject to restrictions on the vortex density (Roux, 2003). However, it became evident that the technique, as described in this chaper is incapable of creating super–resolution images. This is due to the fact that the object wave passes through the specimen prior to superposition with the beams from the two reference arms; hence the vortices never interact directly with the object. Any subwavelength feature would correspond to an evanescent object wave (Felsen, 1976), exponentially decaying within a few wave–lengths (Wolf and Foley, 1998). Thus any subwavelength features of the object will not carried by the wave to the point where it interferes and so the vortices do not "see" these features.

The experimental setup was modified in order to have the optical vortex lattice illuminate the object. In this arrangement the position of each vortex in the lattice is unaffected by the phase object.[1] Originally the optical vortex lattice is perturbed, so that only one wave incurs a phase shift due to the object. With the object under structured illumination by the vortex lattice, each constituent plane wave incurs an identical phase shift from the object. In this case each phasor rotates by the same angle (see Fig 4.1), hence there is no change in the relative phase angles and the condition for vortex localization is also unchanged.

Instead, we chose to localize a single vortex within the lattice and scan it across a subwavelength feature (Masajada, 2005; Masajada et al., 2011a,b; Augustyniak et al., 2012; Szatkowski et al., 2014). Using high precision piezoelectric motors to adjust the reflecting mirrors of the reference wave's mirrors, the optical vortex lattice could be stepped in any transverse direction to high precision. Upon scanning a vortex across a subwavelength feature a transverse shift in the position of the optical vortex was observed, which is consistent with vortex dynamics (Rozas et al., 1997; Brunet et al., 2010; Hemo et al., 2011). Unexpectedly,

[1] Shifts in the positions of the vortices can still be introduced by propagating the vortex lattice. However, measuring the phase shifts from such vortex displacements would require an entirely new approach, which was not investigated.

when scanning the vortex lattice we also observed vortex/anti–vortex nucleation events. These nucleation events occurred often, but consistent and repeatable observation of the phenomenon could not be achieved. At most we were only able to show that the shift in the vortex implied that a subwavelength feature was present. Due to the non–repeatability of these observations, we could not quantify this phenomenon in any meaningful way, forcing us to abandon trying to obtain super–resolution measurement with vortex interferometry.

4.6 Conclusion

In this chapter we presented a new method of phase determination, called vortex lattice interferometry, where three plane waves are interfered to generate a uniform vortex lattice. Distortions in this lattice, due to phase shifts induced in the object wave, are related to the position of each vortex in the interference pattern (see Eqs. (4.5) and (4.8)). Localizing each vortex in the lattice allows for the phase to be algebraically calculated. The technique was demonstrated on a spherical lens and a fly's wing. Both phase reconstructions show good agreement with existing Fourier methods. An advantage of our technique is its robustness to noise and its capacity to reconstruct the phase locally. With recent experiments demonstrating vortex lattices in electron beams (Verbeeck et al., 2012; Petersen et al., 2013b; Dwyer et al., 2015), the implementation of this technique for matter waves is a real possibility. An attempt was also made to extend vortex interferometry to perform super–resolution imaging. Whilst interesting results were observed, the difficulty in quantifying these observations led to this approach to super–resolution being ultimately abandoned.

5 Phase Retrieval Based on Caustic Surface Measurements

In Chap. 4 we presented a method of phase determination using vortices. In the current chapter, another type of phase retrieval method is discussed, which utilizes caustics as the measurement tool. If the optical vortex interferometer was an example of singularimetry using vortices—namely the singularities of waves—then this chapter presents a technique that is an example of singularimetry using caustics, the singularities of rays. Here, a general method of recovering phase is presented for a wave field that has been naturally focused to form a caustic surface corresponding to a fold catastrophe. By definition, a caustic represents singularities in the mapping from the state space to the control space, as described by the caustic condition given by Eq. (3.36). No direct inverse of this mapping exists, which takes points of the initial phase and then maps to points on the caustic surface. The novelty of the type of singularimetry presented in this chapter is that it provides a method of recasting this ill–posed inverse problem to one that is well–posed, by assuming a small amount of *a priori* information about the wave field, which is derived from the formalism of catastrophe theory.

5.1 Introduction

Consider the dancing caustic network of bright lines on the bottom of a swimming pool. This intricate web of caustics is a result of the light rays refracting from the undulating water surface. The corresponding inverse problem can be stated as follows: given a snapshot of the caustic network at some time t_0, can we determine the topographic map of the water's surface? In optics, inverse problems are typically solved by seeking the inverse function, which described how light propagates from the source to the image. Such methods require that the inverse function exists. However, the presence of caustics in an image requires that

the caustic condition (3.36) at each point of the caustic surface is satisfied. It follows from Eq. (3.36) that the Jacobian determinant of the mapping vanishes and no inverse exists at these points. The inverse problem of determining a wavefront from a caustic is ill–posed as there is no direct solution, hence the answer to the original question would appear to be no.

Reconstructing the wavefront from the caustics it forms has several advantages. Over large length scales, caustics typically represent the brightest regions of a light field. Given a noisy signal from a dim source, propagating through a medium with a characteristic length scale much larger than the wavelength, the highest signal–to–noise ratio would typically occur along points on a caustic surface. Determining the wavefront deformation solely from the information provided by a caustic would allow measurements of the phase shifts in atmospheric sciences and astronomy. Phase retrieval based on caustics would also facilitate characterisation and measurement of lens aberrations in an optical system. The aberrations may be described by an aberration function, which is expressed as Zernike polynomials (Zernike, 1942). All germs of the elementary catastrophes are found to be represented by a particular Zernike polynomial. This means that a caustic forming in the focal volume of an aberrated lens depends on the specific type of aberrations present. Chromatic aberration leads to fold catastrophes (Berry and Klein, 1996); spherical aberrations give cusps (Nye, 2005); astigmatism produces the elliptic umbilic (Berry et al., 1979), and coma gives rise to the hyperbolic umbilic (Marston and Trinh, 1984; Nye, 1984). This correspondence between a specific catastrophe and an aberration has been utilized to generate electron vortices within diffraction catastrophes by deliberately inducing aberrations in an electron microscope (Petersen et al., 2013b; Clark et al., 2013). Electron diffraction catastrophes produced by aberrations will be discussed in Chap. 6.

In this chapter we find a solution to the apparent ill–posed inverse problem of determining the phase of a caustic–forming wavefront upon recasting the problem into a well–posed one, by using the *a priori* observation of a generic fold catastrophe. The method exploits a Taylor series expansion of the phase, which is related to measurements of the versal unfolding of the caustic in the control space. Our treatment is limited to the fold catastrophe; however, methods of extending the technique to higher order catastrophes are briefly discussed.

In semi–classical molecular collision theory, similar ideas in which a catastrophe's germ have been determined from information measured from the bifurcation set have been investi-

gated. The asymptotic form of the scattering amplitude has been shown to adopt the form of a catastrophe integral, similar to Eq. (3.29) (Connor, 1976; Horn et al., 1989). Knowledge of the control parameters are used to calculate the elements of the scattering matrix. The type of elementary catastrophe associated with the scattering amplitude depends on the nature of the scattering. To calculate the differential cross–section of He^+–Ne scattering, Connor and Farrelly (1981b) showed that the potential function of the scattering integral could be represented by the canonical form of the cusp catastrophe, expressing the semi–classical scattering amplitude in terms of a Pearcey integral and its first derivatives (Pearcey, 1946; Connor and Farrelly, 1981a). The authors showed that the two unfolding parameters can be determined either by an iterative algorithm or a direct algebraic approach. The algebraic method was also shown to be capable of determining the universal unfolding parameters for the umbilic catastrophes (Uzer and Child, 1982).

The only example of the inverse problem in the literature in the context of diffraction catastrophes is due to Trinkaus and Drepper (1977). These authors used an approximate solution to Eq. (3.29) for the fold diffraction catastrophe, showing that the distance between the first and second interference maximum depends on the third order partial derivative ϕ_{yyy}. Hence ϕ_{yyy} could be determined by measuring the distance between the two intensity maxima. Given a measurement of intensity, the second derivative of the phase ϕ_{xx}, can also be determined. A similar analysis was applied to the cusp diffraction catastrophe. Trinkaus and Drepper (1977) showed that a measurement of the absolute intensity was sufficient to find $\phi_{xx}, \phi_{xxy}, \phi_{xyy}$ and ϕ_{yyy}. If the cusp belongs to a higher order catastrophe, such as one of the umbilics, the problem can be parameterized in such a manner that it is possible to determine the four partial derivatives of the phase.

5.2 Description of the Problem

We assume that an initial wavefront at $z = 0$ results in a fold diffraction catastrophe after propagation, so that the intensity pattern is described by Eq. (3.29). Given that we can only measure the intensity of the fold, we seek to determine the wavefront of the initial wave. The diffraction detail of the caustic is considered to be unnecessary detail in this situation, hence we choose to simplify the problem by only considering the stationary phase approximation to Eq. (3.29). The problem reduces to finding the function $g(x, y)$ in Eq. (3.30), given the

gradient mapping described by Eqs. (3.34) and (3.35). By definition, no direct inverse of this mapping exists as the caustic condition, $\phi_{xx}\phi_{yy} - \phi_{xy}^2 = 0$, implies the Jacobian of the mapping is singular (see Eq. (3.36)). However, given our discussion on equivalency in Chap. 3, g is equivalent to the canonical form of the fold catastrophe, which we denote by f_c. Following §3.1 we also know that local diffeomorphisms exist between the respective bifurcation sets of g and f_c. Using the concept of determinacy and the connection between g and the fold catastrophe germ, we show that g can be retrieved solely from imaging measurements of the bifurcation set.

5.3 Determinacy

To explain the concept of determinacy we must first introduce the notion of jets. The jet of f, denoted by jf, is the Taylor series representation of f at the origin with a vanishing zeroth order term.[1] The k–jet of f, denoted by $j^k f$, is the Taylor series of f truncated up to and including all terms of order k. Hence we may express any real–valued smooth function as

$$f(x_1, \ldots, x_n) = j^k f(x_1, \ldots, x_n) + \hat{f}(x_1, \ldots, x_n), \tag{5.1}$$

where $\hat{f}$ is a function whose partial derivatives of order $< k$ vanish at the origin. Jets play a pivotal role in the derivation of the germs listed in Table. 3.1 and in the application of catastrophe theory. Since catastrophe theory only considers the local nature of a critical point, the stability of a function can be analysed in terms of a local approximation in the form of a truncated Taylor series. Under certain criteria the k–jet of a function can be shown to be an exact representation of the function, not simply an approximation.

A function is said to be k–determined if, whenever $j^k f = j^k g$, it follows that g is equivalent to f. The k–jet $j^k f$ is then said to be k–sufficient (Stewart, 1982). It follows that if $f \sim g$, there exists a smooth local change of coordinates $y : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$f(x_1, \ldots, x_n) = g(y_1(x_1, \ldots, x_n), \ldots, y_n(x_1, \ldots, x_n)). \tag{5.2}$$

If we set

$$g(x_1, \ldots, x_n) = j^k f(x_1, \ldots, x_n), \tag{5.3}$$

[1] We can make this definition of a jet more general by considering a Taylor series expansion about an arbitrary point. However, since we are only considering analytic functions, any point may be transformed to the origin under a diffeomorphism, described by Eq. (3.3). Thus the jet about any point is equivalent to the jet at the origin.

then the condition for k-determinacy is clearly satisfied for Eq. (5.3). By making the coordinate transformation (5.2), Eq. (5.3) becomes

$$f(x_1, \ldots, x_n) = j^k f(y_1(x_1, \ldots, x_n), \ldots, y_n(x_1, \ldots, x_n)). \tag{5.4}$$

Therefore if f is k–determined, it may be expressed exactly by its k–jet in the y coordinate system. Note that if f is k–determined, the function f must also be $(k+1)$–determined. A truncated Taylor series generally represents an approximation of a function given some local neighbourhood. Even if an arbitrarily large number of terms is included in the expansion, it still remains a approximation. The significance of Eq. (5.4) is that a truncated Taylor series represents the behaviour of a function exactly, provided the function is k–sufficient.

5.3.1 Rules for ascertaining determinacy

Here we present a set of rules that may be applied to rigorously test the determinacy of a function (Poston and Stewart, 1976). Given a function $f : \mathbb{R}^n \to \mathbb{R}$, we assume $f(0, \ldots, 0) = 0$. The order to which f is determined at 0 is denoted by $\sigma(f)$ and can be found as follows

(i) Calculate the gradient ∇f at $\mathbf{x} = 0$. If the gradient is non–zero then $\sigma(f) = 1$. By Theorem 3, f is equivalent to the non–critical canonical form.

(ii) If $\nabla f = 0$, compute the Hessian matrix $\mathbf{H}$ of f at 0. If $\det(\mathbf{H}) \neq 0$ then $\sigma(f) = 2$.

(iii) If $\det(\mathbf{H}) = 0$, then we expand f in terms of $j^k f$, following the prescription in Eq. (5.1). We define the first order partial derivatives of each term in the $j^k f$ as

$$f_i(\mathbf{x}) = \frac{\partial j^k f}{\partial x_i}(\mathbf{x}) \qquad (1 < i < n), \tag{5.5}$$

where $\mathbf{x} \equiv (x_1, \ldots, x_n)$. Hence f_i is a polynomial of order $\leq k-1$.

(iv) We now list all polynomials of the form

$$^{k+1}\overline{m(\mathbf{x}) f_i(\mathbf{x})}, \tag{5.6}$$

where m is any monomial of degree ≥ 2; the bar denotes that the product is truncated to order $k+1$. We label these polynomials by P_i.

(v) If all possible monomials of order $k+1$ in $\mathbf{x}$ can be expressed as linear combinations of P_i, then f is k–determined.

If it is found that f is not k–determined then this algorithm is iterated to test for $(k+1)$ determinacy. The way a function behaves under perturbations is described by its derivatives. These rules test whether any of the derivatives of f is a factor of any monomial that has a higher order than k. If any of the derivatives of f are not a factor of all monomials of higher order than k, then f has an unfolding that is not determined to order k; hence f isn't k–determined. This also shows how determinacy and unfolding are connected. This point is illustrated later in this section, but first we give an example to make clear the application of the above rules.

Consider the polynomial of the germ of the fold catastrophe in two–dimensions, i.e., $p = \frac{1}{3}x^3 + \frac{1}{2}y^2$, whose derivatives are $f_x = x^2$ and $f_y = y$. By rules (i) and (ii) p is not 1– or 2–determined, thus we test for 3–determinacy. To this end, we then list all the monomials P_i (see Table 5.1).

$O(m)$	${}^4\overline{m(x,y)f_x}$			${}^4\overline{m(x,y)f_y}$			
1	x^3	x^2y		xy	y^2		
2	x^4	x^3y	x^2y^2	x^2y	xy^2	y^3	
3				x^3y	x^2y^2	xy^3	y^4

Table 5.1: List of all P_i monomials corresponding to the fold catastrophe germ $\frac{1}{3}x^3 + \frac{1}{2}y^2$.

All monomials in (x, y) of degree 4, i.e., x^4, x^3y, x^2y^2, xy^3 and y^4, are contained in the set of P_i polynomials in Table 5.1. We conclude that the fold catastrophe germ is 3–determined.[2]

For polynomials in two dimensions, a method known as Siersma's trick (Siersma, 1974) provides a simpler method to test for determinacy. Let $f = xy^2$ with $f_x = y^2$ and $f_y = 2xy$, and suppose we again wish to find monomials that have f_x and f_y as a factor. Consider the arrangement of monomials in Fig. 5.1(a). The "shadows" cast by the monomials, xy and y^2 contain all monomials of which xy and y^2 are factors, respectively. If all monomials of the same order are shaded by the monomials contained in the first order derivative of a

[2]This test does not always give the lowest order determinacy of a polynomial. The definition of determinacy can be further classified into strong and weak determinacy. These measures of determinacy are often the same, but can differ when considering higher order polynomials (see e.g., Poston and Stewart (1996)).

polynomial, then the polynomial has determinacy of that order. For xy^2, the entire left side of Fig. 5.1(a) containing monomials of the form x^n remains unshaded. Therefore xy^2 is not determined to any finite order and is said to be indeterminate. Now consider the germ of the elliptic umbilic catastrophe, $f = \frac{1}{3}x^3 - xy^2$ with $f_x = x^2 - y^2$ and $f_y = -2xy$. Because f_x is a linear combination of terms we cannot draw its shadow in the table of monomials. We instead seek monomials that may be written as linear combination of the first derivatives of f:

$$x^3 = xf_x + \frac{1}{2}yf_y, \tag{5.7}$$

$$y^3 = \frac{1}{2}xf_y - yf_x. \tag{5.8}$$

Thus, for the polynomial $x^3 - xy^2$, we draw the shadows of x^3, y^3 and xy monomials, shown in Fig. 5.1(b). The third row of monomials is fully shaded indicating that the polynomial $x^3 - xy^2$ is 3–determined. This example also highlights how Siersma's trick may be used to find the universal unfolding of a singularity. Note that the four monomials that are not shaded in Fig. 5.1(b) correspond to the universal unfoldings given by the canonical form of the elliptic umbilic catastrophe. This shows how the concept of determinacy is central to catastrophe theory and to Thom's theorem in particular.

5.4 3–jet of the Initial Wavefront

We now return to the task of retrieving the initial wavefront from a fold caustic surface. To proceed, we express g as a Taylor expansion about the point $x = y = 0$:

$$\begin{aligned} g(x,y) = g_x x + g_y y + \frac{1}{2!}\left(g_{xx}x^2 + 2g_{xy}xy + g_{yy}y^2\right) \\ + \frac{1}{3!}\left(g_{xxx}x^3 + 3g_{xxy}x^2y + 3g_{xyy}xy^2 + g_{yyy}y^3\right) + O(4), \end{aligned} \tag{5.9}$$

where subscripts denote partial differentiation and g is the phase of the initial wave as defined in Eq. (3.24). We neglect the constant term $g(0,0)$, since it only adds a uniform phase shift and therefore has no effect on the formation of the fold caustic. Given that g produces a fold caustic in the control space, it is therefore equivalent to the canonical form of the fold catastrophe. We have shown in §5.3.1 that the germ of the fold catastrophe is 3–determined, if $g \sim f_c$; it then follows that g is also 3–determined and from Eq. (5.5) g is described exactly by its 3–jet.

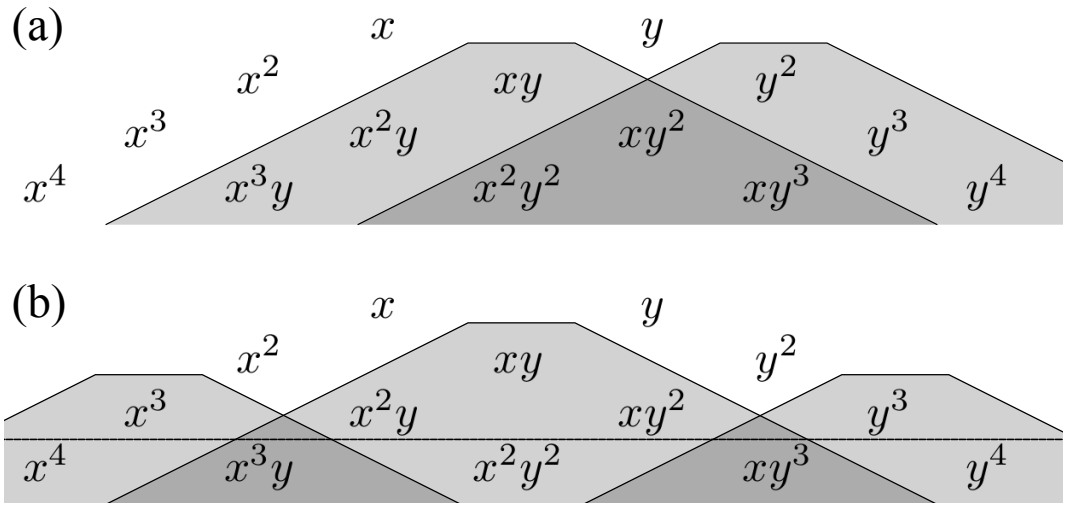

Figure 5.1: Table of monomials. The "shadow" cast by the lowest order monomial within the shadow is a factor of all other monomials contained within it. A polynomial is finitely determined when the "shadows" cast by its first order derivatives contain all monomials of the same order. (a) The" shadows" of the first order derivatives of xy^2. Since the left hand side of the table is left unshaded, xy^2 is therefore indeterminate and has infinite codimension. (b) The polynomial $x^3 \pm xy^2$. Since all monomials are shaded at 3rd order, the polynomial is 3–determined.

Although we have shown that we are justified in truncating Eq. (5.9) to third order, j^3g is still not equivalent to the canonical form of the fold. We need to also demand that the singularity in g must be of corank of 1, i.e., identical to that of the fold. This restriction on g amounts to neglecting all third order terms, except one, in j^3g. We may also choose to rotate our coordinate system, defined by the control space variables ξ, η and ζ as described in §3.6, such that the caustic at the focus in the control space is parallel with the η axis; this dictates that the third order term in j^3g is x^3. In §5.3.1 it was shown that the terms x^2y and xy^2 are indeterminate. If they were to be included in the expansion of g then there is the possibility g may not be 3–determined, given a particular set of values for its coefficients. Hence for an initial wavefront that forms a caustic, g must have the following form:

$$g(x, y) = Ax^3 + \delta\zeta_a x^2 + \delta\zeta_b y^2 + Bxy + \delta\xi x + \delta\eta y, \tag{5.10}$$

where $A, B, \delta\zeta_a, \delta\zeta_b, \delta\xi$ and $\delta\eta$ are real coefficients. The expansion coefficients have been re–labelled to better describe their effect on the caustic surface. By using the principles of catastrophe theory, we have shown that the problem of determining g has been reduced to determining a small finite number of coefficients.

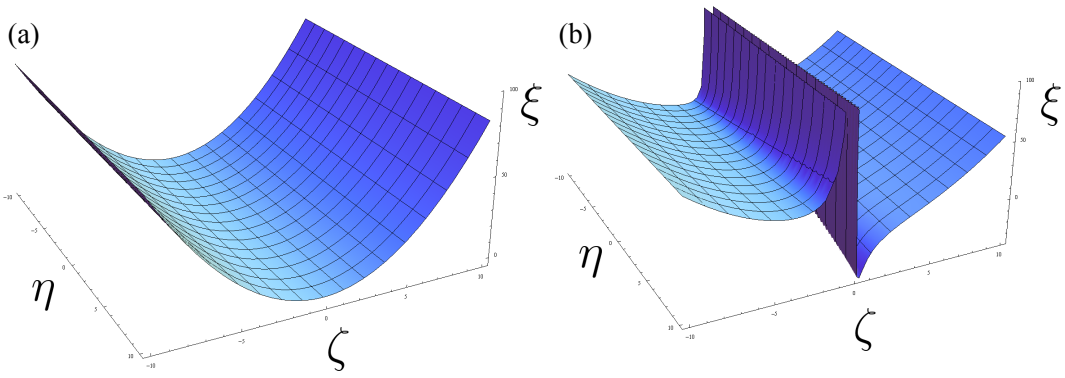

Figure 5.2: (a) The caustic surface formed by a family of rays from a wavefront whose phase corresponds to the canonical form of the fold catastrophe. (b) Bifurcation set of Eq. (5.13).

Before examining the caustic surface produced by the initial wavefront, we briefly describe the geometry of the caustic surface of the canonical fold. To this end we substitute the canonical form of the fold catastrophe into Eq. (3.30), giving

$$\phi(x, y; \xi, \eta, \zeta) = x^3 + y^2 + \zeta(x^2 + y^2) + \xi x + \eta y. \tag{5.11}$$

By using the caustic condition, the state variables may be eliminated from Eq. (5.11), which gives the equation for the caustic surface in the control space, defined using the variables (ξ, η, ζ), as

$$\zeta^2 - 3\xi = 0. \tag{5.12}$$

Thus the fold caustic unfolds parabolically through its focus and is parallel to the η–axis. The surface described by this equation is shown in Fig. 5.2(a). This parabolic unfolding may also be observed in the focal volume of an aberrated lens as shown in Fig. 1.5(a). The form of g in Eq. (5.10) represents a versal unfolding of the singularity associated with the fold catastrophe. To find the geometry of the caustic surface corresponding to g, Eq. (5.10) is substituted into Eq. (3.30), to give

$$\phi(x, y; \xi, \eta, \zeta) = Ax^3 + Bxy + (\zeta + \delta\zeta_a)x^2 + (\zeta + \delta\zeta_b)y^2 + (\xi + \delta\xi)x + (\eta + \delta\eta)y. \tag{5.13}$$

By applying the ray and caustic conditions to Eq. (5.13), we obtain three equations, which

are used to eliminate the state variables. This gives the bifurcation set of Eq. (5.13) as

$$\frac{1}{A(\zeta+\delta\zeta_b)}\Big\{B^4 - 8B^2(\zeta+\delta\zeta_a)(\zeta+\delta\zeta_b) + 24AB(\zeta+\delta\zeta_b)(\eta+\delta\eta) + 16(\zeta+\delta\zeta_b)^2\left[(\zeta+\delta\zeta_a)^2 - 3A(\xi+\delta\xi)\right]\Big\} = 0. \quad (5.14)$$

Equation (5.14) represents the most general surface of a fold catastrophe in the control space arising from a wavefront with initial phase g (see Fig. 5.2(b)). The bifurcation set of Eq. (5.12) varies in a similar manner to that of the canonical fold catastrophe at large values of ζ, but exhibits a singularity at $\zeta = -\delta\zeta_b$. The behaviour of the caustic near this point has also been noted by Peregrine and Smith (1979). Equation (5.14) shows that the caustic surface is well defined if $A \neq 0$. This is a consequence of the fact that if $A = 0$, then the x^3 term in g vanishes; g would then become 2–determined and not equivalent to the canonical form of the fold catastrophe. An explanation as to how each expansion coefficient affects the caustic surface is left to the next section, where our method of determining the expansion coefficients is outlined.

5.5 Determination of the Expansion Coefficients from the Surface of the Fold Caustic

The Bxy term in Eq. (5.13) is the origin of the singularity of the caustic surface shown in Fig. 5.2(b). If $B = 0$ in Eq. (5.14) the bifurcation set reduces to the square brackets, which is identical to the form of the bifurcation set of the canonical fold caustic, given by Eq. (5.12). Essentially, B dictates how rapidly the caustic surface diverges. The information pertaining to the coefficient B is encoded in the variation of the caustic close to focus. To proceed we write Eq. (5.14) as an explicit function of ξ and ζ:

$$\eta(\xi,\zeta) = \frac{-1}{24AB(\zeta+\delta\zeta_b)}\Big\{B^4 - 8B^2(\zeta+\delta\zeta_a)(\zeta+\delta\zeta_b) + 24AB\delta\eta(\zeta+\delta\zeta_b) + 16(\zeta+\delta\zeta_b)^2\Big[(\zeta+\delta\zeta_a)^2 - 3A(\xi+\delta\xi)\Big]\Big\}. \quad (5.15)$$

The coefficient B is therefore given by

$$\frac{\partial\eta}{\partial\xi} = \frac{2}{B}(\zeta+\delta\zeta_b), \quad (5.16)$$

whence we have

$$B = 2\left(\frac{\partial^2\eta}{\partial\xi\partial\zeta}\right)^{-1}. \quad (5.17)$$

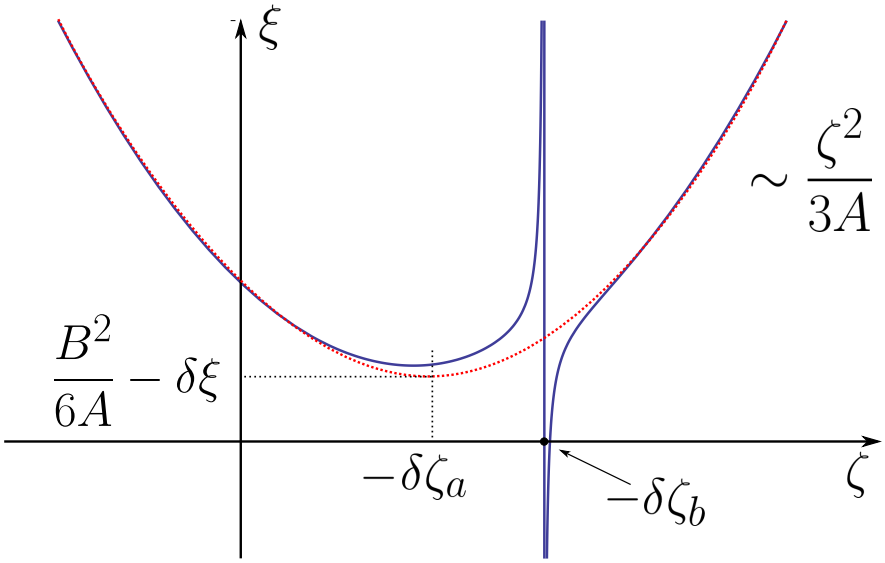

Figure 5.3: Curve of the bifurcation set defined in Eq. (5.14) at constant η. The red dotted curve represents the asymptotic approximation, described by Eq. (5.19). The dilation factor for the curve is $1/3A$. The position of the local minimum is at the coordinates $(-\delta\zeta_a, B^2/6A - \delta\xi)$, whilst the curve becomes undefined when $\zeta = -\delta\zeta_b$.

We see that the gradient of η in the ξ–direction varies linearly with ζ. Therefore B may be determined from Eq. (5.17) by measuring the variation in the gradient of the caustic in a through–focal series. However, a more practical method of measuring B consists of fitting a linear curve to Eq. (5.16) and calculating its gradient.

Consider the term $\delta\zeta_b y^2$ in Eq. (5.13). Due to the pre–factor in front of the braces in Eq. (5.14), the bifurcation set becomes undefined at $\zeta = -\delta\zeta_b$, as seen in Fig. 5.3. If the full three–dimensional caustic surface was imaged, the value of $\delta\zeta_b$ could be measured simply by determining at what focal value this occurs. However, since the rate at which the caustic surface diverges depends on B, a larger value would result in less precision in a measurement of $\delta\zeta_b$. Note that the ζ–axis intercept of Eq. (5.16) is $-\zeta_b$, therefore $\delta\zeta_b$ may be found from a linear fit to Eq. (5.16), in the same manner as B was determined, with the value of ζ at $\partial\eta/\partial\xi = 0$ giving $-\delta\zeta_b$.

In the Ax^3 term, the coefficient A serves as a dilation factor for the caustic surface. Although the xy term in g causes a large variation in the surface near $\delta\zeta_b$, at points far away from the singularity the behaviour of the surface is approximately parabolic; this is similar to the bifurcation set of the canonical fold caustic. Therefore to measure A, we seek an

asymptotic representation of Eq. (5.14) for $|\zeta| >> |\delta\zeta_a|, |\delta\zeta_b|$. To this end we write Eq. (5.14) as an explicit function of η and ζ, i.e.,

$$\xi(\eta,\zeta) = \frac{B^4}{48A(\zeta+\delta\zeta_b)^2} + \frac{B(\eta+\delta\eta)}{2(\zeta+\delta\zeta_b)} - \frac{B^2(\zeta+\delta\zeta_a)}{6A(\zeta+\delta\zeta_b)} - \frac{1}{3A}\left[3A\delta\xi - (\zeta+\delta\zeta_a)^2\right]. \quad (5.18)$$

For $|\zeta| >> |\delta\zeta_a|, |\delta\zeta_b|$, the first two terms of Eq. (5.18) are negligibly small relative to the terms in the square brackets. Equation (5.18) then reduces to

$$(\zeta+\delta\zeta_a)^2 - 3A(\xi+\delta\xi) + \frac{B^2}{2} \approx 0. \quad (5.19)$$

This shows that the generic caustic surface is approximated by the equation of the canonical fold catastrophe with an additional constant term. From Eq. (5.19), A can be found from:

$$\frac{\partial\xi}{\partial\zeta} = \frac{2}{3A}(\zeta+\delta\zeta_a), \quad (5.20)$$

whence

$$A = \frac{2}{3}\left(\frac{\partial^2\xi}{\partial\zeta^2}\right)^{-1}. \quad (5.21)$$

Thus by measuring the caustic at points far from its singularity, A can be retrieved using Eq. (5.21). However, a more practical method for determining A would consist of fitting a linear curve to the variation of the caustic at a constant ξ though focus.

Now consider the term $\delta\zeta_a x^2$. Equation (5.19) shows that the coefficients $\delta\zeta_a$ and $\delta\xi$ displace the caustic surface in the $-\zeta$ and $-\xi$ direction, respectively (see Fig. 5.3). Hence, the critical point of the bifurcation set occurs at $\zeta = -\delta\zeta_a$, as shown in Fig. 5.3. The coefficient $\delta\zeta_a$ can therefore be found from the coordinates of the extrema of Eq. (5.19). Thus $\delta\zeta_a$ is determined from the derivative $\partial\xi/\partial\zeta$ calculated from Eq. (5.19) by setting $\partial\xi/\partial\zeta = 0$ and solving for ζ. The coefficient $\delta\xi$ in the term $\delta\xi x$ may be measured determined once $\delta\zeta_a$ is known. The corresponding ξ coordinate for the critical point of Eq. (5.19) is $B^2/6A - \delta\xi$ (see Fig. 5.3). By using the known values of B, A and $\delta\zeta_a$ and any point on the caustic surface such that $|\zeta| >> |\delta\zeta_a|$ and $|\delta\zeta_b|$, Eq. (5.19) can be solved to determine the value of $\delta\xi$.

Finally, consider the term $\delta\eta y$. The coefficients $\{A, B, \delta\zeta_a, \delta\zeta_b, \delta\xi\}$ can be determined independently from $\delta\eta$. This latter coefficient can be found by choosing any point on the caustic surface and substituting the values of those coefficients already determined into Eq. (5.14). This procedure can be performed multiple times for a large number of points to give better precision in the measurement of $\delta\eta$.

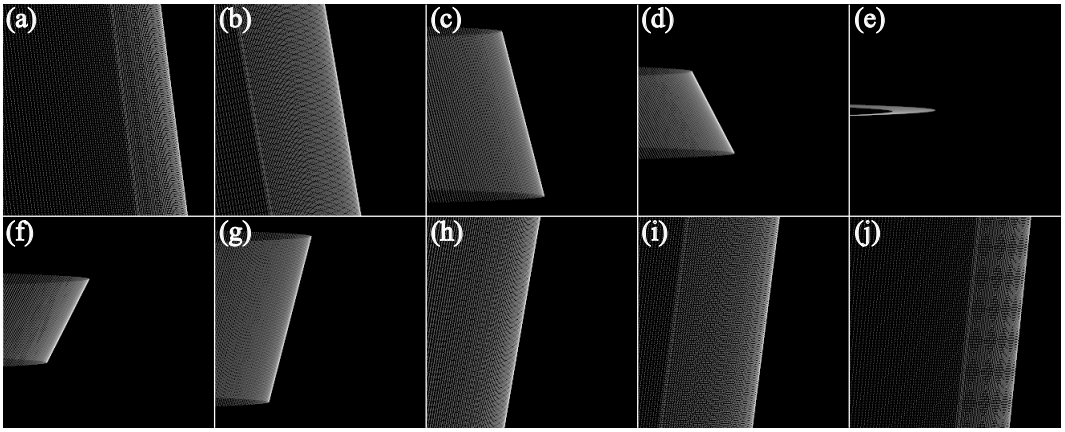

Figure 5.4: A through–focal series of a simulated fold caustic observed in the (ξ, η)–plane, with $\Delta\zeta = 1.75\ \mu\text{m}^{-1}$ between each consecutive image. (a)–(d) As ζ tends towards $-\delta\zeta_b$ the caustic rotates counter–clockwise due to a non–zero xy term in Eq. (5.13). (e) At $\zeta = \delta\zeta_b$ the fold caustic disappears from the field of view as it has diverged to infinity. (f)–(j) The fold caustic comes back into the field of view and rotates clockwise and shifts in the positive ξ–direction as ζ increases.

5.6 Simulation of a Fold Caustic Surface as it Passes Through Focus

The methods discussed in §5.5 were applied to simulated data for an arbitrary fold caustic. The initial wavefront was defined according to Eq. (5.10) with the values of the coefficients chosen randomly and with uniform probability over $[-2, 2]$. This range of values was chosen for purely computational reasons to prevent the unfoldings of the caustic surface from being too large. Otherwise the mapping of the germ from state space to control space will result in sparse data coverage of the caustic surface in control space. However, large variations in a physical continuous wave field would not obviate the utility of the methodology.

The caustic surface was then constructed using the gradient mapping defined by Eqs. (3.34) and (3.35). An example of the simulated data is shown in Fig. 5.4. The fold is viewed in the (ξ, η)–plane and its unfolding observed along the ζ–axis. The position of the fold is shifted along the ξ–axis before beginning to rotate counter–clockwise as ζ tends towards the point $-\delta\zeta_b$. Note that this figure is only a representative section of the actual caustic surface used to determined the coefficients. The caustic surface was isolated from the rest of the image using the Canny edge detection algorithm (Canny, 1986). The edge detection scale was set by convolving all images with a Gaussian of width matching the ap-

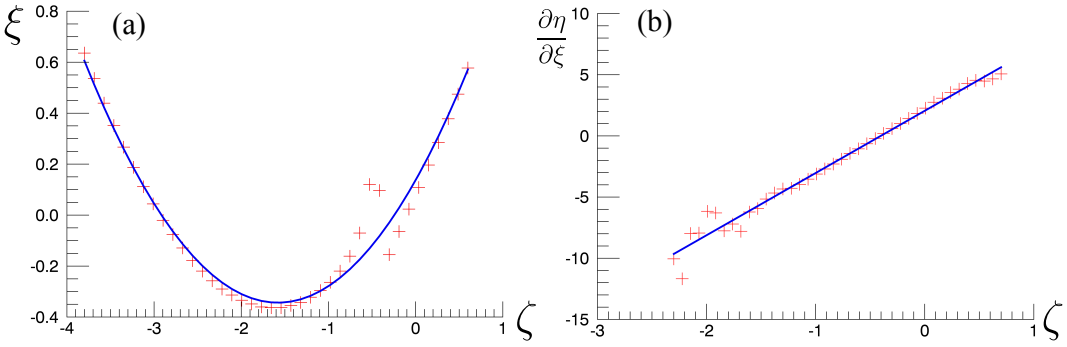

Figure 5.5: (a) The red crosses give the measured ξ coordinate of the caustic surface as a function of ζ for a fixed η. The blue curve represents a fit to the data based on Eq. (5.19). This curve was used to determine A, $\delta\zeta_a$ and $\delta\xi$. (b) A linear fit to $\partial\eta/\partial\zeta$ as a function of ζ. The fitted curve corresponds to Eq. (5.16) and was used to calculate the coefficients B and $\delta\zeta_b$.

	A	B	$\delta\zeta_a$	$\delta\zeta_b$	$\delta\xi$	$\delta\eta$
Actual value	1.615	0.558	1.584	0.412	0.360	−0.451
Measured value	1.730	0.393	1.579	0.403	0.375	−0.475
% Error	1.121	29.570	0.316	2.184	4.167	5.322

Table 5.2: Measured coefficients for a fold catastrophe based on simulated data. The actual values used in the simulation were chosen randomly.

proximate width of the fold singularity (Lindeberg, 1998). This had the additional advantage of smoothing the pixelation present in the images. The gradients in each direction, G_ξ and G_η, were subsequently calculated by convolution with the Sobel gradient operators (Sobel, 1990). The edge strength S, and angle Θ, at each pixel in the image was found using

$$S = \sqrt{G_\xi^2 + G_\eta^2} \tag{5.22}$$

$$\Theta = \arctan(G_\eta/G_\xi). \tag{5.23}$$

Since caustics represent the brightest part of the image, the caustic surface corresponds to points where the edge strength is a maximum.

Table 5.2 shows the measured coefficients determining using our methodology. These results show good agreement between the simulated and measured values, with B exhibiting the largest discrepancy. The coefficients A, $\delta\zeta_a$ and $\delta\xi$ were found by fitting a function based

on Eq. (5.19) to the ξ–position of the caustic at a constant value of η. The value of η was chosen based on where the measured edge strength was the largest. The curve fitted to the simulated data is shown in Fig. 5.5(a). The fit matches the data well with the exception of the points near $-\delta\zeta_b$; this is due to the caustic surface diverging. A parabolic fit was used, instead of fitting a linear curve to Eq. (5.20), as it avoided the need to calculate the partial derivative $\partial\xi/\partial\zeta$. Figure 5.5(b) shows a linear fit to $\partial\eta/\partial\xi$. This derivative was found from the angle the caustic makes with the ξ–axis, which was calculated using Eq. (5.23). Both B and $\delta\zeta_b$ were determined from this graph since the equation of the fitted curve corresponds to Eq. (5.16). To determine $\delta\eta$, the point on the caustic surface that gave the largest edge strength for each image at each ζ value was found. These coordinates and the values of all other coefficients were substituted into Eq. (5.14) and solved to give a value for $\delta\eta$ for each image. The final value of $\delta\eta$ was taken as an average of the values determined for each image.

5.7 Discussion

The caustic phase retrieval methods demonstrated in this chapter can also be applied to the undulating fold diffraction catastrophe. Although the technique was demonstrated for a caustic surface based on simulated data using ray theory (see Eqs. (3.34) and (3.35)), the variation of the intensity maximum of the diffraction catastrophe is identical. The only difference in applying this method to a diffraction catastrophe is a practical one. The catastrophe can no longer be isolated using an edge detection algorithm, as the caustic surface will be blurred by interference. Instead, a ridge detection algorithm is required to measure the position and variation of the intensity maximum (Lindeberg, 1998). Another consideration is that according to Eqs. (3.34) and (3.35) the location of the edge of the caustic does not coincide with the position of the intensity maximum of a diffraction catastrophe. At the focus the canonical fold caustic intersects the ξ–axis at 0; however, when interference effects are manifest, this intensity maximum shifts slightly in the direction of $-\xi$, depending on the wavelength. This wavelength shift would need to be taken into account in determining coefficients. An experimental implementation of phase retrieval for diffraction catastrophes has not been carried out and is subject to ongoing work.

Extending our caustic phase retrieval technique to higher order catastrophes requires

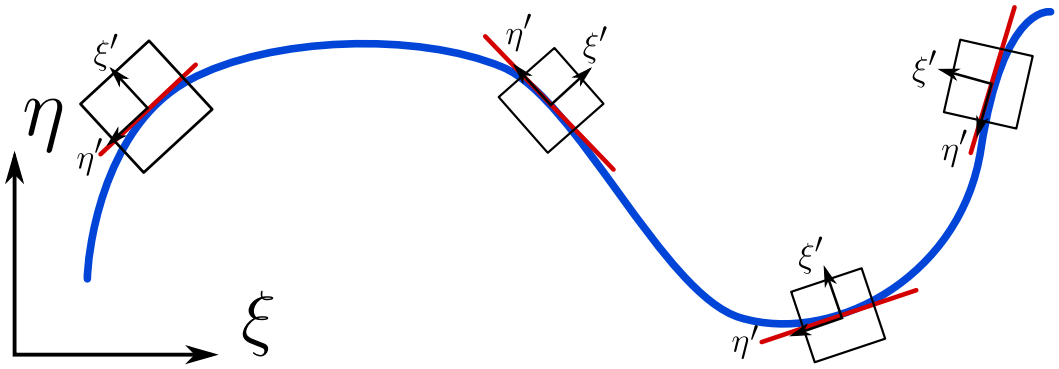

Figure 5.6: Schematic of the "patching" method to measure the phase of a wave field that gives rise to a curved caustic. At each smooth point on the caustic the local neighbourhood is approximately linear and parameterised by (ξ', η'). The initial phase g, of the wave field can be retrieved within each local neighbourhood by applying the methods described in §5.5 to the variation of the tangents (shown in red) with respect to ζ. Each local measure, or "patch" of g, could then be combined to determine g globally in terms of the coordinates (ξ, η).

the expansion of g to include all third–order terms; the critical point in g would then have a corank of 2 and therefore correspond to an unfolding of an umbilic catastrophe. Truncation of g as a higher order polynomial, such as a 4–jet, would then be equivalent to the cusp catastrophe. In our treatment of the inverse problem we assumed that the only third order term in the expansion of g involved x^3. This means that the caustic in the observation plane is linear and the parabolic variation of the caustic along ζ occurs in the ξ–direction. To extend the scope of this chapter to include caustics with curvature in the (ξ, η)–plane, we would need to consider all third order terms in j^3g. However, the inclusion of these additional terms means that g would no longer be equivalent to the canonical form of the fold catastrophe. Instead g would be diffeomorphically equivalent to one of the umbilic catastrophes, since these are 3–determined and of corank 2. The specific subsection of the catastrophe that j^3g represents (indicated by the squares in Fig. 5.6) would be dictated by the relative magnitudes of each expansion coefficient. Suppose the coefficient of the x^2y term was the largest, then the curved caustic observed in the field of view would consist of a subsection of the bifurcation set of any of the umbilic catastrophes. The specific catastrophe to which the caustic belongs would not be apparent from observing the caustic surface alone. If the full bifurcation set of an umbilic catastrophe was observed, then the third–order terms in the expansion could be determined immediately, since we know that the third–order terms of j^3g would have to match those in the germ of the umbilic catastrophes. It should still be possible to employ a

method similar to that presented in this chapter, in order to retrieve the phase from higher order catastrophes. However, the geometry of these catastrophes is more elaborate compared to the fold catastrophe. Because of this, such an algorithm would require more sophisticated image processing in order to fully quantify any versal unfolding of g.

When all third order terms are included in $j^3 g$ we propose an alternative means of extending our technique to catastrophes that are non–linear in the (ξ, η)–plane. Consider an image of a single fold caustic that possesses some curvature.[3] At every smothh (ξ, η) point along the caustic there exists a local neighbourhood within which the caustic is approximately linear (see Fig. 5.6). Within this neighbourhood we may define a new set of local coordinates (ξ', η'), such that the tangent to the caustic is parallel with the η'–axis. Using this parameterization, the local neighbourhood g is approximated in the (ξ', η') coordinate system by Eq. (5.10). The phase of the initial wavefront can therefore be determined using the method outlined in §5.5, so that g is measured locally along all points of the caustic using knowledge of the size, position and orientation of the local neighbourhood defined by (ξ', η'), relative to the global coordinates (ξ, η). Each local measure of g can then be "patched" together to recover the global form of g for the entire image (see Fig. 5.6). However, the way in which multiple local measurements are combined to give g is non–trivial; consequently further work is required to determine the algorithms needed to realise this "patching" method.

A major benefit in determining the exit phase of a wave field from its caustic surface is that it enables phase retrieval even in the presence of significant noise. This is because caustics will typically represent the point of the wave field with the greatest intensity. An example is the dim light from a twinkling star, which is also aberrated by propagation through the Earth's atmosphere (Mercier, 1962; Jakeman et al., 1976).

In a statistical analysis of caustics formed by light originating from a star, Berry (1977) considered the mth moment of a random wave field $\langle |\psi|^{2m} \rangle$. Moments of order $m \geq 2$ become non–Gaussian due to caustic formation. For a codimension of 2, it was shown that only cuspoidal catastrophes contributed to the intensity, with the fold catastrophe dominating the second and third moments. In three dimensions, higher order catastrophes can also arise in the phase of a random Gaussian wave field. In this situation the fold catastrophe

[3] It is slightly misleading to call this caustic a fold catastrophe as any curved caustic is actually some small subsection of the caustic surface of a higher order catastrophe. The observation of a single caustic in the field of view is a consequence of the terms x^3 or y^3 in the expansion of the wavefront having relatively large coefficients.

was again shown to dominate in the second moment, whilst the fold, elliptic umbilic and hyperbolic umbilic catastrophes dominate the third moment. These results suggest that for caustics arising from propagation through inhomogeneous media, the fold catastrophe is the most likely. So although our method of retrieving the phase from a caustic is only currently applicable to the fold catastrophe, this limitation may not be a issue for the majority of physical scenarios, as the fold catastrophe has the greatest probability of forming.

5.8 Conclusion

In this chapter we described a new form of singularimetry, which allows the phase of a wave field to be determined from measurements of the caustic surface of a fold catastrophe in the focal volume. The solution to this well–posed inverse problem that was recast from the ill–posed inverse problem is made possible through Thom's classification theorem (see §3.5), which essentially provides us with *a priori* information; namely, if a fold caustic is observed in the control space, then the phase of the wave must be equivalent to the germ of the canonical form of the fold catastrophe. This reduces the problem of retrieving the phase of a wave field to one of determining a finite set of coefficients in the Taylor expansion of the phase. We have outlined a method for determining each coefficient in the expansion from the variation and the position of the corresponding caustic surface in the control space. The technique was demonstrated on simulated data, using the ray formalism to simulate the caustic surface of the fold catastrophe. This formalism was chosen to simplify the image processing steps; however, this singularimetry method is applicable when applied to diffraction catastrophes. All six of the expansion coefficients describing the phase were retrieved to within an error of $< 6\%$. The coefficient of the xy term was the only exception, with an error of approximately 30%. This is because the algorithm required the derivative $\partial\eta/\partial\xi$ to be calculated numerically in order to determine B. Although this technique in its current form is applicable to isolated linear fold catastrophes, it is an important first step in retrieving the phase of the wavefunction for higher order catastrophes, or for more complex caustic networks observed in nature.

6 Phase Singularities Associated with Caustics and Diffraction Catastrophes

In this chapter we will consider different types of singularities and phase anomalies in electron waves. The first half of the chapter describes work published in Petersen et al. (2013b).[1] This work exploited the relationship between vortices and diffraction catastrophes to experimentally produce vortices in the electron wavefunction. This was motivated by our early attempts to produce electron vortices as discussed in §2.1.3.1, which also lead to the work presented in Chap. 4. Diffraction catastrophes are created by aberrating the lenses within a Transmission Electron microscope (TEM). The analysis of the resulting diffraction catastrophes indicated that intensity zeros occurred near the edges of the caustic surfaces. Phase retrieval was then used to explicitly show that electron vortices were present in the wave field.

The second part of this chapter deals with the Gouy phase anomaly—an abrupt phase shift that occurs when a wave passes through a focus (Gouy, 1891). Caustics, representing the natural focus of a wave are therefore inextricably linked with the Gouy phase anomaly. This part of the chapter is based on two published papers on the Gouy phase anomaly by Petersen et al. (2013a) and Petersen et al. (2014).[1] In the first paper, the Gouy phase shift was measured for electron waves in the TEM. An astigmatic lens was used to focus an electron wave into two line foci. In–line holography was then used to show that the phase of the electron wave underwent a phase shift of $\pi/2$ through each line focus. The second of these papers unites various interpretations of the Gouy phase anomaly within a single theoretical framework.

[1] Relevant papers are included in Appendix C. See the declaration on page vi for my percentage contribution to these publications.

6.1 Electron Vortex Production and Control using Aberration Induced Diffraction Catastrophes

Quantized vortices in propagating electron waves are of interest for electron wavefunction phase mapping, since in–line holography approaches can fail in the presence of these wavefront dislocations (Allen et al., 2001a). In this context, electron vortices were theoretically shown to arise from the transmission of fast electrons through an atomic lattice (Allen et al., 2001b). However, the simulations in §2.1.3.1 of carbon phase objects with such thickness variations suggest that specimen geometry is important and that electron vortices are not readily produced. The coherent superposition of distorted plane waves is expected to give rise to vortices. By analogy with three-wave interference in light optics (Masajada and Dubik, 2001), we expect to experimentally observe electron vortices for Bragg diffraction from crystalline specimens. Indeed, in biprism interference patterns from crystalline specimens, one can find examples of forked dislocations (Ravikumar et al., 1997), which are signatures of quantized phase vortices. Similarly, the hallmarks of three–wave electron interference (Nicholls and Nye, 1987) are evident for three–beam electron diffraction from crystals (Vincent et al., 1993; Terauchi et al., 1994; Moodie et al., 1996; Nakashima et al., 2007).

Throughout this thesis it has been discussed how diffraction catastrophes give rise to caustics, which are generic to optical wave fields and stable with respect to perturbations. Berry et al. (1979) demonstrated the elliptic umbilic catastrophe using a triangular lens formed from water. Further, it was shown that the associated diffraction detail can be accurately approximated using a superposition of plane waves, which create a lattice of phase vortices (see §2.2). The analysis of umbilic and cusp catastrophes by Berry et al. (1979) demonstrates that vortical wave fields can be formed by lenses with aberrations.

Optical caustics formed by primary aberrations have been widely recognized since the pioneering work of Nijboer and Nienhuis (Nijboer, 1947; Nienhuis and Nijboer, 1949; Wolf, 1951). Caustics arising from lens aberrations in the TEM have also been characterized theoretically using geometric optics (Hawkes and Kasper, 1996). Primary aberrations can induce diffraction catastrophes; in particular, the umbilic foci for astigmatism and coma have been shown to be hyperbolic (Berry and Upstill, 1980). Caustics are routinely observed when

TEM illumination apertures are removed to include electron trajectories deflected through large angles. These rays do not produce significant interference effects in the probe intensity, since the removal of probe–forming apertures degrades the spatial coherence, which would seem to exclude the possibility of observing phase dislocations arising from diffraction catastrophes. However, partial coherence can be improved if small illumination apertures are used. Diffraction catastrophes can then be explored by imposing significant probe forming aberrations to disturb the point focus and create severely distorted coherent electron probes.

We have conducted experiments to produce and manipulate electron vortices in the specimen plane of a TEM, employing lens aberrations to create diffraction catastrophes. We used a Titan^3 80–300 TEM (FEI) that provides dual aberration correction (CEOS GmbH) of both the illumination and imaging lenses. Operating at an acceleration voltage of 300 kV, imaging lens aberrations were corrected to third order and the microscope was set up in the bright field imaging mode to minimize the semiangle subtended by the field emission gun source. A 10 μm condenser aperture was selected; the resulting intensity distribution was a small yet parallel probe. By adjusting the condenser lenses, the illumination was focused in the specimen plane to produce a far–field diffraction pattern of the circular condenser aperture in the form of Airy rings. The circular symmetry of these rings was then broken by increasing twofold condenser astigmatism, producing a sharp (sub–nanometre) line focus streaking along one direction. Through–focus propagation of the resulting distorted electron probe, using the imaging lens, revealed the presence of four umbilic foci, which outlined the caustic of the probe. The electron probe cross section and decorating diffraction detail maintained form as the probe was imaged throughout the focal series. This observed persistence of form and stability under perturbations is a key aspect of a diffraction catastrophe.

Figure 6.1(a) shows the logarithm of an astigmatic experimental electron probe imaged near one of the line foci. Figure 6.1(b) shows the same probe imaged several hundred nanometres further along the optic axis. Since the electron optical configuration was chosen to optimize the spatial coherence, the diffraction pattern was quite dim. Accordingly, the acquisition times were 100 seconds per probe pattern. Comparing Figs. 6.1(a) and 6.1(b), the stability of the umbilic features with respect to smooth variation of the electron wave (changing focus) implies the formation of a diffraction catastrophe. Electron trajectories make very small angles with the optic axis and we can assume paraxial imaging conditions. If we

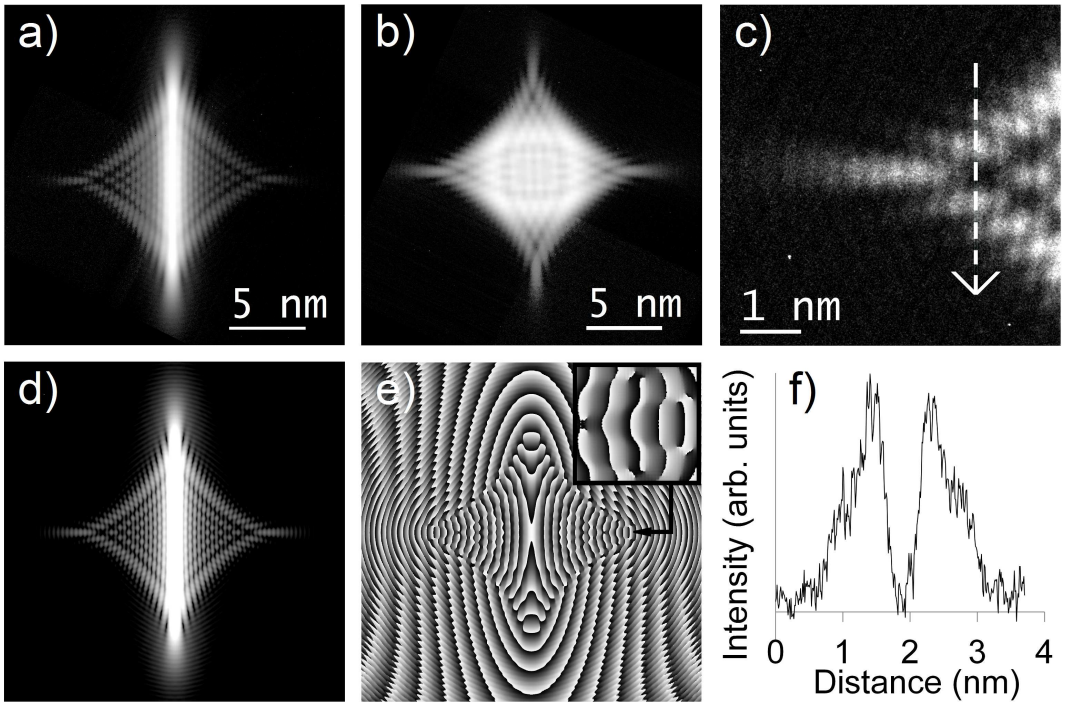

Figure 6.1: (a) Experimental intensity (logarithm) near a line focus caused by condenser astigmatism. (b) Experimental intensity (logarithm) at a defocus half way between two line foci. (c) Enlarged experimental intensity for the leftmost umbilic focus in (a). (d) Logarithm of the caustic intensity computed from Eq. (6.1). (e) Corresponding phase computed from Eq. (6.1), exhibiting an array of phase vortices. (f) Five–pixel wide line plot measured from (c) showing interior intensity zeros threading vortex cores.

further ignore chromatic aberrations or the spread of energies in the electron source, we can utilize the diffraction theory of aberrations based upon the Huygens-Fresnel principle (Born and Wolf, 1999). To this end, we consider the condenser lens aberrations as modifying the optical path lengths of spherical waves originating from a circular aperture $A(x, y)$. Fresnel propagation from the aperture plane describes the electron wave at a distance z along the optic axis, downstream from the aperture:

$$\Psi(x, y, z) = S(x, y, R_2)\mathcal{F}^{-1}\left\{\mathcal{F}\left[A(x, y)S(x, y, R_1)e^{(2\pi i/\lambda)[C(x^2+3y^2)+B(x^2y+y^3)]}\right]e^{-i\pi\lambda z(k_x^2+k_y^2)}\right\}, \tag{6.1}$$

where k_x and k_y are the Fourier coordinates conjugate to the Cartesian aperture–plane coordinates x and y, $\Psi(x, y, z)$ is the scalar wavefunction and λ is the wavelength. The symbols $\mathcal{F}$ and $\mathcal{F}^{-1}$ in Eq. (6.1) denote forward and inverse Fourier transforms, respectively. Primary astigmatism in the aperture plane is parameterized by the coefficient C using the polynomial form given by Kingslake (1925). The coefficient B parameterizes the degree of coma aberra-

tion, which is set to zero here; this is also relevant to another experiment we outline in this chapter, which used coma to induce diffraction catastrophes. The function $S(x, y, R_1)$ models the electron source as a simple spherical wave in the paraxial approximation with radius R_1, which numerically serves to condense the wave within an image array of fixed size (Paganin, 2006). The spherical wave $S(x, y, R_2)$ removes residual scaling of $\Psi(x, y, z)$ after propagation. Setting $B = 0$ and after careful adjustment of the free parameters C and R_1 in Eq. (6.1), we obtain excellent agreement between the experimental results in Fig. 6.1(a) and the theoretical predictions in Fig. 6.1(d). Diffraction catastrophes may also be simulated without the need to account for the functions $S(x, y, R_1)$ and $S(x, y, R_2)$ by substituting the corresponding catastrophe germ as the initial phase into Eq. (3.23) and using only one Fourier transform to calculate the diffraction integral (Paganin, 2006).

The intensity logarithm and phase calculated from the diffraction integral in Eq. (6.1) are shown in Figs. 6.1(d) and 6.1(e), respectively. Between the umbilic foci in Fig. 6.1(d), spontaneously nucleated phase vortices decorate the outer edges of the caustic and some are within the interior, where intensity zeros occur in Figs. 6.1(a), 6.1(c), and 6.1(d). Several minima within the umbilic foci in Fig. 6.1(a) contain intensity zeros, which is consistent with the creation of electron vortices. However, none of the minima inside the caustic in Fig. 6.1(b) correspond to intensity zeros, in agreement with diffraction integral calculations. Figure 6.1(f) shows a five pixel wide line plot of the intensity shown in Fig. 6.1(c). This line plot confirms the presence of intensity zeros near the line focus, inside the caustic. The observation of intensity zeros is a necessary condition for the existence of vortices, as discussed in §1.1.

Another experiment was performed to investigate the caustic associated with coma aberration. Astigmatism was minimized and the probe corrector was adjusted to induce primary coma. This aberration produces the hyperbolic umbilic catastrophe. Astigmatism could not be completely eliminated as strong excitation of the corrector lenses can lead to parasitic aberrations. To explore the coma induced diffraction catastrophe, the coma caustic was further enlarged by increasing the size of the condenser aperture to 150 μm, which diminished the contrast of the pattern. To compensate for the reduced coherence, the first condenser lens was excited to the maximum nominal setting after which the fringe contrast increased significantly. Additional coma caustics were then recorded using multiple frames

(ten) and shorter exposures (3 seconds per frame) to offset residual beam drift.

Figure 6.2 compares experiment with both catastrophe theory for the hyperbolic umbilic and the diffraction integral for the coma aberration. The experimentally measured intensity in Fig. 6.2(c) is an average of 50 images, each exposed for 1 s and then post–aligned with respect to each other. The horizontal line in Fig. 6.2(c) marks a single-pixel wide line plot, which is displayed in Fig. 6.2(g), showing significant fringe visibility and an array of intensity zeros. For Fig. 6.2(d), 10 images were acquired for 3 s, post–aligned and averaged. Arrays of intensity zeros were also observed at this defocus setting, which was nominally 100 nm from that shown in Fig. 6.2(c). The logarithm of the intensity in Fig. 6.2(d) compares well to that of the hyperbolic umbilic catastrophe in Fig. 6.2(a) and the diffraction integral intensity logarithm in Fig. 6.2(e). The hyperbolic umbilic catastrophe is described by

$$\Psi(x,y,z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(x^3+y^3-zs_1s_2-xs_1-ys_2)} ds_1 ds_2, \tag{6.2}$$

where s_1 and s_2 are the relevant state variables and the wavefunction Ψ is computed at focal depth z and plotted in the plane spanned by (x, y) (Berry and Upstill, 1980). Equation (6.2) was evaluated numerically with the coordinates x, y normalized by the number of pixels, to span ± 30 dimensionless units over 512×512 pixels. The integration variables s_1, s_2 were truncated to ± 3.5 and incremented in steps of 0.014 dimensionless units. The wavefunction Ψ was then cropped to 256×256 pixels to approximately match the field of view of the experimental data. The phase of Ψ, determined from Eq. (6.2), is shown in Fig. 6.2(b), where arrays of phase vortices decorate the interior and outer portions of the coma caustic. Similar vortices are evident in Fig. 6.2(f), which was computed from Eq. (6.1) with $C = 0$. The parameter B was varied until the caustic intensity pattern visually matched the experimental data.

Electron phase maps were experimentally determined from a through-focus series of images using the Gerchberg–Saxton–Misell phase retrieval algorithm (Gerchberg and Saxton, 1972; Misell, 1973). Figure 6.3(a) shows the logarithm of the retrieved intensity for the first image in the through-focus series for the astigmatism caustic. Comparing Fig. 6.3(a) and 6.3(b), it is evident that most of the diffraction detail was captured by the retrieval algorithm, although the intensity minima are not as distinct. The retrieved phase map in Fig. 6.3(b) has a similar form to that in Fig. 6.3(e). Propagation of the reconstructed wave over many

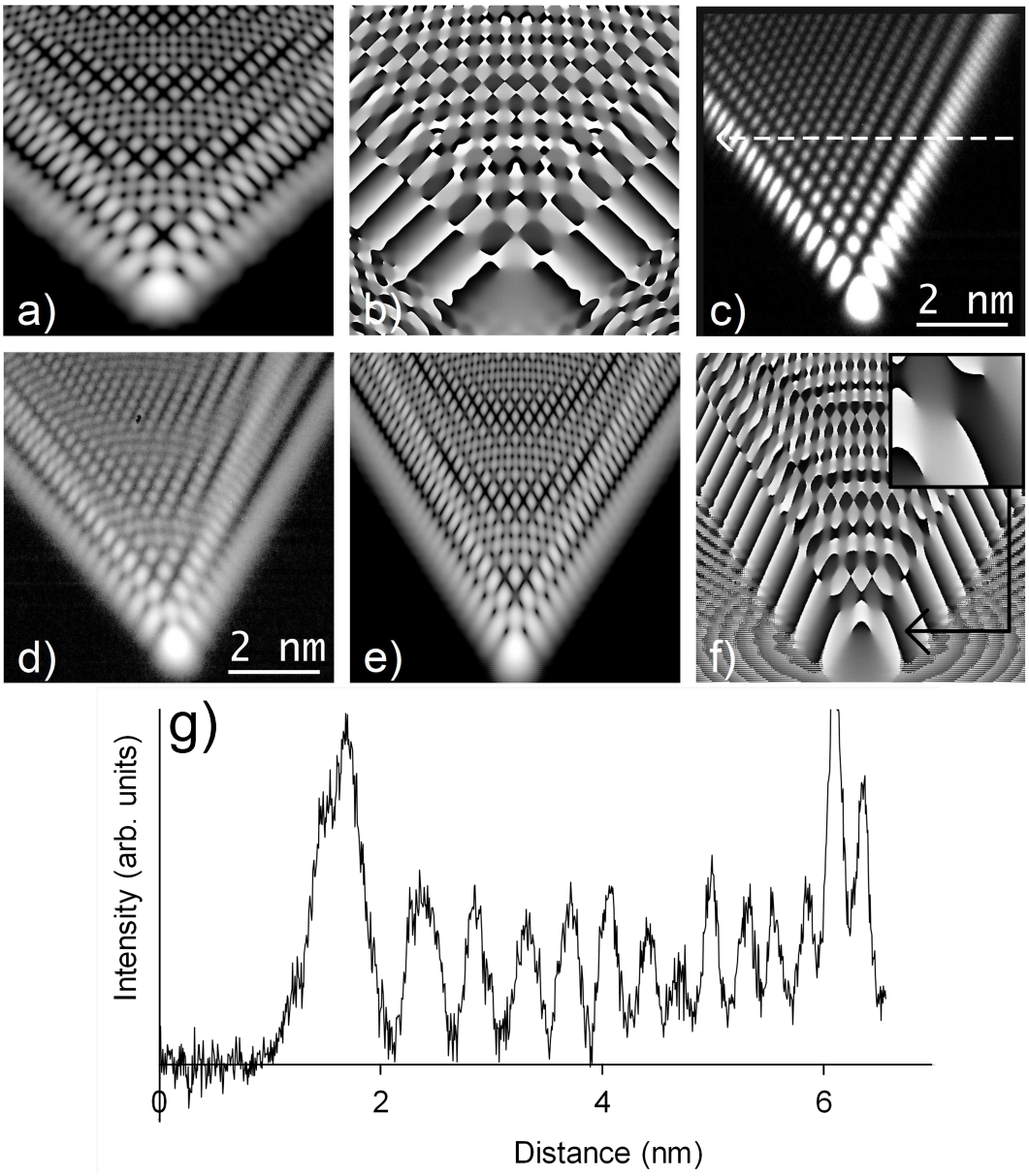

Figure 6.2: (a) Logarithm of the intensity calculated using diffraction catastrophe theory. (b) Phase of the diffraction catastrophe showing a distorted lattice of vortices and anti–vortices. (c) Experimental intensity of coma caustic near the diffraction focus. (d) Logarithm of the experimental intensity nominally 100 nm away from the diffraction focus. (e) Logarithm of the coma caustic intensity computed from the diffraction integral in Eq. (6.1) with $C = 0$. (f) Phase of the coma caustic from the diffraction integral in Eq. (6.1) with $C = 0$. (g) Line profile of (c), as indicated by the dashed line.

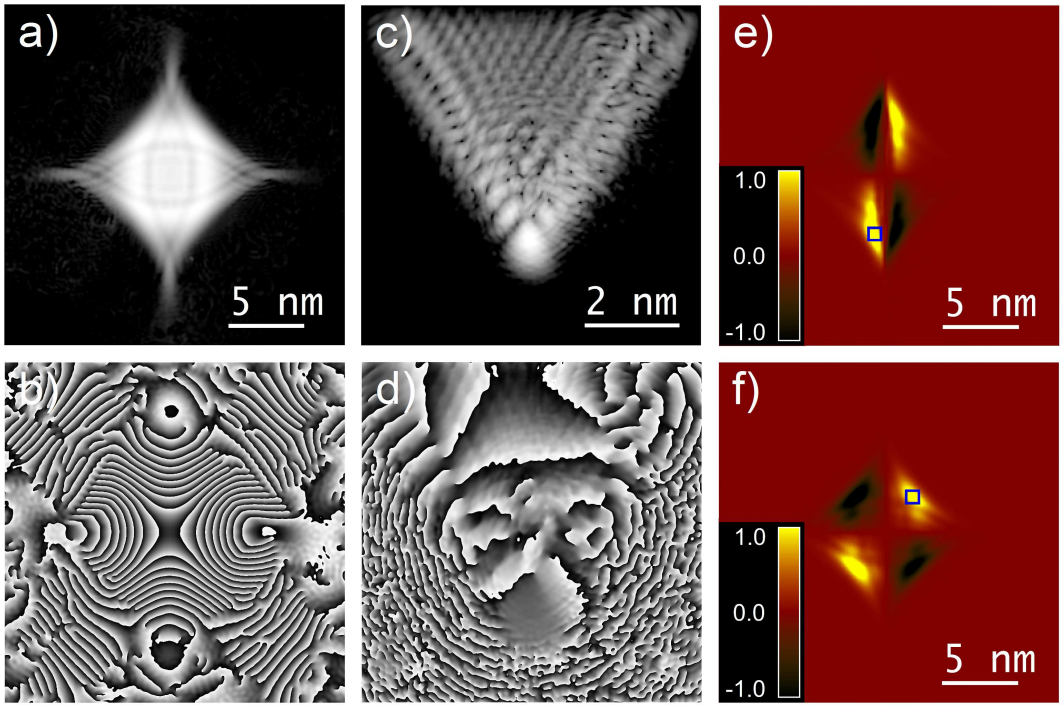

Figure 6.3: (a) Logarithm of the reconstructed intensity for the first image in the through focus series of astigmatic images. (b) Corresponding retrieved phase with vortices surrounding the caustic. (c) Logarithm of the reconstructed intensity for the first image in the through focus series of five coma images. (d) Corresponding retrieved phase. (e) The z–component of the orbital angular momentum density for the retrieved astigmatism caustic wavefunction near one line focus, and (f) a defocus between the two line foci.

focal planes correctly predicted a pair of mutually orthogonal sharp line foci, as well as preservation of the caustic shape and diffraction detail with varying focus, all of which were observed in our experiments.

Figure 6.3(c) shows the logarithm of the retrieved intensity for the first image in a through-focus series for the coma caustic. Vortices are seen to decorate the exterior of the coma caustic, with several inside the caustic shown in Fig. 6.3(d). Again, the intensity minima are not as distinct when compared to the experimental input. Nonetheless, the reconstructed wave reproduced much of the detail observed in the experimental images over the entire focal range. Figures 6.3(e) and 6.3(f) show the experimentally measured z–component of the orbital angular momentum (OAM) density, determined for the retrieved wavefunction about a central cross section of the astigmatism caustic, displayed in SI units of 1.0×10^{16} kg s^{-1} per electron. The root mean square value of the OAM density in Fig. 6.3(f) is $0.15\hbar$ nm^{-2}. The OAM density in Figs. 6.3(e) and 6.3(f) varies significantly across the

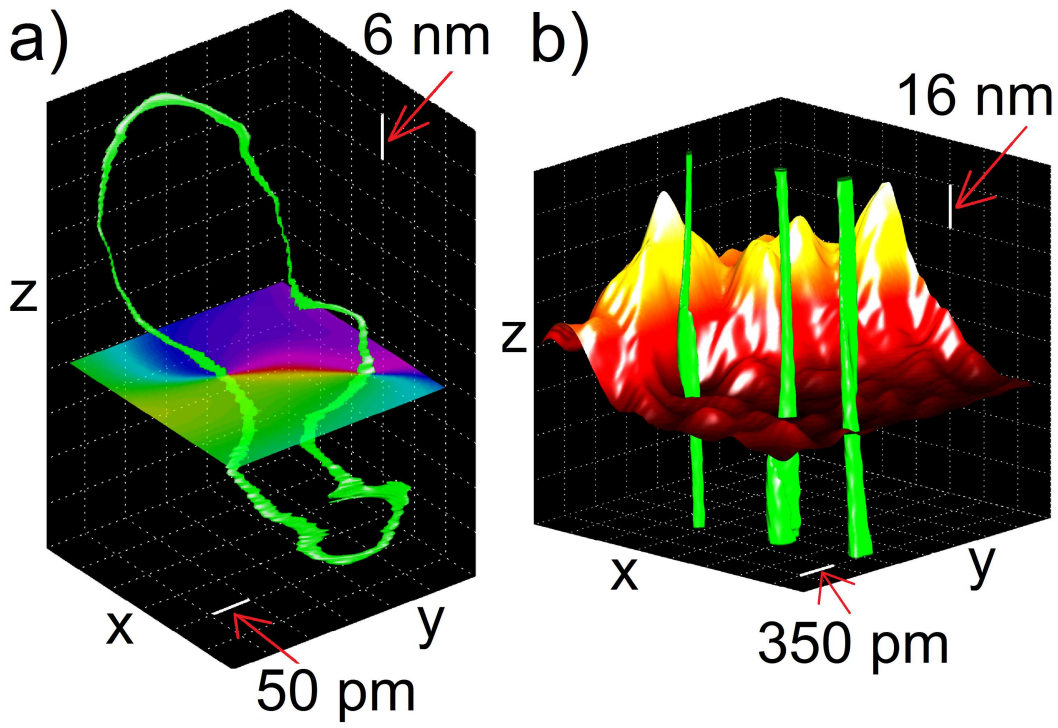

Figure 6.4: (a) Electron vortex loop observed in the interior of the coma caustic in Fig. 6.3, where the colour—coded phase map shows a vortex/anti–vortex pair, which nucleates and annihilates at specific points along the optic z axis. (b) Vortex lines puncturing the image plane for the astigmatism caustic of Fig. 6.3(b); the colour–code (excluding the vortex lines) represents the intensity.

beam, which is analogous to the optical interferometry measurements of Courtial et al. (1997) for elliptical Gaussian beams. The small boxes shown in Figs. 6.3(e) and 6.3(f) have side lengths of 0.7 nm, within which the integrated OAM per electron is $42\hbar$ and $35\hbar$, respectively. Within both boxes, the corresponding average probability densities are greater than 75% of the maximum intensity over the entire field of view.

Using the reconstructed experimental wavefunctions, vortices were tracked throughout a propagated focal series comprising 2048^3 points and detected by measuring points of nonzero circulation in two–dimensional phase maps. For our retrieved wavefunctions, we examined particular nodal lines in detail by cropping out all tracks except those in the sub–regions marked in the phase maps of Figs. 6.3(b) and 6.3(d). The electron vortex loop from the coma caustic in Fig. 6.4(a) shows nodal line excitations and Crow–like instabilities (Crow, 1970), apparently initiating dissociation into several vortex loops. The nodal loops within the hyperbolic umbilic catastrophe are similar to the nodal line topology

for four plane–wave interference, as discussed in §2.2.1, as the vortices within the coma caustic are a result of a similar four–wave interference pattern (Nye, 2006). Similar nodal loops within an elliptic umbilic catastrophe were studied by O'Holleran et al. (2009b). The coloured plane in Fig. 6.4(a) shows the phase windings around adjacent sides of the loop, which highlight vortices of opposite topological charge. Figure 6.4(b) shows three electron vortices represented as intensity iso–surfaces, which are adjacent to the umbilic focus of the astigmatism caustic shown in Fig. 6.3(b).

Our work shows that we are able to induce electron diffraction catastrophes, thereby creating distorted lattices of spontaneously nucleated electron vortices, using an aberration corrected electron microscope. We have demonstrated that phenomena, such as Crow-type nodal-line instabilities and nonlinear effects, such as pair creation or annihilation, can be measured for matter waves that obey a linear wave equation. Our singular electron optics observations closely parallel experimental and theoretical findings in light optics and hence raise the possibility of creating topologically knotted electron waves (Dennis et al., 2010). Using experimentally retrieved electron wavefunctions, we have also mapped the OAM density per electron for astigmatic wave fields. In addition to vortices, diffraction catastrophes have also been known to exhibit the Gouy phase anomaly (Pang et al., 2011; Rolland et al., 2012). In the next section we discuss this phenomenon in the context of electron diffraction catastrophes.

6.2 Measurement of the Gouy Phase Anomaly for Electron Waves

The Gouy phase anomaly (Gouy, 1891), which describes the additional phase shift accumulated by a wave packet upon focusing, has been of fundamental interest in light optics for more than a century, and the diverse literature on this phenomenon continues to grow. Understanding, measuring, and ultimately exploiting the Gouy phase in a variety of experimental contexts has been pivitol to the development of particular optical technologies. Gouy's original observations were made using mirrors and white-light interferometry (Gouy, 1891; Siegman, 1986). Visible–light lasers have since been used to measure the effect (Kandpal et al., 2007). Measurements of the Gouy phase for cylindrically focused waves were

reported in visible-light optics (Rolland et al., 2012), where the anomaly was generalized for astigmatic wave fields (Visser and Wolf, 2010). Specifically, Visser and Wolf (2010) derived the following expression for the on–axis Gouy phase anomaly $\delta(u)$ for a scalar wave diffracted from an aperture of size a, focused by an astigmatic lens of focal length f with coefficient of astigmatism A_0:

$$\delta(u) = \arg\left[\int_0^1 e^{i[(kA_0-u)\rho^2/2]}\rho J_0(kA_0\rho^2/2)d\rho\right] - \frac{\pi}{2}, \tag{6.3}$$

where J_0 is a zeroth–order Bessel function of the first kind and the integration variable ρ is dimensionless. The parameter u is proportional to the distance z along the optic axis according to $u = 2\pi(a/f)^2 z/\lambda$, where λ is the wavelength and k is the wave number. Equation (6.3) describes two sequential Gouy phase shifts of $\pi/2$ rad along the optic axis, associated with a pair of mutually orthogonal line foci, separated by $\Delta u = 2kA_0$ (see Fig. 6.5).

Recently, experiments were proposed for measuring the Gouy phase in matter waves, such as coherent atomic beams, using cylindrical focusing of Rydberg atoms (da Paz et al., 2011). Inspired by this proposal, we have measured the Gouy phase for astigmatic electron matter waves using phase retrieval (Gerchberg and Saxton, 1972; Misell, 1973). Aberration correction lenses in a TEM were used to induce astigmatic pairs of line foci with transverse cross sections narrower than 1 nm, separated by more than 1 μm along the longitudinal optic axis. A through–focal series of images was used to retrieve the phase of the wavefunction, with the Gouy anomaly through each line focus measured by propagating the retrieved electron wavefunction.

6.2.1 Experimental creation of two line caustics used for measuring the Gouy phase

Figure 6.5 shows a schematic of the experiment in which electron matter waves were diffracted by a circular aperture and focused by an astigmatic lens; this produced a caustic volume containing a pair of mutually orthogonal line foci at different points along the optic axis in the vicinity of the backfocal plane of the aberrated lens.

With a small 10–μm condenser aperture, a thin disordered carbon specimen was used to correct the aberrations in the imaging lens of a Titan^3 80–300 (FEI) aberration–corrected (CEOS GmbH) TEM, operating at 300 kV. The probe–corrector stigmator coils were grossly

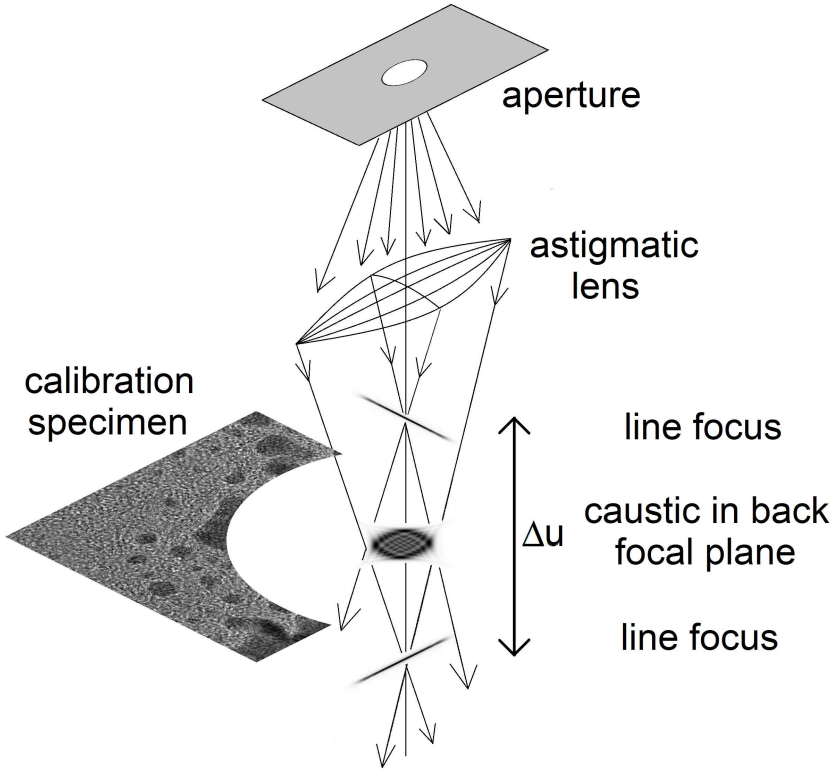

Figure 6.5: Schematic of electron matter waves diffracted by a circular aperture and focused by a lens with astigmatism. The Gouy phase anomaly describes the variation in the electron wave along the longitudinal (vertical) optic axis, through the transverse centre of the aberrated focal volume as compared to the linear variation of phase predicted by geometric optics. A thin specimen is used to correct imaging lens aberrations with a parallel probe, prior to the imposition of astigmatism in the illumination and focusing of the probe in the specimen plane. The parameter Δu is proportional to the distance along the optic axis as described in the text.

excited to produce two astigmatic line foci of sub–nanometre width, 1.5 μm apart along the optic axis. Using 100 s acquisition times, a focal series of 12 images was collected at points along the optic axis between the two line foci, with nominal defocus increments of 60 nm.

The electron phase was retrieved using the Gerchberg–Saxton–Misell algorithm (Gerchberg and Saxton, 1972; Misell, 1973). Despite the long exposure times and down–sampling the data to increase the signal–to–noise ratio, the algorithm failed to converge. To obviate this problem, rather than use a guess for the initial phase, we approximated the caustic using a diffraction integral (i.e., Eq. (3.23)) to compute the astigmatic wavefunction, which was then used to seed the initial phase in the retrieval algorithm. Furthermore, the retrieval algorithm excluded intensities far from the caustic, where the experimental data contained

only noise and were devoid of phase information. Satisfactory convergence[2] occurred after 10^3 iterations. To improve the spatial resolution of diffraction detail within the caustic, the retrieved wavefunction was then resampled to 1024×1024 pixels and used to seed a further 10^3 iterations with the experimental data down–sampled to the same number of pixels.

6.2.2 Analysis

In principle, any in-line holographic method can be used to reconstruct a desired monochromatic scalar wavefunction, provided that the solution reproduces the experimental data when propagated between focal planes. Allen et al. (2004) define a sum–squared error (SSE) to characterize the convergence for iterative phase retrieval, which has the form

$$\mathrm{SSE}_j = \sum_n \sum_m \frac{\left(\sqrt{I_j^{exp}} - \sqrt{I_j^{rec}}\right)^2}{I_j^{exp}}, \tag{6.4}$$

where the double sums are over all pixels in the jth experimental intensity I^{exp} and reconstructed intensity I^{rec}. The average of SSE_j over all N images in the focal series defines an averaged sum–square error, i.e., $\mathrm{SSE}_{av} = \sum_j \mathrm{SSE}_j/N$. Over the entire field of view for all 12 images in the focal series, the caustic reconstruction converged to $\mathrm{SSE}_{av} = 5.6 \times 10^{-2}$. For all pixels with intensity above the threshold used for iterative replacement, the error metric was measured to be $\mathrm{SSE}_{av} = 1.4 \times 10^{-2}$.

Figure 6.6 shows the results of the phase–retrieval for the astigmatic caustic. Figure 6.6(a) shows the retrieved phase of the propagated wavefunction at a focal distance in–between the two line foci, with the corresponding retrieved intensity shown in Fig. 6.6(b); these compare favourably with the experimental data in Fig. 6.6(c). Figure 6.6(d) shows the retrieved phase near one of the line foci, with the associated retrieved intensity shown in Fig. 6.6(e). The experimental data is shown in Fig. 6.6(f). Figures 6.6(e) and 6.6(f) are in good agreement, which is consistent with small values of SSE_{av} and demonstrates satisfactory convergence of the iterative phase–retrieval algorithm.

Using the retrieved wavefunction, the centres of the phase maps were tracked by propagating the electron wave in steps of 30 nm along the optic axis. The positions of the sharp line foci in the transverse intensity distributions were used to accurately determine the position of the optic axis, along which the retrieved on-axis phase was then plotted, as

[2]The convergence criterion is based on the work by Allen et al. (2004).

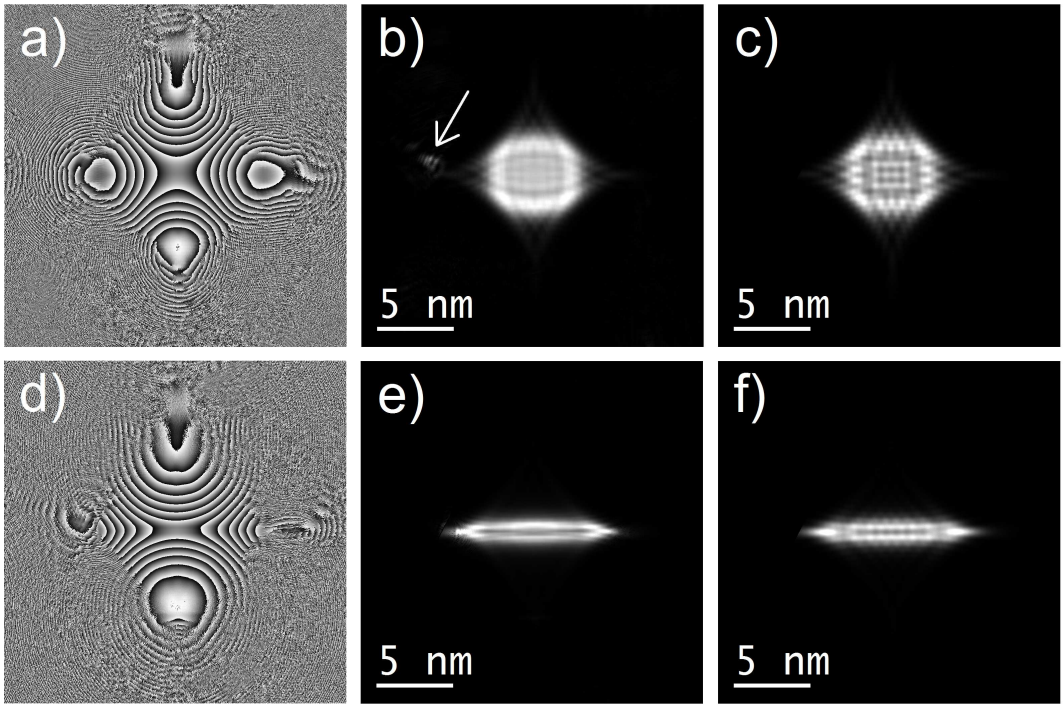

Figure 6.6: (a) Retrieved electron phase map at a focal distance in–between the two astigmatic focal lines. (b) Corresponding propagated intensity. The arrow highlights a outlying discrepancy, which is also visible in (e), arising from the finite field of view used to experimentally record the caustic. (c) Experimental intensity distribution at the same focal distance. (d) Retrieved electron phase map near one of the astigmatic focal lines. (e) Corresponding propagated intensity. (f) Experimental intensity distribution at the same focal distance.

shown by the dashed profile in Fig. 6.7. The solid–line profile in Fig. 6.7 was calculated from Eq. (6.3). Apart from the electron wavelength, three additional parameters were required for the computation; namely, the ratio of the effective aperture size to the focal length, a/f , the coefficient of astigmatism A_0, and the position of the line foci. All of these parameters were robustly determined from the intensities of the experimentally retrieved wavefunction by computing the Heisenberg uncertainties $\Delta x(z)$ and $\Delta y(z)$. Away from the line foci, $\Delta x(z)$ and $\Delta y(z)$ behave asymptotically as $T|z|$, where T is a constant representing the gradient of the uncertainty as a function of the distance z. Each measured value of T was averaged to determine A_0 using Eq. (6.3) with $a/f \cong 2T$ (Visser and Wolf, 2010). Both phase profiles contain an arbitrary vertical offset, which we have systematically chosen so that the on-axis variations approach zero in-between the two line foci.

The experimental on–axis phase profile in Fig. 6.7 follows closely the theory of Visser and

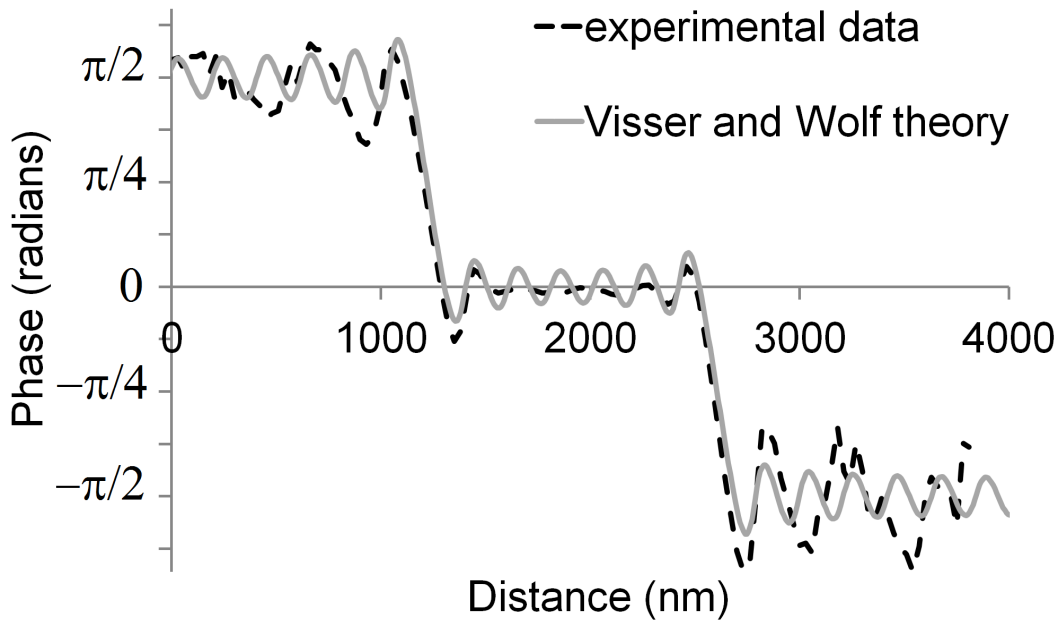

Figure 6.7: Experimentally retrieved on–axis phase (dashed curve) for a focused astigmatic electron wave compared with the Gouy anomaly predicted by Visser and Wolf (2010). Two sequential phase shifts of $\pi/2$ rad occur at each line focus due to the Gouy anomaly. Parameters required for the theory given by Eq. (6.3) were measured from the experimentally retrieved intensity distributions alone.

Wolf (2010); in particular, the slopes and horizontal positions of the rapid phase variations near each line focus match. Some differences are evident between the undulations in Fig. 6.7, which are sensitive to the effects of diffraction. These discrepancies could be ascribed to systematic errors in the phase retrieval. However, we expect differences on account of the fact that large excitation of the TEM stigmator coils does not produce pure twofold astigmatism but, rather, an astigmatic beam with a pair of line foci, which is also perturbed by coma and higher–order aberrations.

In addition to the scalar diffraction theory of focused paraxial waves, a variety of interpretations exist for the Gouy anomaly. These interpretations are discussed in detail and are brought together into a single theoretical framework in §6.3. One persistent theme is the idea that the Gouy effect arises from fluctuations in the transverse momentum, induced by variations in the uncertainty of the beam at different focal points along the optic axis (Boyd, 1980; Hariharan and Robinson, 1996; Feng and Winful, 2001). For Gaussian beams, the variation in the standard deviation is characterized by the evolution of the beam waist, and the

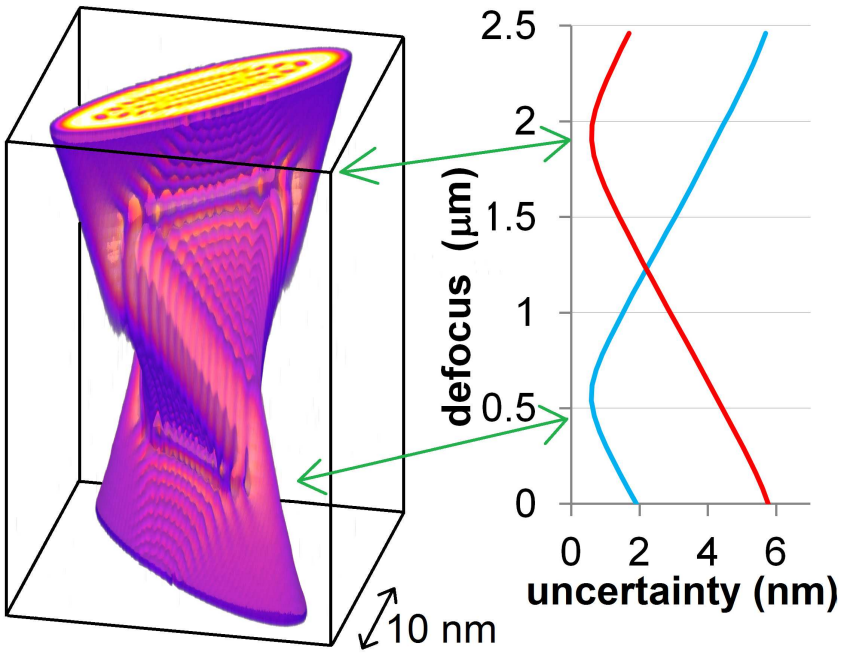

Figure 6.8: Electron probability density measured from a through-focus series of 32 images, which were stacked and interpolated along the vertical direction. The uncertainties along two transverse directions, orthogonal to each line focus, were calculated from each image. The paraxial electron wave evolves slowly along the optic axis as evidenced by the 10 nm transverse scale bar, which is to be contrasted with the micron–scale focal range along the vertical axis.

transverse intensity distribution maintains the same shape at different points along the optic axis (Siegman, 1986). Accordingly, the Gouy phase evolves along the optic axis, varying most rapidly near the focus and more abruptly for smaller beam waists. We performed another experiment to demonstrate the three–dimensional nature of this standard deviation, organized by the diffraction detail within an astigmatic volume. Similar illumination conditions were chosen, with the exception of the field-emission gun lens, the strength of which was halved, thereby increasing the probe intensity to reduce the Poisson noise in the recorded images. Thirty–two images were then acquired using 100 s exposures and nominal defocus increments of 80 nm. These images were stacked to create a tomogram of the electron wave probability density. Figure 6.8 shows a false–colour isosurface of the tomogram. The accompanying Heisenberg uncertainties were measured along directions orthogonal to the line foci and fall to a minimum at each line focus.

To summarize, we have measured the Gouy phase anomaly for astigmatic matter waves using electron wavefunction phases inferred from experimental intensities. Successive Gouy

phase shifts of $\pi/2$ rad were observed for fast electrons travelling along the optic axis, passing through two sub–nanometre lines of focus. These longitudinal phase variations were compared with wave optics theory, and consistency with experiment was demonstrated.

The origin of the Gouy phase anomaly has been analysed using a number of alternative descriptions. In the next section many of these interpretations are considered. We show how these descriptions may be unified into a single theoretical framework.

6.3 Unifying Interpretations of the Gouy Phase Anomaly for Electron Waves

For over 100 years, Gouy's quantized phase changes incurred by focused rays (Gouy, 1891) have resisted intuitive interpretation with many accurate, yet contrasting, theoretical descriptions detailed in the literature. Analytical formulations in light optics have been developed for Gaussian beams (Boyd, 1980) where the Gouy phase follows an inverse tangent function of a scaled distance along the optic axis. For beams that propagate in free space without changing shape, this characteristic inverse tangent phase variation has been derived from wave optics as a "universal form" (Borghi et al., 2004). From a quantum-mechanical point of view, a phase anomaly arises for Gaussian beams because the effective longitudinal momentum is reduced due to transverse confinement of the wave, which increases the spread of the transverse momentum (Hariharan and Robinson, 1996). Indeed, the Gouy phase has been associated with the uncertainty principle to describe changes in an effective propagation constant (Feng and Winful, 2001; Yang and Winful, 2006). Gaussian beams have also been used to describe the Gouy anomaly (Subbarao, 1995) in terms of Berry's geometric phase (Berry, 1984; Simon and Mukunda, 1993).

Certain optical devices can collapse rays from all spatial directions towards a common focal point, from which rays then diverge (Tyc, 2012). For such wide–angle lenses, a Gouy phase anomaly protects against an unphysical singularity, since the superposition of converging and diverging spherical waves must differ in phase by π radians at the focal point in order to satisfy the wave equation (Tyc, 2012). Such phase shifts more generally arise on passing through a caustic where singularities occur in geometric optics as rays overlap (Keller, 1958, 1985) and are characterized by Maslov indices (Keller, 1985; Orlov,

1981).

In this section we provide a theoretical analysis of the Gouy effect to describe quantized phase changes within the context of focused scalar waves. Motivated by the experiments presented in §6.2, we relate our observations to the rich variety of physical interpretations detailed in the literature. For fast electrons obeying the Schrödinger equation, we sought to unify these different points of view to better understand the Gouy effect and ultimately identify new ways in which to measure and exploit the phenomenon. We begin with a geometric construction to visualize the standard deviation of the transverse probability density of a focused beam and its evolution along the optic axis. Using this model, we show that the intuitive notion of the Gouy phase is incorrect; namely that it is induced by second moments of transverse momentum, measured over the entire transverse plane. Examination of local transverse momentum fluctuations in a neighbourhood containing the optic axis leads to quantitative derivations of the Gouy phase in terms of paraxial waves. Upon further scrutiny, we are able to connect several disparate interpretations of the Gouy phase phenomenon within a unified theoretical framework. Specifically, for focused paraxial waves, we relate the anomaly to statistical fluctuations, a geometric phase, semiclassical phase changes, and statistical confinement. Although Gaussian beams are used for certain numerical evaluations, we show that the quantized on–axis phase variation is independent of the explicit form of the probability density.

6.3.1 Theoretical descriptions of the Gouy phase anomaly

Figure 6.9 shows the transverse spread of rays in the vicinity of a line focus. The solid curves delineate the standard deviation in the transverse position $\Delta x(z)$, which varies along the longitudinal optic axis z. The grey dashed curves mark the asymptotes of the standard deviation $\Delta x(z) = \pm T|z|$, where $T = \tan(\theta)$ is a positive real number. The curved portion of the focal region represents the effects of diffraction in which the dashed boundary, governed by geometric optics, is broadened to account for a finite wavelength, thereby avoiding an unphysical singularity at the line focus ($z = 0$). For the distribution of ray trajectories defining the angular range θ, $T = \tan(\theta)$, and the mean transverse position $\langle x(z) \rangle$ is zero (enforcing symmetry about the z axis). The dashed lines define a quadratic variance $\left\langle x^2(z) \right\rangle = T^2 z^2$. The angular brackets refer to averages over the entire transverse plane such that, for a given

function $f(x)$ and probability density $P(x)$,

$$\langle f(x)\rangle = \int_{-\infty}^{+\infty} f(x)P(x)dx. \tag{6.5}$$

One simple way to account for nonzero variance at the line focus is to add a small positive constant ϵ so that $\left\langle x^2(z)\right\rangle = T^2z^2 + \epsilon^2$. The expression for the standard deviation $\Delta x(z)$ in the transverse position is now given by

$$\Delta x(z) = \sqrt{T^2z^2 + \epsilon^2}. \tag{6.6}$$

The spread in the transverse momentum p_x is constrained by the Heisenberg uncertainty principle $\Delta x\Delta p_x \geq \hbar/2$. In terms of the transverse component k_x of the wave–vector $\mathbf{k} = (k_x, k_y)$, de Broglie's relation gives $\Delta k_x \geq 1/(2\Delta x)$. Under paraxial conditions, we have $k_x = k\theta$ so that the angular spread $\Delta\theta$ defines the spread in k_x and Eq. (6.6) becomes

$$k\Delta\theta = \Delta k_x \geq \frac{1}{2\sqrt{T^2z^2 + \epsilon^2}}. \tag{6.7}$$

Along the optic axis, the z component k_z of $\mathbf{k}$ varies statistically on account of the variation in k_x and constant energy in the waves, where $|\mathbf{k}|^2 = k^2 = k_x^2 + k_z^2$. Averaged over the transverse axis x, we can define the propagation constant $\langle k_z\rangle$ in terms of the dispersion in transverse momentum as

$$\langle k_z\rangle = \sqrt{k^2 - \left\langle k_x^2\right\rangle} \cong k - \frac{1}{2k}\left\langle k_x^2\right\rangle = k - \frac{1}{2k}(\Delta k_x)^2. \tag{6.8}$$

The phase accumulates along the optic axis in accordance with variations in k_z. The phase predicted by ray optics is the wave number k multiplied by the distance along the optic axis. Upon subtracting this contribution we can define the average accumulated phase $\langle\phi(z)\rangle$ by

$$\begin{aligned}\langle\phi(z)\rangle &:= \int_{z_{min}}^{z}\left\langle\frac{\partial\phi}{\partial z'}\right\rangle dz' - k\int_{z_{min}}^{z} dz' \\ &= k\int_{z_{min}}^{z} dz' - \frac{1}{2k}\int_{z_{min}}^{z}\left\langle k_x^2\right\rangle dz' - k\int_{z_{min}}^{z} dz' + C,\end{aligned} \tag{6.9}$$

where C is an arbitrary constant and z_{min} is the leftmost point on the optic axis.

Equations (6.6), (6.8) and (6.9) can be combined, whence,

$$\langle\phi(z)\rangle \geq -\frac{1}{8k}\int_{z_{min}}^{z}\frac{1}{T^2z'^2 + \epsilon^2}dz' \tag{6.10}$$

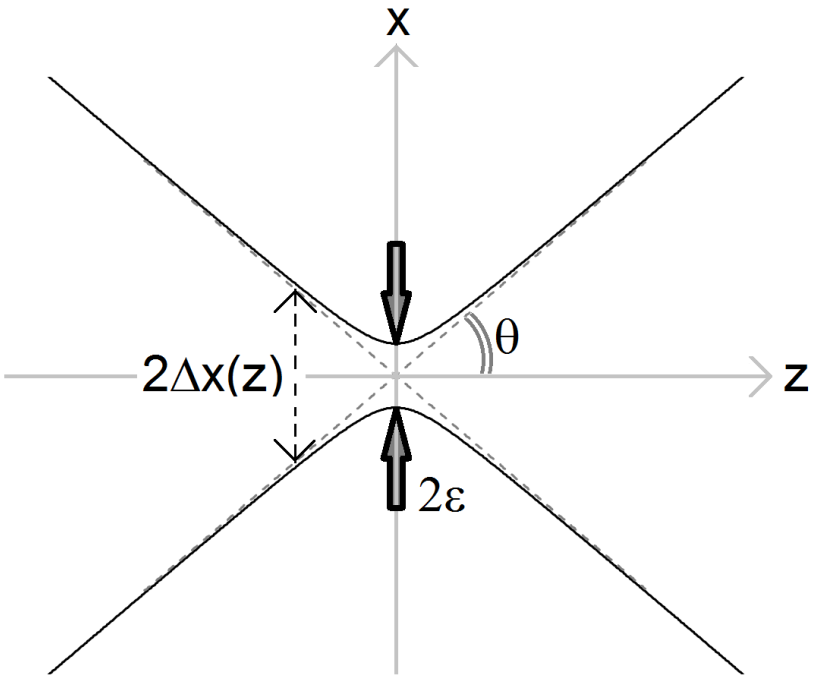

Figure 6.9: Longitudinal variation in the standard deviation $\Delta x(z)$ along the optic axis near a focal point. The parameter ϵ accounts for the effect of diffraction and the minimum transverse spread in the electron beam. The minimum spread ϵ is proportional to the beam waist for a Gaussian beam. For large z the standard deviation $\Delta x(z)$ behaves asymptotically as $T|z|$, where $T = \tan\theta$.

where the constant C has been omitted by fixing the average phase to be zero at the focus.

Taking the limit $z_{min} \to -\infty$, the definite integral in Eq. (6.10) evaluates to give

$$\langle \phi(z) \rangle \geq -\frac{1}{8k\epsilon T} \tan^{-1}\left(\frac{Tz}{\epsilon}\right). \tag{6.11}$$

At the origin, $z = 0$, the standard deviation $\Delta x(z = 0)$ in the transverse position is defined to be ϵ. Assuming that the angular spread $\Delta\theta$ is approximately equal to the semiangle ($\theta << 1$), Eq. (6.7) provides bounds on the constants in Eq. (6.11), i.e.,

$$k\theta \approx k\tan(\theta) = kT \geq \frac{1}{2\epsilon} \to \epsilon kT \geq \frac{1}{2}. \tag{6.12}$$

Equation (6.11) becomes

$$\langle \phi(z) \rangle \geq -\frac{1}{4} \tan^{-1}\left(\frac{Tz}{\epsilon}\right). \tag{6.13}$$

In three dimensions the same arguments produce an additional, but identical, term on the right-hand side of Eq. (6.10), which doubles the right–hand side of Eq. (6.13). On account of this inverse tangent variation in z, it is tempting to directly associate $\langle \phi(z) \rangle$ with the Gouy

phase anomaly. However, this bound for $\langle\phi(z)\rangle$ was derived from the minimum product in the uncertainty principle, which occurs at $z = 0$ in Fig. 6.9. The explicit value of $\langle\phi(z)\rangle$ for all z can be computed using, say, the Gaussian beam solution of the paraxial wave equation either directly or with expectation values of the transverse momentum through Eq. (6.9). In either case, we find $\langle\phi(z)\rangle = z/(2z_R)$ in two transverse dimensions where $z_R = \pi w_0^2/\lambda$ is the Rayleigh range for a Gaussian beam with minimum beam waist w_0 and wavelength λ (Siegman, 1986). Consequently, the transverse average of the phase is not equal to the Gouy phase anomaly. Generally, the linear variation in $\langle\phi(z)\rangle$ is expected from the paraxial approximation, which requires $\langle k_z^2\rangle = 0$. Hence, we must have $\langle k_z\rangle$ = constant so that, by analogy with Ehrenfest's theorem, the momentum (or energy) remains unchanged with respect to z in the absence of any forces.

We have shown that the Gouy phase ϕ_G is not equal to the integrated transverse average of the longitudinal phase gradient $\langle\phi(z)\rangle$. Using the same theoretical framework, we can inquire as to the longitudinal variation in the phase about the optic axis in the presence of statistical fluctuations. Within the paraxial approximation, one has the following three–dimensional form of Eq. (6.8):

$$\langle k_z\rangle = k - \frac{1}{2k}\left\langle k_x^2 + k_y^2\right\rangle. \tag{6.14}$$

By identifying the wave–vector components with their operators, $k_x \to -i\partial/\partial x$ and $k_y \to -i\partial/\partial y$, the expectation value can be written out in full as

$$\langle k_z\rangle = k\int_{\mathbb{R}^2} A^2 dxdy + \frac{1}{2k}\int_{\mathbb{R}^2} Ae^{-i\phi}\nabla_\perp^2 Ae^{i\phi}dxdy, \tag{6.15}$$

where A is real and represents the amplitude of a paraxial electron wave, i.e., $\xi = Ae^{i\phi}$.

By equating Eq. (6.14) and Eq. (6.15) and evaluating the integral over an infinitesimal patch of the transverse plane, the real part gives the paraxial eikonal equatiom[3] (Gureyev et al., 1995):

$$\frac{\partial\phi(z)}{\partial z} = k + \frac{1}{2k}\nabla_\perp^2 \ln(A). \tag{6.16}$$

Appendix A provides an explicit calculation of Eq. (6.16).

By choosing a Gaussian function to model an arbitrary amplitude distribution, we can show that Eq. (6.16) predicts the Gouy phase anomaly. Replacing the probability density

[3]Note that the full paraxial eikonal equation contains more terms than Eq. (6.16). However, we have assumed that the first order transverse gradients in the phase and amplitude are zero. This is justified since Eq. (6.16) is defined only over the optic axis, on which the caustic maintains an extremum for all z.

$I = |A|^2$ with a paraxial Gaussian beam (Boyd, 1980) having the same standard deviation, we obtain

$$\begin{aligned}\phi_G(z) &= \frac{\partial \phi}{\partial z} - k \\ &= \frac{1}{2k}\nabla_\perp^2 \ln\left\{\exp\left[-\frac{1}{4}\left(x2 + y^2\right) / \left(T^2 z^2 + \epsilon^2\right)\right]\right\} \\ &= -\frac{1}{2k\left(T^2 z^2 + \epsilon^2\right)}. \end{aligned} \tag{6.17}$$

Integrating the longitudinal phase changes along z, we obtain the Gouy phase anomaly, i.e.,

$$\begin{aligned}\int \left\{\frac{\partial \phi(z)}{\partial z} - k\right\} dz &= -\int -\frac{1}{2k\left(T^2 z^2 + \epsilon^2\right)} dx \\ &= -\frac{1}{2k\epsilon T}\tan^{-1}\left(\frac{Tz}{\epsilon}\right) = -\tan^{-1}\left(\frac{Tz}{\epsilon}\right) \\ &= \phi_G(z), \end{aligned} \tag{6.18}$$

where $\epsilon kT = 1/2$ for a Gaussian distribution.

Therefore, we see that transverse local momentum fluctuations arising from the uncertainty principle produce the Gouy phase anomaly $\phi(x = 0, y = 0, z) = \phi_G(z)$. Using similar considerations, we can also interpret the effect in terms of the Lévy–Leblond longitudinal phase shift due to statistical confinement in the transverse dimensions (Lévy-Leblond, 1987). The Lévy–Leblond phase shift is discussed in Appendix B.

Within the same theoretical framework, we can associate the phase anomaly with the quantized caustic phase changes described by Keller (1985), as well as the Berry phase (Berry, 1984). We begin with Keller's phase anomaly for rays that touch caustics, which are governed by Maslov indices (Keller, 1985). Keller substitutes a Wentzel–Kramers-Brillouin wave form into the Schrödinger equation in terms of a phase with a classical action (in units of $\hbar$) and with an amplitude that can be complex and multivalued. The substitution leads to semiclassical corrections for asymptotic solutions with integer or half–integer quantum numbers, applicable for separable and nonseparable systems. Following Keller, we choose the classical action, $S = kz$, of a ray with momentum directed along the optic axis and factor the wavefunction ψ as a product of this phase contribution with a complex amplitude ξ, which we will refer to as an "envelope", such that $\psi = \xi \exp(ikz)$. By demanding ψ to be single–valued, the phase of ξ is found to change by $\pi/2$ rad along a trajectory touching a

caustic (Keller, 1985).[4] For a closed–curve trajectory which touches a caustic m times, the circulation integral describes Keller's phase anomaly ϕ_K:

$$\phi_K = -i \oint_\Gamma \nabla \ln(\xi) \cdot d\mathbf{l} = -2\pi \frac{m}{4}, \tag{6.19}$$

where m is the Maslov index, $\mathbf{l}$ is a vector along the trajectory Γ, and ∇ is the three–dimensional gradient operator. We can associate Keller's anomaly ϕ_K with the Gouy phase ϕ_G by noting that the envelope ξ, by construction, obeys the paraxial equation $(\nabla_\perp^2 + 2ik\partial/\partial z)\xi = 0$ for large wave number k. If we integrate along the entire optic axis and assume paraxial conditions $|\partial^2\xi/\partial z^2| << k|\partial\xi/\partial z|$ (Allen and Oxley, 2001), Eq. (6.19) can be rewritten as

$$\begin{aligned} \phi_K &= -i \lim_{R\to\infty} \int_{-R}^{R} \frac{\partial \ln(\xi)}{\partial z} \equiv -i \oint \frac{\partial \ln(\xi)}{\partial \tilde{z}} d\tilde{z} \\ &= \frac{1}{2k} \oint_\Gamma \frac{1}{\xi} \nabla_\perp^2 \xi d\tilde{z}, \end{aligned} \tag{6.20}$$

where the limit represents the Cauchy principal value at infinity (Ablowitz and Fokasz, 1997) and we have extended the integral over the Argand plane for complex $\tilde{z}$. Equation (6.20) is formally equivalent to Eq. (6.15) evaluated over an infinitesimal patch for the Gouy phase anomaly, deduced from fluctuations in transverse momentum; this follows since the imaginary part of ξ contributes nothing to the circulation due to conservation of probability density.

We can generalize our analysis further if we reduce the essence of the Gouy anomaly to the mere existence of a quantized phase change without concern for the explicit functional form of this phase variation, which is accumulated as the wave field evolves along the optic axis. Equation (6.20) and implicitly Eq. (6.18) are proportional to the logarithmic derivative of the paraxial wave, $\xi'(z)/\xi(z) = \partial_z \ln[\xi(z)]$, where both the dash and ∂_z both refer to the longitudinal derivative with respect to z. Using the same contour of integration, we can evaluate the logarithmic derivative using Cauchy's argument principle (Ablowitz and Fokasz, 1997) and determine the scale of the phase shift without recourse to the explicit details of the Gaussian beam. Suppose that the transverse probability distribution is Gaussian, but the scale of the Gouy phase and phase variation in Eq. (6.18) is not specified, such that $\phi_G(z) = \alpha \tan^{-1}(\beta z)$, with α and β unknown at present. For a given standard deviation $\Delta x(z) = \Delta y(z) = \sigma(z)$, conservation of probability requires that the beam amplitude varies in

[4] A complex amplitude giving rise to a "Gouy type" phase shift is discussed further in Chap. 7.

proportion to $1/\sigma(z)$ in two transverse spatial dimensions. Accordingly, the natural logarithm of $\xi(z)$ takes the following form on the optic axis:

$$\begin{aligned}\ln\left[\delta\xi(\beta z)\right] &= \ln\left[\frac{\delta}{\sigma(\beta z)}e^{i\alpha\phi_G(\beta z)}\right] \\ &= \frac{1}{2}\ln\left[\frac{1}{\beta^2T^2z^2+\epsilon^2}\right] + \alpha\ln\left[\frac{1+i\beta z}{1+(\beta z)^2}\right] + \ln[\delta],\end{aligned} \tag{6.21}$$

where α, β, and δ are all constants of proportionality; the factor of $1/2$ arises from the square root in Eq. (6.6), and the term multiplying α derives from the inverse tangent form of $\phi_G(z)$. Cauchy's argument principle states that the closed contour integral of the logarithmic derivative in the complex plane is equal to $2\pi i(N-P)$ with the number of zeros N and poles P weighted by their respective multiplicities and orders (Ablowitz and Fokasz, 1997). Both N and P are trivially zero for the term containing δ in Eq. (6.21) (as is the derivative of this term). Similarly, N and P are both unity for the second term on the right in the upper half of the complex plane, so $\phi_G(z)$ does not contribute to the closed contour integral. The remaining term has no zeros and one simple pole in the upper half of the complex plane, hence the closed contour integral evaluates to give a total Gouy phase shift of $\phi_G = -\frac{1}{2}i(2\pi i)(0-1) = -\pi$ rad where the prefactor of $-i$ in Eq. (6.20) has been included. In one transverse dimension the beam amplitude varies as $[\sigma(z)]^{-1/2}$, in which case the same analysis gives $\phi_G = -\pi/2$ rad for the total Gouy shift. Hence, the argument principle fixes the scale of the Gouy anomaly.

Berry (1984) showed that, for quantum states which evolve slowly (adiabatically) and, therefore, remain in the same instantaneous eigenstate at later times, there is an associated phase factor $\gamma(t)$ which also varies in time. Upon changing a parameter in the Hamiltonian cyclically, the wavefunction will return to the same state at some later time; however, the accumulated phase will not generally match the initial value of $\gamma(t)$. In such cases, a measurable geometric phase arises. Let us associate slow adiabatic evolution with the paraxial equation for the envelope ξ, where the fast oscillation of e^{ikz} has been factored out of the wavefunction ψ. In this context, evolution refers to propagation along the optic z axis since the paraxial equation is formally identical to the time–dependent Schrödinger equation in two spatial dimensions if z replaces the time parameter t . Following Berry (1984), we can determine the evolution of the phase factor $\gamma(z)$ by substituting a formal solution for $\xi(z)$ (Teague, 1983), i.e.,

$$\xi(z) = e^{(iz/2k)\nabla_\perp^2}\xi(z=0), \tag{6.22}$$

into the paraxial wave equation,

$$\left(\nabla_\perp^2 + 2ik\frac{\partial}{\partial z}\right)e^{i\gamma(z)}e^{(iz/2k)\nabla_\perp^2}\xi(z=0) = 0, \tag{6.23}$$

which implies that

$$2ke^{i\gamma(z)}e^{(iz/2k)\nabla_\perp^2}\left(i\frac{\partial\xi}{\partial z} - \frac{\partial\gamma}{\partial z}\xi\right) = 0. \tag{6.24}$$

From Eq. (6.24), we deduce that

$$\begin{aligned}\phi_B &= \int_{-\infty}^{+\infty}\frac{\partial\gamma(z)}{\partial z}dz = -i\int_{-\infty}^{+\infty}\frac{1}{\xi}\frac{\partial\xi}{\partial z}dz \\ &= -i\int_{-\infty}^{+\infty}\frac{\partial\ln(\xi)}{\partial z}dz,\end{aligned} \tag{6.25}$$

where ϕ_B denotes the integration of phase changes described by $\gamma(z)$. These integrals can be written as circulations if we integrate over the complex plane in the same manner as for the Keller phase anomaly, since the results are formally identical. From here, Green's theorem is used to express Eq. (6.25) as a closed contour integral in the Argand plane. By stereographically back projecting this closed contour onto the Riemann sphere, Eq. (6.25) may be written as a surface integral, i.e.,

$$\phi_G = \phi_B = -i\oint\frac{\partial\ln(\xi)}{\partial\tilde{z}}d\tilde{z} = -\frac{1}{2}\iint dA = -\frac{\Omega}{2}, \tag{6.26}$$

where Ω is the solid angle subtended by the patch on the Riemann sphere and $\tilde{z}$ is the z–variable extended to the Argand plane. Note that in one dimension, the phase shift ϕ_B given in Eq. (6.26) is halved, in accordance with the observations of the Gouy phase shift for astigmatic line foci. For a full derivation of Eq. (6.26) see Appendix A.

As with our derivation based upon the argument principle, we find that the phase changes in Eq. (6.26) are invariant to an arbitrary change in the Gaussian beam phase scale and to that in the complex focus $\tilde{z}$. Similarly, we are free to distort the Gaussian wave envolope by multiplying ξ by any function that has an as analytic logarithmic derivative, without changing the value of the quantized phase; this is a direct consequence of Cauchy's integral theorem. In other words, when we integrate over half the Cartesian plane or perform a contour integrating along the entire real (optic) axis, the magnitude of the Gouy phase anomaly is explicitly $\pi/2$ rad in one transverse dimension and π rad in two transverse dimensions.

6.3.2 Conclusion

Within the context of focused matter waves, our theoretical description of the Gouy anomaly unites Heisenberg's uncertainty principle, Berry's geometric phase (Berry, 1984) and Maslov indices for rays that touch caustics (Keller, 1958, 1985). If we denote the phase change from transverse fluctuations in momentum as a "Heisenberg phase," then, for focused paraxial electron beams, the Gouy, Heisenberg, Keller, Berry, and Lévy–Leblond phases all are seen to be manifestations of the same phenomenon.

7 Catastrophes on Order–parameter Manifolds

7.1 Introduction

Up to this point we have highlighted singularities in the context of optical and electron wave fields. The current chapter applies these ideas to more abstract scenarios, in which caustics are described as occurring in order–parameter spaces. The concept of an order–parameter was introduced in §1.1.2, where its role in determining the topological stability of a vortex was discussed. Here we show that the order–parameter caustics are a topologically stable defect. However, in this case the stability of these defects originates from the intrinsic stability of the caustics, rather than from the topology of the order–parameter space. Aided by the principles of catastrophe theory discussed in Chap. 3, we outline a theoretical formalism for the formation and unfoldings of entirely new and hitherto unexplored catastrophes on order–parameter manifolds.

Order–parameters are pivotal to the study and classification of topological defects and phase transitions (Sethna, 2006). Many important areas of physics can be understood using this construct, e.g., crystal growth (Mermin, 1979), quantum computing (Golovach et al., 2010), nematic liquid crystals (Pieranski et al., 2013), magnetic textures (Röszler et al., 2006; Mühlbauer et al., 2009; Ezawa, 2010), multiferroics (Wang et al., 2003; Kimura et al., 2003; Chu et al., 2008), superconductors (Gull et al., 2013; Seibold et al., 2012) and superfluid Helium (Levitin et al., 2013), to name but a few. The order–parameter may be viewed as a function that maps points in the physical space to the order–parameter space. The topological properties of the order–parameter space are important in the study of defects, since these properties govern the stability of the defect (Ruben et al., 2010).

Here we demonstrate the existence of a new type of singularity that becomes manifest

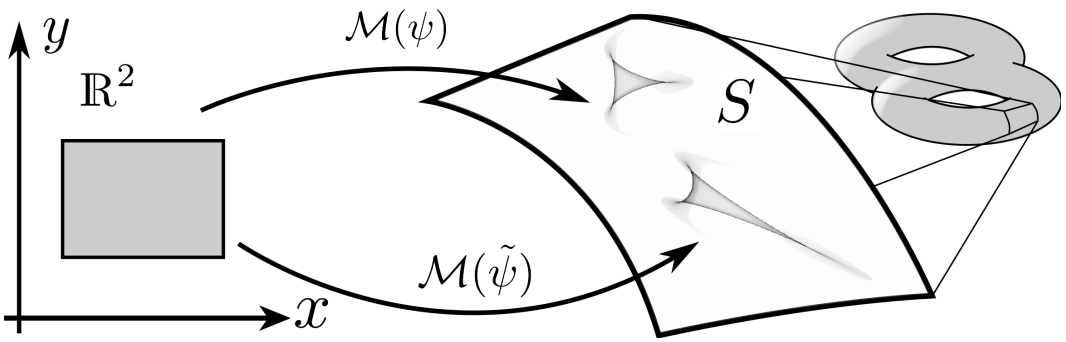

Figure 7.1: Schematic showing how the field ψ in an $\mathbb{R}^2$ patch is mapped to its order–parameter manifold, leading to a catastrophe on S. Here S is a local patch of the manifold, which may possess a non-trivial global topology. Due to the intrinsic stability of the catastrophe, a mapping of the patch of the perturbed field $\tilde{\psi}$, would yield a deformed catastrophe. The features of the field that are mapped to the catastrophe persist in the presence of perturbations.

when a field is mapped to its order–parameter space. These defects are governed by the framework of catastrophe theory (Arnol'd, 1975; Saunders, 1980; Poston and Stewart, 1996; Thom, 1983). Since catastrophes are stable with respect to perturbation (see §3.1), features of the field that map to a catastrophe on the order–parameter manifold must also be stable to perturbations. This singularity represents a new type of topological defect, which we refer to as an order–parameter catastrophe defect (OPCD) (see Fig. 7.1).

Catastrophes possess their own topology regardless of the topology of the order–parameter manifold on which they exist. Their stability is intrinsic to the catastrophe defect, rather than depending on the topology of the order–parameter manifold. This implies that local regions of the field that map to the order–parameter catastrophe defect, transcend the topology of the order–parameter space, due to the intrinsic topology of the catastrophe itself. Catastrophe theory may be applied to any system whose order–parameter space is "tattooed" with such catastrophes.

As we have seen in previous chapters, catastrophes appear naturally in many physical systems and are generic structures (Arnol'd, 1975; Berry, 1976; Thom, 1977; Nye, 1978; Berry and Upstill, 1980; Hannay, 1982; Maeda et al., 1994; Nye, 2003a,b, 2006; Petersen et al., 2013b; Simula et al., 2013). OPCDs are also expected to be universal phenomena, which arise as local topological defects that persist within the order–parameter manifold associated with the relevant field. In addition to the broad range of systems mentioned above, we can also expect to find OPCDs in abstract physical systems involving quantum fields,

cosmological phase transitions and stochastic fields, or indeed in any area of physical science in which a field can be mapped to its order–parameter manifold. For example, evidence of such catastrophes has recently been observed in scalar optical fields (Rothschild et al., 2012, 2014), in which the Argand plane of the wave field is identified with the order–parameter manifold. An exposition of how OPCDs arise on order-parameter manifolds is non–trivial and worthy of a detailed exploration. From the wide variety of applicable physical systems, here we choose to illustrate order-parameter catastrophes for a 2D spin–1/2 field, for which the corresponding order–parameter manifold is the Bloch sphere.

7.2 Bloch Sphere Mapping

Consider the general wavefunction of a two–component spinor

$$|\Psi(\mathbf{x},t)\rangle = \psi_0(\mathbf{x},t)\,|0\rangle + \psi_1(\mathbf{x},t)\,|1\rangle\,, \tag{7.1}$$

where $\mathbf{x}$ is the spatial variable and t denotes the time coordinate; ψ_0 and ψ_1 are the probability amplitudes of the spin up state $|0\rangle$ and spin down state $|1\rangle$, respectively. Without loss of generality we may assume $|\psi_0|^2 + |\psi_1|^2 = 1$, and consider only pure spin states that map to the surface of the Bloch sphere. A single spin state is represented graphically as a unit position vector of the Bloch sphere, denoted by $|\mathscr{B}\rangle$ (see Fig. 7.2). It is parametrized by the spherical polar angles (ϕ, θ), i.e.,

$$|\mathscr{B}\rangle = e^{i\chi}[e^{-i\phi(x,y)/2}\cos\left(\theta(x,y)/2\right)|0\rangle + e^{i\phi(x,y)/2}\sin\left(\theta(x,y)/2\right)|1\rangle], \tag{7.2}$$

where the global phase factor $\exp(i\chi)$ is an unobservable of the system (Martinis et al., 2003; Xiang et al., 2005; Nielsen and Chuang, 2010).

Here we are concerned with the mapping of the Bloch vector, located at every point in coordinate space, to the Bloch sphere. Consider the mapping of an (x, y)–patch of Bloch vectors, corresponding to a spin–$1/2$ field, to the surface of the Bloch sphere. If the patch contains a texture defect it will wrap the surface of the Bloch sphere an integer number of times (Mermin, 1979; Mäkelä et al., 2003; Kawaguchi et al., 2008; Zhang et al., 2009). Perturbing the field locally will not destroy the texture structure, since it is topologically protected. Instead, the mapping is altered in such a way that a local (x, y)–patch can be sheared, dilated or rotated by the mapping, whilst still fully wrapping the Bloch sphere. To

accommodate transformations of this type, a small region on the Bloch sphere must fold back onto itself, resulting in a fold catastrophe. However, the converse is not necessarily true. To see this, consider a two–dimensional periodic spin arrangement. This field has the topology of a torus, and so cannot wrap the Bloch sphere. The catastrophes are insensitive to the global topology of the order–parameter manifold and hence may still be present in local regions of the mapping.

7.3 Conditions for a Catastrophe

The explicit form of the Bloch sphere map is determined by the inverse of Eq. (7.2), given by

$$\phi(x, y) = \arg\left(\psi_1/\psi_0\right), \tag{7.3}$$

$$\theta(x, y) = \arccos\left(\eta\right). \tag{7.4}$$

The quantity $\eta = \psi_0^*\psi_0 - \psi_1^*\psi_1$ is the spin asymmetry, which represents the difference in probability density of the spin up and spin down components. The Bloch sphere image is a 2D histogram of ϕ and θ, each of which are calculated for every (x, y) point of the spin field. The mapping takes a patch of the (x, y)–plane and calculates the complex functions $\psi_0(x, y)$ and $\psi_1(x, y)$. Together these functions completely define the spin–$1/2$ field over the Euclidean patch in $\mathbb{R}^2$. These states are then mapped to the corresponding (ϕ, θ)–patch on the Bloch sphere.

We demonstrate the preceding ideas via a generic example. A randomly varying spinor wavefunction was generated by low–pass filtering an image of white noise for each spin component, producing the field shown in Fig. 7.3(a). By applying Eqs. (7.3) and (7.4) the field was mapped to the Bloch sphere to produce Fig. 7.3(b). The presence of several phase vortices can be seen in the Bloch sphere coordinate $\phi(x, y)$, shown in real space in Fig. 7.3(c). These screw–type topological defects come from the vortices in the phase of the two–component spinor definied by Eq. (7.1), which are preserved when $\phi(x, y)$ is calculated.

A Bloch sphere singularity corresponds to a many–to–one mapping. This singularity occurs when the Jacobian determinant of the mapping vanishes, where the Jacobian matrix is defined by

$$J(x, y) = \begin{pmatrix} \frac{\partial\phi}{\partial x} & \frac{\partial\phi}{\partial y} \\ \\ \frac{\partial\theta}{\partial x} & \frac{\partial\theta}{\partial y} \end{pmatrix}. \tag{7.5}$$

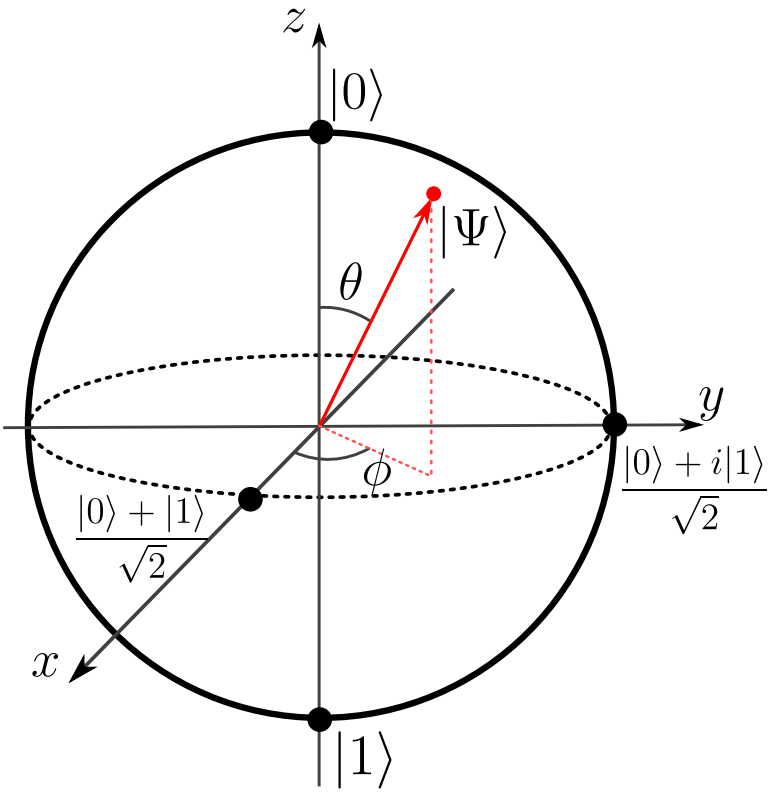

Figure 7.2: The Bloch sphere is the order–parameter manifold for a two–component wavefunction, since all possible states of Eq. (7.1) correspond to points on its surface (for pure states) or points within its volume (for mixed states). It is parameterised by the two basis states, $|0\rangle$ (spin up) and $|1\rangle$ (spin down). The Bloch vector $|\mathscr{B}\rangle$, which describes each spin state as a point on the Bloch sphere, may also be represented in spherical polar coordinates.

Setting $\det[J(x,y)] = 0$ gives the condition for a particular spin state to map to a catastrophe, i.e.,

$$\left[\left(|\psi_0|^2\,\boldsymbol{\nu}_1 - |\psi_1|^2\,\boldsymbol{\nu}_0\right) \times \nabla\eta\right] \cdot \hat{\boldsymbol{z}} = 0, \tag{7.6}$$

where $\hat{\boldsymbol{z}}$ is a unit vector in the z–direction; $\boldsymbol{\nu}_0$ and $\boldsymbol{\nu}_1$ are the respective probability current densities of ψ_0 and ψ_1, given by

$$\boldsymbol{\nu}_i = \frac{i\hbar}{2m}\left(\psi_i \nabla\psi_i^* - \psi_i^* \nabla\psi_i\right). \tag{7.7}$$

Equation (7.6) specifies the condition for the formation of a Bloch sphere catastrophe and represents the main result of this chapter. The interpretation of Eq. (7.6) is aided by constructing an auxiliary function, defined by

$$\Gamma(x,y) = \sqrt{\eta}\exp(i\phi). \tag{7.8}$$

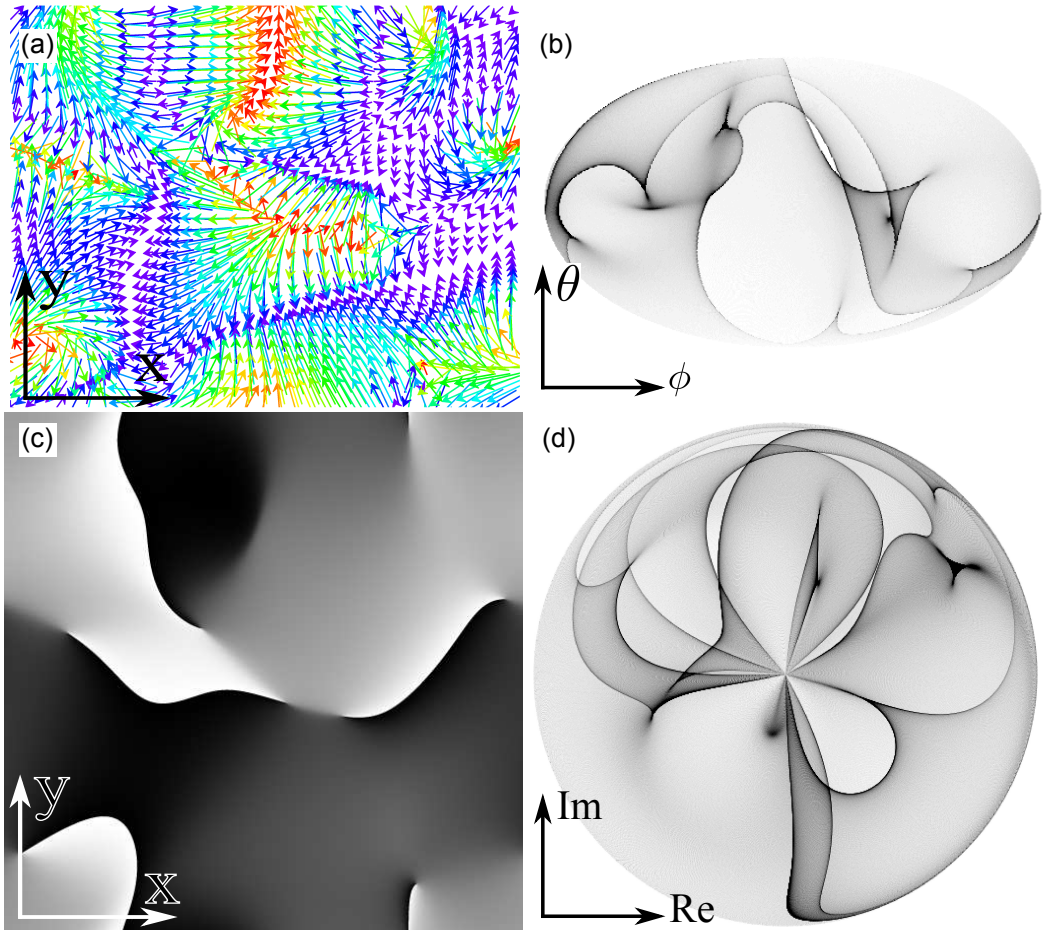

Figure 7.3: (a) Bloch vectors of a random spin–1/2 field distributed in real space with the colour of the vector corresponding to the z–component. (b) Mollweide projection of the Bloch sphere map. Here the density of the plot is proportional to the number of spin states of the spin–1/2 field shown in (a). (c) The Bloch sphere coordinate ϕ displayed as a 2D function of (x, y). (d) The Γ–map of the spin field in (a), which can be thought of as a projection mapping the Bloch sphere to the Argand plane (see main text).

It can be shown that, up to a scaling factor, Eq. (7.6) is equivalent to the z–component of the curl of the probability current density of Γ, i.e., the vorticity of Γ.

Equation (7.8) is a projection that maps the Bloch sphere to the Argand plane; this Γ–map is shown in Fig. 7.3(d). Zeros of Γ are mapped to the equator of the Bloch sphere, while the entire equator is mapped to the origin of the Argand plane. Note that the mapping in Eq. (7.8) preserves the local structure of the catastrophes, as can be inferred by comparing Figs. 7.3(a) and 7.3(b). Each hemisphere is projected over a unit disc, with the poles mapping to the disc's circumference at $|\Gamma| = 1$. However, each hemisphere of the Bloch sphere is rotated

relative to the other by $\pi/2$ radians. This is a consequence of Γ having a complex amplitude. Since η may be negative, corresponding to where the Bloch vector crosses the equator of the Bloch sphere, Γ undergoes an abrupt phase shift of $\pi/2$. We note the close similarity between this phase shift in Γ and the Gouy phase shift discussed in Chap. 6. Specifically, the function ψ as defined in §6.3.1, also possesses a complex amplitude corresponding to a ray passing though a caustic (Keller, 1985). For Bloch sphere caustics, Γ acquires a "Gouy type" phase shift when the Bloch vector touches the equator.

The function Γ allows us to express the equation governing order–parameter space defects of the spin–$1/2$ field in terms of its curl. The current density of Γ has the form $\eta\nabla\phi$, which may be interpreted as describing the flow of spin asymmetry. Catastrophes are mapped to the Bloch sphere along points where the curl of Γ vanishes.

7.4 Relation to the Canonical Form

Thom's theorem identifies a number of elementary catastrophes — these being all possible topologically distinct catastrophes that occur when the codimension of the catastrophe is less than or equal to four (Thom, 1983) (see Chap. 3). Every elementary catastrophe has an associated function, known as the canonical form (Thom, 1983; Arnol'd, 1975; Saunders, 1980; Poston and Stewart, 1996). If a local patch of a spin–$1/2$ field maps to a particular elementary catastrophe on the Bloch sphere, the field itself must be related to the canonical form of the catastrophe. The Bloch sphere catastrophe that appears in the coordinates (ϕ, θ) is only a two–dimensional slice of the entire caustic surface associated with one of Thom's elementary catastrophes.

The catastrophes on the Bloch sphere are singularities of the Bloch sphere mapping. Hence points of the spin field, where the Jacobian determinant vanishes, map to a Bloch sphere catastrophe. In catastrophe theory it is the degenerate critical points of the canonical form that map to the caustic surface. Points where the spin–field has a vanishing Jacobian determinant are locally equivalent to the Hessian of the germ of the catastrophe $V(x, y)$, i.e.,

$$\begin{pmatrix} \frac{\partial\tilde{\phi}}{\partial x} & \frac{\partial\tilde{\phi}}{\partial y} \\ \\ \frac{\partial\tilde{\theta}}{\partial x} & \frac{\partial\tilde{\theta}}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 V}{\partial x^2} & \frac{\partial^2 V}{\partial y \partial x} \\ \\ \frac{\partial^2 V}{\partial x \partial y} & \frac{\partial V}{\partial y^2} \end{pmatrix}, \tag{7.9}$$

where $\tilde{\phi}$ and $\tilde{\theta}$ are the transformed coordinates of the Bloch sphere, which account for

the local translation and rotation of the catastrophe from its canonical form and $V(x, y)$ is chosen to represents an arbitrary catastrophe germ as it may also be thought of as a type of potential function. To account for smooth deformations of the catastrophe of the Bloch sphere mapping, the coefficients of the transformations of ϕ and θ must be functions of the state space coordinates (x, y). Since the transformation matrix has an inverse, this allows us to express the Bloch sphere coordinates in terms of the transformed coordinates as

$$\begin{pmatrix} \phi \\ \theta \end{pmatrix} = \begin{pmatrix} a(x,y) & b(x,y) \\ c(x,y) & d(x,y), \end{pmatrix} \begin{pmatrix} \tilde{\phi} \\ \tilde{\theta} \end{pmatrix}, \tag{7.10}$$

where the inverse transformation matrix elements a, b, c and d are arbitrary functions of (x, y), which encapsulate the local rotations, translations or deformations of the coordinate system.

Given Eqs. (7.9) and (7.10) we find that $\phi(x, y)$ and $\theta(x, y)$ may be expressed in terms of the partial derivatives of the canonical form of the catastrophe:

$$\phi(x, y) = a(x, y)V_x(x, y) + b(x, y)V_y(x, y) \tag{7.11}$$

$$\theta(x, y) = c(x, y)V_x(x, y) + d(x, y)V_y(x, y), \tag{7.12}$$

where $V_x \equiv \partial V/\partial x$ and $V_y \equiv \partial V/\partial y$. Equations (7.11) and (7.12) show that the Bloch sphere catastrophe is indeed a form of gradient mapping. If a small patch on the spin field in state space maps to a catastrophe, the field in that patch of state space must relate to the canonical form of the catastrophe. The field in these local regions has the functional form:

$$|\Psi\rangle = e^{-\frac{i}{2}[(aV_x+bV_y)}\cos\left[(cV_x + dV_y)/2\right]|0\rangle + e^{\frac{i}{2}(aV_x+bV_y)}\sin\left[(cV_x + dV_y)/2\right]|1\rangle\,. \tag{7.13}$$

The Bloch sphere coordinates (ϕ, θ) are deformed gradient maps of the canonical form of the catastrophe; the structure of the zeros of the Jacobian determinant dictates the type of elementary catastrophe to which the field will map. In essence, the lines of zeros determine where to fold a patch in $\mathbb{R}^2$, as it is mapped to the Bloch sphere. The way in which the patch is folded yields a particular type of catastrophe. Examples of elliptic and hyperbolic umbilic catastrophes are shown in Fig. 7.4, where the geometry of the Jacobian determinant of the mapping is highlighted in the insets. The canonical form of the elliptic umbilic catastrophe is $x^3 - xy^2$, where the null set for the Hessian determinant of this form corresponds to the equation of an ellipse, i.e., the zeros of the Jacobian determinant map to an elliptic umbilic

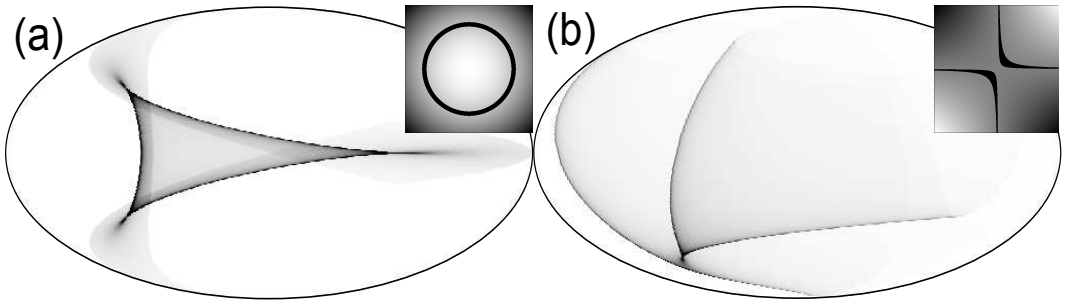

Figure 7.4: Mollweide plots of (a) elliptic, and (b) hyperbolic umbilic Bloch sphere catastrophes. The insets on each figure show the Jacobian determinant of the mapping, with the zeros highlighted in black. When the zeros of the Jacobian form an ellipse, the map of that region forms an elliptic umbilic catastrophe. Similarly, when the zeros form a hyperbolic curve, the field maps to the hyperbolic umbilic catastrophe.

catastrophe. Similarly the Hessian determinant of the hyperbolic umbilic catastrophe is the equation of a hyperbola, given by its canonical form $x^3 + y^3$. We can also use the geometries of the zeros, associated with Eq. (7.6), to visualize the spin arrangements corresponding to the elliptic umbilic and hyperbolic umbilic OPCDs (see Fig. 7.4). Figures 7.5(a) and (b) display the respective spatial distribution of the Bloch vectors, which create OPCDs in Figs. 7.4(a) and (b). The corresponding spin arrangements of the elliptic umbilic and hyperbolic umbilic OPCDs can be visualized by masking these spin fields where the Jacobian determinant vanishes. The resulting isolated spin defects correspond to only those Bloch vectors that map to the caustic surfaces in Fig. 7.4. Consider the OPCD of the elliptic umbilic catastrophe shown in Fig. 7.5(c) and choose an arbitrary point which lies on the defect. As we traverse the path where Eq. (7.6) is satisfied, the vectors undergo a full 2π rotation in both order–parameter variables θ and ϕ. The same is observed with the OPCD of the hyperbolic umbilic catastrophe, given in Fig. 7.5(d). The distinction, here, is that only in the x–direction do the vectors rotate though the order–parameter variable ϕ. Whilst the θ rotation of the vectors occurs in the y–direction. This behaviour alludes to a type of wrapping, similar to conventional topological defects. In the latter case the phase winding about a vortex point fully wraps the 1D sphere (see §1.1.2). However, whilst these vectors rotate a full 2π within the spin arrangement of the OPCDs, they map to caustic surfaces. This behaviour is due to the fact that whilst we visualize the Bloch vectors on a 2–sphere, the spinors exists in SU(2), which has the topology of a 3-sphere. This extra degree of freedom that spinors possess is made evident by the Bloch sphere mapping, which highlights why the Bloch sphere mapping

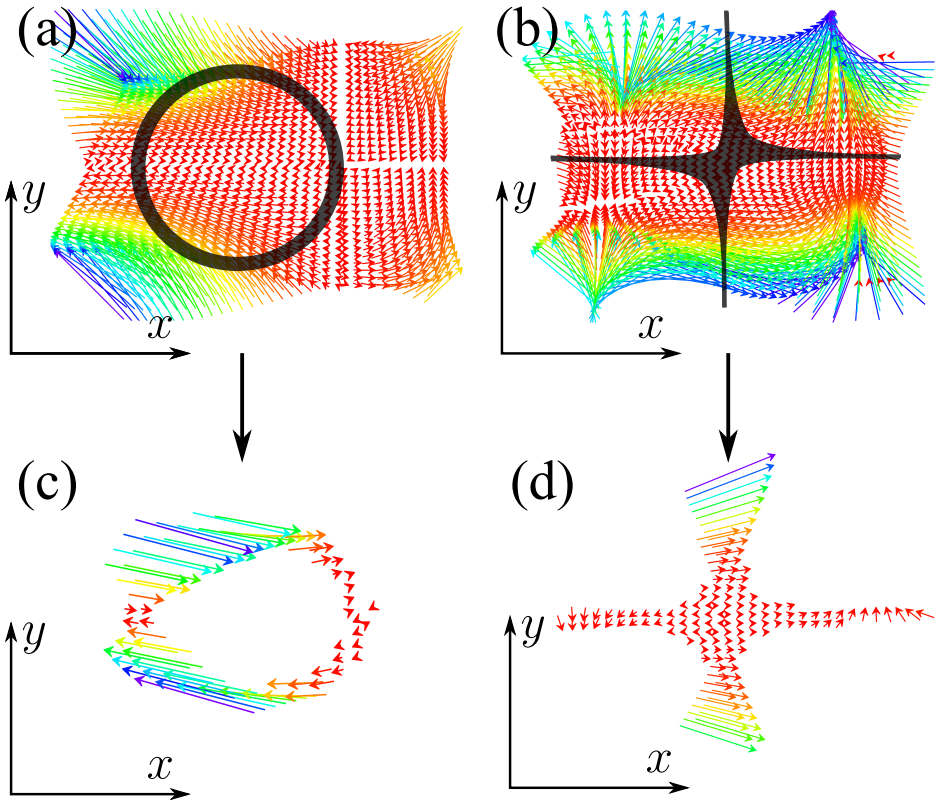

Figure 7.5: Visualization of the elliptic umbilic and hyperbolic umbilic OPCDs. (a) and (b) show the respective spatial distributions of the Bloch vectors that produce the Bloch sphere mappings in Figs. 7.4(a) and (b). The semi–transparent black masks on the spin field indicate where the Jacobian determinant vanishes. Using these masks to isolate spins corresponding to points where the Jacobian determinant vanishes allows the visualization of the elliptic umbilic OPCD (c) and the hyperbolic umbilic OPCD (d).

must be used in order to identify these "hidden" topological defects. Since we have shown that the spin field is related to the canonical form of the catastrophe by a gradient map, perturbations of the spin–1/2 field behave as a particular unfolding of the canonical form.

7.5 Stability of the Caustic Surface of Bloch Sphere Catastrophes

Bloch sphere catastrophes represent 2D slices of the full caustic surface associated with each type of elementary catastrophe, which are topologically protected. Perturbation of the spin field preserves the catastrophe and simply changes the section of the caustic surface on the Bloch sphere. To demonstrate this consider the full caustic surface of the hyperbolic

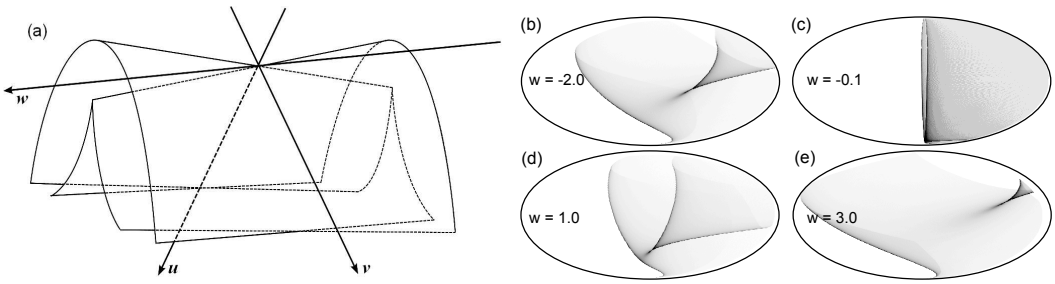

Figure 7.6: (a) Caustic surface of the hyperbolic umbilic catastrophe. (b)–(e) Bloch sphere projections showing a hyperbolic umbilic catastrophe perturbed through various values of w. Note that the order–parameter manifold is only partially covered in this example. The presence of the OPCD highlights that, whilst the spin arrangement as a whole is unstable, the local region associated with the hyperbolic umbilic catastrophe is stable.

umbilic catastrophe shown in Fig. 7.6(a). By inputting the known canonical form and using Eq. (7.13), we can map an isolated hyperbolic umbilic catastrophe to the Bloch sphere and control its unfolding. This is demonstrated in Figs. 7.6(b)–(e). This method of spin–engineering by way of using a catastrophe's canonical form could be used to generate any Bloch sphere mapping at a given unfolding. In this example the cross sections denote the $u - v$ plane. As the control parameter w is varied, the $u - v$ plane shifts along the w–axis, which perturbs the Bloch sphere catastrophe. In general, a hyperbolic umbilic Bloch sphere catastrophe will express itself as cross sections of the caustic surface through some arbitrary plane.

7.6 Discussion

Order–parameter catastrophe defects are a new form of topological defect that are associated with fields on order–parameter manifolds. Catastrophes are extremely general structures that almost always arise when mapping physical fields to their corresponding order–parameter manifold. Equations (7.11) and (7.12) show that the order–parameter coordinates are proportional to the first partial derivatives of a catastrophe's canonical form, which are always simple polynomial functions. For the spin–1/2 exemplar, this means that both ϕ and θ will also have a local polynomial form for regions that map to a catastrophe. If we consider any sufficiently small local expansion about a point of an arbitrary spin field, the patch only needs to be diffeomorphic to the canonical form of one of the elementary catastrophes. This

criterion is satisfied for the creation of a catastrophe and demonstrates that a mapping of a generic field to its order–parameter space will result in caustic "tattoos" on its manifold. This is a general result that can be found in other types of mappings. For example, catastrophes manifest themselves as Argand–plane caustics, which arise when a complex scalar field is mapped to the Argand–plane (Rothschild et al., 2012, 2014). In this context, the condition, $\det[J(x, y)] = 0$, that a point maps to a caustic is equivalent to the optical vorticity vanishing.

Whilst we have chosen to illustrate OPCDs using pure states, the general concept can be readily extended to include mixed states, when the magnitude of the spinor component in Eq. (7.2) adopts values less than unity. The additional Bloch sphere coordinate then provides another dimension in which to view the caustic surface. For example, if a three-dimensional spinor field is mapped to the Bloch sphere, the entire geometry of three–dimensional catastrophes (e.g., the elliptic and hyperbolic umbilic catastrophes), could be observed in full.

Whilst we have used the specific example of 2D spin–1/2 fields and the Bloch sphere, it is important to emphasize that all the concepts presented in this chapter are very general. We need only ask what is the mapping of any field to its order–parameter space to be able to view the field through the new perspective of order–parameter catastrophe defects. This concept is not restricted to 2–dimensions; the 2D surface of the Bloch sphere allows us to view 1D or 2D catastrophes, such as folds or cusps, or 2D cross sections of higher order catastrophes, such as the elliptic or hyperbolic umbilic catastrophes (see Fig. 7.4). For mappings to order–parameter spaces of higher dimensionality, the full bifurcation sets or higher dimensional cross sections of catastrophes with higher codimension would also become observable.

The stability of OPCDs is attributed to the intrinsic topology of the catastrophes themselves, rather than that of the order–parameter space, as is the case for conventional topological defects. Whilst our treatment of OPCDs considered the Bloch sphere as an exemplar, the principal finding has wide applicability; it provides a novel framework for elucidating phenomena in any physical system in which a real or complex field is mapped to its order–parameter manifold.

8 Conclusions and Future Work

The occurrence of singularities in nature is a generic phenomenon. In this thesis we have seen how caustics and vortices — the singularities of rays and waves — form complementary singularities, whence the natural formation of one does not occur in isolation to the formation of the other. However, they represent singularities of different theories, which operate at different length scales, giving rise to a type of singularity "uncertainty principle". We have shown that vortices and caustics may be exploited separately, using their unique properties for applications in imaging and measurement.

Vortices have been considered undesirable in optical fields, when imaging is concerned. However, Chap. 4 presented a modified holography technique, which used vortices to reconstruct an image of a phase–amplitude object. Using a three–beam interferometer, scalar plane waves were interfered to produce a uniform optical vortex lattice. Passing one of the beams through an object perturbed the optical vortex lattice, allowing the phase of the object to be determined locally by measuring the displacement of each vortex. Our method was demonstrated on a spherical lens and the wing of a house fly. With its robustness to noise being a key feature of this method, it would be interesting to quantify the impact of noise on the reconstructed images.

In Chap. 5 we applied the ideas of phase measurement using vortices to consider singularimetry with caustics. Using catastrophe theory, we showed that it was possible to recast the ill–posed inverse problem of determining the phase of a wave from measurements of the fold caustic, into a well–posed one. Numerical simulations were used to validate the proposed technique. However, further work is required to extend the technique to explore higher codimensional catastrophes, as only a fold catastrophe was considered in this thesis.

One of the many goals of the thesis was to produce vortices and vortex lattices using a transmission electron microscope. Chapter 6 we presented an experiment that exploited the duality between vortices and caustics. Aberrations were introduced to the lenses of the electron microscope, which focused the beam, giving rise to electron diffraction catastrophes. Phase retrieval performed on a through–focal series showed that vortices had indeed formed. In future work we intend to exploit specifically designed apertures to create "tailored" vortex lattices in a TEM, thereby allowing the singularimetry technique to be applied to high resolution imaging of electromagnetic fields. In Chap 6 also presented the first direct measurement of the Gouy phase anomaly for matter waves. Astigmatism applied to the magnetic lenses of a conventional TEM was used to create two perpendicular line caustics. Iterative phase retrieval determined that the phase shifted by $\pi/2$ once the electron beam passed though each line focus. This chapter also presented multiple interpretations of the Gouy phase anomaly, unifying them within a single theoretical framework.

Finally, in Chap. 7 we investigated caustics in more abstract fields. It was shown that caustics "tattoo" the surface of the manifold, when a field is mapped to its order–parameter space. Furthermore, it follows from catastrophe theory that these new catastrophes on the order–parameter manifold are stable with respect to perturbations. We have called this new and largely unexplored topological phenomenon order–parameter catastrophe defects (OPCDs). Chapter 7 investigated OPCDs on the Bloch sphere — the order–parameter space corresponding to spin–1/2 fields. Equations governing the conditions for the existence of OPCDs and the relationship to Thom's theorem were also presented. Since we have only considered a single application of OPCDs, there is still much scope for investigating this concept in various other physical contexts.

One of the encompassing aspects of singular optics is how all types of singularities and defects, in many various guises, share an underlying interconnectedness. This aspect is no more apparent than for the duality between caustics and vortices — the singularities of rays and waves. The research reported in this thesis reflects on this duality between caustics and vortices, with this theme permeating and influencing all key ideas of the research programme.

Appendix A

A.1 Derivation of the Paraxial Eikonal Equation

We begin with the expectation values of the z component of a paraxial wave–vector, given by Eqs. (6.14) and (6.15) in §6.3, i.e.,

$$\begin{aligned}\langle k_z \rangle &= k - \frac{1}{2k}\left\langle k_x^2 + k_y^2 \right\rangle \\ &= k\int_{\mathbb{R}^2} A^2 dxdy + \frac{1}{2k}\int_{\mathbb{R}^2} Ae^{-i\phi}\nabla_\perp^2 Ae^{i\phi} dxdy, \end{aligned} \tag{A.1}$$

where A is real and represents the amplitude of a paraxial electron wave, i.e., $\xi = Ae^{i\phi}$.

The two-dimensional Laplacian operator $\nabla_\perp^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$ creates a mixture of first– and second–order partial derivatives in the amplitude A and phase ϕ. For a symmetric beam, all first-order partial derivatives are zero on the optic axis, so the integrand in Eq. (A.1) can be written as

$$Ae^{-i\phi}\nabla_\perp^2 Ae^{i\phi} = A\nabla_\perp^2 A + iA^2\nabla_\perp^2\phi. \tag{A.2}$$

The effective propagation constant can be complex, with the imaginary part $\mathrm{Im}(\tilde{k}_z)$ describing the amplitude variation and the real part $\mathrm{Re}(\tilde{k}_z)$ giving the longitudinal phase derivative. With this in mind, we now consider the expectation value in Eq. (A.1) over an infinitesimal patch of the transverse plane and equate the integrands to obtain

$$\begin{aligned} A^2\tilde{k}_z = A^2\frac{\partial\phi(z)}{\partial z} - iA^2\frac{\partial A^2}{\partial z} &= kA^2 + \frac{A^2}{2k}\frac{1}{\xi}\nabla_\perp^2\xi \\ &= kA^2 + \frac{1}{2k}A\nabla_\perp^2 A + i\frac{1}{2k}A^2\nabla_\perp^2\phi, \end{aligned} \tag{A.3}$$

where $\tilde{k}_z$ denotes a complex variable and Eq. (A.2) has been used to produce the final expression. The longitudinal phase gradient in Eq. (A.3) is consistent with the local momentum, $\mathrm{Re}(\xi^*\partial\xi/\partial z)/A^2$ of the paraxial envelope ξ (Berry, 2009), for which we also have

$\mathrm{Im}(\xi^* \partial\xi/\partial z)/A^2 = \ln(\partial A/\partial z)$ since transverse derivatives of ξ are zero on the optic axis for a symmetric beam. The real part of Eq. (A.3) gives the paraxial eikonal equation (Gureyev et al., 1995):

$$\begin{aligned}\frac{\partial\phi(z)}{\partial z} &= k + \frac{1}{2k}\frac{1}{A}\nabla_\perp^2 A \\ &= k + \frac{1}{2k}\nabla_\perp^2 \ln(A),\end{aligned} \tag{A.4}$$

where the natural logarithm has been introduced by discarding first–order on–axis partial derivatives in the amplitude A due to symmetry of the beam.

A.2 Gouy Phase Evaluated on the Riemann Sphere

We begin by extending the integral for ϕ_B in Eq. (6.25) to the Argand plane. For complex defocus $\tilde{z}$, we can decompose an integration contour Γ in the complex plane as a succession of line segments Γ_{Re} and Γ_{Im}, which are parallel to either the real or the imaginary axis, respectively. Writing $\tilde{z} = z_R + iz_I$, Eq. (6.25) becomes

$$\begin{aligned}-i\int_\Gamma \frac{\partial \ln(\xi)}{\partial\tilde{z}} dz &= -i\left\{\int_{\Gamma_{Re}} \frac{\partial \ln(\xi)}{\partial z_R} dz_R + \int_{\Gamma_{Im}} \frac{\partial \ln(\xi)}{\partial z_I} dz_I\right\} \\ &= \int_\Gamma \left\{\frac{\partial \ln(\xi)}{\partial z_R} - i\frac{\partial \ln(\xi)}{\partial z_I}\right\} d\tilde{z}\end{aligned} \tag{A.5}$$

For closed contour integrals, Green's theorem can be used on both the real and the imaginary parts of Eq. (A.5). For a differentiable complex function $g(z_R, z_I)$, Green's theorem is given by

$$\oint g d\tilde{z} = i \iint \left\{\frac{\partial g}{\partial z_R} + i\frac{\partial g}{\partial z_I}\right\} dz_R dz_I. \tag{A.6}$$

Substitution of Eq. (A.5) into Eq. (A.6) gives

$$\begin{aligned}\iint_S \left[\frac{\partial^2 \ln(\xi)}{\partial z_R^2} + i\frac{\partial^2 \ln(\xi)}{\partial z_R \partial z_I}\right] dz_R dz_I - \iint_S \left[\frac{\partial^2 \ln(\xi)}{\partial z_R \partial z_I} + i\frac{\partial^2 \ln(\xi)}{\partial z_I^2}\right] dz_R dz_I \\ = \iint_S \left[\frac{\partial^2 \ln(\xi)}{\partial z_R^2} + \frac{\partial^2 \ln(\xi)}{\partial z_I^2}\right] dz_R dz_I,\end{aligned} \tag{A.7}$$

so that Eq. (6.25) can be expressed as

$$\phi_B = -i\oint \frac{\partial \ln(\xi)}{\partial\tilde{z}} d\tilde{z} = \iint \nabla^2_{z_R, z_I} \ln(\xi) dz_r dz_I, \tag{A.8}$$

where the two-dimensional Laplacian is defined by the double derivatives with respect to z_R and z_I in Eq. (A.7). To reveal the geometric aspect of Eq. (A.8), consider again our

archetypal Gaussian beam on the optic axis in two transverse dimensions (Boyd, 1980). Evaluating the Laplacian, we obtain:

$$\begin{aligned}\nabla^2_{\tilde{z}} \ln(\xi) &= \nabla^2_{x,y} \ln\left[\left(1+|\tilde{z}|^2\right)^{-\frac{1}{2}} e^{t\tan^{-1}(\tilde{z})}\right] \\ &= -\frac{4}{\left(1+z_R^2+z_I^2\right)^2} dz_r dz_I. \end{aligned} \tag{A.9}$$

Mapping points from the Riemann sphere to the complex plane (Ablowitz and Fokasz, 1997) allows the areal element dA on the unit sphere to be written as

$$dA = \frac{4}{\left(1+z_R^2+z_I^2\right)^2} dz_r dz_I. \tag{A.10}$$

Combining Eqs. (A.8)—(A.10), we have

$$\phi_B = -i \oint \frac{\partial \ln(\xi)}{\partial \tilde{z}} d\tilde{z} = -\frac{1}{2} \oiint dA = -\frac{\Omega}{2}, \tag{A.11}$$

where Ω is the solid angle subtended by the patch on the Riemann sphere.

Appendix B

B

B.1 The Lévy–Leblond Phase Shift and Statistical Confinement

In this section we describe the Gouy phase anomaly by considering the effects of geometric confinement, which supports the interpretation given by Hariharan and Robinson (1996). Far from any obstructions, an electron wave is free and propagates along the z–axis with energy E given by

$$E = \frac{h^2 k^2}{2m}, \tag{B.1}$$

where the wave number $k = 1/\lambda$ and m is the electron mass.

Confinement along the transverse axis within a distance L produces a change in the magnitude of the wave–vector, from k to k' (Lévy-Leblond, 1987) hence

$$E = \frac{h^2 k'^2}{2m} + \frac{h^2 n^2}{8mL^2}, \tag{B.2}$$

where n is an integer greater than zero. Using conservation of energy, we can equate Eqs. (B.1) and (B.2) to obtain

$$E = \frac{h^2 k^2}{2m} = \frac{h^2 k'^2}{2m} + \frac{h^2 n^2}{8mL^2} \rightarrow (k - \Delta k)^2 + \frac{n^2}{4L^2}, \tag{B.3}$$

where $h\Delta k$ is the change in the longitudinal momentum k_z.

Assuming small changes in k, $(\Delta k)^2 << k\Delta k$, we have

$$\Delta k = \frac{n^2}{8kL^2}. \tag{B.4}$$

The integer n^2 must be replaced by $n^2 + m^2$ for waves propagating in two transverse dimensions, which doubles the change in momentum when both n and m are unity. Hence,

the essence of the Lévy-Leblond (1987) phase shift is a geometric effect, which reduces the momentum upon confinement of the wave in the transverse coordinate, orthogonal to the longitudinal propagation direction.

In the absence of boundaries or cavities, focused beams conserve kinetic energy, and we expect no change in the average longitudinal momentum. However, we can describe the Gouy anomaly in terms of the Lévy–Leblond phase shift if we identify the standard deviation in Fig. 6.9 with statistical confinement of the wave, which varies continuously along the z–axis. To examine local phase variations, we must consider an expectation value of the energy or momentum in the vicinity of the optic axis instead of Eq. (B.3). Since the paraxial wave equation is formally equivalent to the time–dependent Schrödinger equation with two spatial variables, we can replace the expectation value of the energy with $\langle \partial/\partial z \rangle$, integrated over an infinitesimal patch of the transverse plane containing the optic axis, which is equivalent to Eq. (A.3). Hence, from the perspective of the Lévy–Leblond effect, focusing produces statistical confinement, which, in turn, induces local momentum fluctuations that give rise to longitudinal phase variations in the optic axis. Naively, this can be inferred from Eq. (B.4) if the confinement parameter L is made proportional to the uncertainty in the beam. Integration along the optic axis then gives an inverse tangent function, which is proportional to the Gouy phase.

Bibliography

Ablowitz M J and Fokasz A S (1997). *Complex Variables, Introduction and Applications*. Cambridge University Press, Cambridge, U.K.

Allen L J, Faulkner H M L, Nugent K A, Oxley M P and Paganin D (2001a). Phase retrieval from images in the presence of first–order vortices. *Physical Review E* **63**(3), 037602.

Allen L J, Faulkner H M L, Oxley M P and Paganin D (2001b). Phase retrieval and aberration correction in the presence of vortices in high-resolution transmission electron microscopy. *Ultramicroscopy* **88**(2), 85–97.

Allen L J, McBride W, O'Leary N L and Oxley M P (2004). Exit wave reconstruction at atomic resolution. *Ultramicroscopy* **100**(1–2), 91–104.

Allen L J and Oxley M P (2001). Phase retrieval from series of images obtained by defocus variation. *Optics Communications* **199**(1–4), 65–75.

Angelsky O V, Maksimyak P P, Maksimyak A P, Hanson S G and Ushenko Y A (2004). Role of caustics in the formation of networks of amplitude zeros for partially developed speckle fields. *Applied Optics* **43**(31), 5744–5753.

Arnol'd V I (1975). Critical points of smooth functions and their normal forms. *Russian Mathematical Surveys* **30**(5), 1–75.

Augustyniak I, Popiołek-Masajada A, Masajada J and Drobczyński S (2012). New scanning technique for the optical vortex microscope. *Applied Optics* **51**(10), C117–C124.

Barnett S M (2008). On the quantum core of an optical vortex. *Journal of Modern Optics* **55**(14), 2279–2292.

Bazhenov V Y, Soskin M S and Vasnetsov M V (1992). Screw dislocations in light wavefronts. *Journal of Modern Optics* **39**(5), 985–990.

Bazhenov V Y, Vasnetsov M V and Soskin M S (1990). Laser beams with screw dislocations in their wavefronts. *JETP Letters* **52**(8), 429–431.

Becker J, Rose P, Boguslawski M and Denz C (2011). Systematic approach to complex periodic vortex and helix lattices. *Optics Express* **19**(10), 9848–9862.

Beijersbergen M W, Coerwinkel R P C, Kristensen M and Woerdman J P (1994). Helical-wavefront laser beams produced with a spiral phaseplate. *Optics Communications* **112**(5–6), 321–327.

Berry M V (1976). Waves and Thom's theorem. *Advances in Physics* **25**, 1–26.

Berry M V (1977). Focusing and twinkling: critical exponents from catastrophes in non-Gaussian random short waves. *Journal of Physics A: Mathematical and General* **10**(12), 2061.

Berry M V (1981). Singularities in waves and rays. *Physics of Defects* **35**, 453–543.

Berry M V (1984). Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **392**(1802), 45–57.

Berry M V (1994a). Evanescent and real waves in quantum billiards and Gaussian beams. *Journal of Physics A: Mathematical and General* **27**(11), L391.

Berry M V (1994b). Faster than Fourier. In J S Anandan and J L Safko, editors, *Quantum coherence and reality; in celebration of the 60th birthday of Yakir Aharonov*, chapter 2, pages 55–65. World Scientific, Singapore.

Berry M V (1998). Much ado about nothing: optical distortion lines (phase singularities, zeros, and vortices). *SPIE Proceedings* **3487**, 1–5.

Berry M V (2000). Making waves in physics. *Nature* **403**(6765), 21–21.

Berry M V (2009). Optical currents. *Journal of Optics A: Pure and Applied Optics* **11**(9), 094001.

Berry M V and Balazs N L (1979). Nonspreading wave packets. *American Journal of Physics* **47**(3), 264–267.

Berry M V and Dennis M R (2004). Quantum cores of optical phase singularities. *Journal of Optics A: Pure and Applied Optics* **6**(5), S178.

Berry M V and Howls C (2006). Chap. 36: Integrals with coalescing saddles. In R W J Oliver, D W Lozier, R F Boisvert and C W Clark, editors, *NIST Handbook of Mathematical Functions*. Cambridge University Press, Cambridge.

Berry M V and Klein S (1996). Colored diffraction catastrophes. *Proceedings of the National Academy of Sciences* **93**(6), 2614–2619.

Berry M V, Nye J F and Wright F J (1979). The elliptic umbilic diffraction catastrophe. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **291**(1382), pp. 453–484.

Berry M V and Popescu S (2006). Evolution of quantum superoscillations and optical superresolution without evanescent waves. *Journal of Physics A: Mathematical and General* **39**(22), 6965.

Berry M V and Upstill C (1980). Catastrophe optics: morphologies of caustics and their diffraction patterns. *Progress in Optics* **18**, 257–346.

Bliokh K Y, Bliokh Y P, Savel'ev S and Nori F (2007). Semiclassical Dynamics of Electron Wave Packet States with Phase Vortices. *Physical Review Letters* **99**, 190404.

Bohren C F and Huffman D R (1983). *Absorption and scattering of light by small particles*. John Wiley & Sons, Inc., New York.

Bondi H and Pirani F A E (1989). Gravitational waves in general relativity. XIII. Caustic property of plane waves. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **421**(1861), 395–410.

Bonse U and Hart M (1965). An X-ray interferometer. *Applied Physics Letters* **6**(8), 155–156.

Borghi R, Santarsiero M and Simon R (2004). Shape invariance and a universal form for the Gouy phase. *Journal of the Optical Society of America A* **21**(4), 572–579.

Born M and Wolf E (1999). *Principles of optics*. Cambridge University Press, 7th edition.

Borwińska M, Popiołek-Masajada A and Kurzynowski P (2007). Measurements of birefringent media properties using optical vortex birefringence compensator. *Applied Optics* **46**(25), 6419–6426.

Boyd R W (1980). Intuitive explanation of the phase anomaly of focused light beams. *Journal of the Optical Society of America* **70**(7), 877–880.

Braunbek W (1951). Zur Darstellung von Wellenfeldern. *Zeitschrift Naturforschung Teil A* **6**, 12.

Brosseau C (1998). *Fundamentals of polarized light: a statistical optics approach*. Wiley, New York.

Brunet T, Thomas J L and Marchiano R (2010). Transverse shift of helical beams and

subdiffraction imaging. *Physical Review Letters* **105**, 034301.

Canny J (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(6), 679–698.

Chen Y F, Liang H C, Lin Y C, Tzeng Y S, Su K W and Huang K F (2011). Generation of optical crystals and quasicrystal beams: Kaleidoscopic patterns and phase singularity. *Physical Review A* **83**, 053813.

Chu Y H, Martin L W, Holcomb M B, Gajek M, Han S J, He Q, Balke N, Yang C H, Lee D, Hu W, Zhan Q, Yang P L, Fraile-Rodriguez A, Scholl A, Wang S X and Ramesh R (2008). Electric-field control of local ferromagnetism using a magnetoelectric multiferroic. *Nature Materials* **7**(6), 478–482.

Clark L, Béché A, Guzzinati G, Lubk A, Mazilu M, Van Boxem R and Verbeeck J (2013). Exploiting lens aberrations to create electron–vortex beams. *Physical Review Letters* **111**(6), 064801.

Connor J N L (1976). Catastrophes and molecular collisions. *Molecular Physics* **31**(1), 33–55.

Connor J N L and Farrelly D (1981a). Molecular collisions and cusp catastrophes: three methods for the calculation of pearcey's integral and its derivatives. *Chemical Physics Letters* **81**(2), 306–310.

Connor J N L and Farrelly D (1981b). Theory of cusped rainbows in elastic scattering: Uniform semiclassical calculations using Pearcey's integral. *The Journal of Chemical Physics* **75**(6), 2831–2846.

Courtial J, Dholakia K, Allen L and Padgett M J (1997). Gaussian beams with very high orbital angular momentum. *Optics Communications* **144**(4–6), 210–213.

Cowley J M (1995). *Diffraction physics*. Elsevier, Amsterdam, 2nd edition.

Crow S C (1970). Stability theory for a pair of trailing vortices. *AIAA journal* **8**(12), 2172–2179.

da Paz I G, Saldanha P L, Nemes M C and de Faria J G P (2011). Experimental proposal for measuring the Gouy phase of matter waves. *New Journal of Physics* **13**(12), 125005.

De Graef M (2003). *Introduction to conventional transmission electron microscopy*. Cambridge Univeristy Press, Cambridge.

Dennis M R (2003). Braided nodal lines in wave superpositions. *New Journal of Physics* **5**(1), 134.

Dennis M R and Götte J B (2012). Topological aberration of optical vortex beams: Determining dielectric interfaces by optical singularity shifts. *Physical Review Letters* **109**, 183903.

Dennis M R and Götte J B (2013). Beam shifts for pairs of plane waves. *Journal of Optics* **15**(1), 014015.

Dennis M R, Hamilton A C and Courtial J (2008). Superoscillation in speckle patterns. *Optics Letters* **33**(24), 2976–2978.

Dennis M R, King R P, Jack B, O'Holleran K and Padgett M J (2010). Isolated optical vortex knots. *Nature Physics* **6**, 118–121.

Dirac P A (1931). Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* pages 60–72.

Dreischuh A, Chervenkov S, Neshev D, Paulus G G and Walther H (2002). Generation of lattice structures of optical vortices. *Journal of the optical society of America B* **19**(3), 550–556.

Dwyer C, Boothroyd C B, Chang S L Y and Dunin-Borkowski R E (2015). Three–wave electron vortex lattices for measuring nanofields. *Ultramicroscopy* **148**, 25–30.

Eastwood S A, Bishop A I, Petersen T C, Paganin D M and Morgan M J (2012). Phase measurement using an optical vortex lattice produced with a three-beam interferometer. *Optics Express* **20**(13), 13947–13957.

Eastwood S A, Paganin D M and Liu A C Y (2011). Automated phase retrieval of a single-material object using a single out–of–focus image. *Optics Letters* **36**(10), 1878–1880.

Eastwood S A, Paganin D M, Petersen T C and Morgan M J (2014). Bloch sphere catastrophes. arXiv:1406.0583 [physics.optics].

Elliot J L, French R G, Dunham E, Gierasch P J, Veverka J, Church C and Sagan C (1977). Occultation of e Geminorum by Mars. II. The structure and extinction of the Martian upper atmosphere. *The Astrophysics Journal* **217**, 661–679.

Ezawa M (2010). Giant Skyrmions stabilized by dipole–dipole interactions in thin ferromegnetic films. *Physical Review Letters* **105**, 197202.

Felsen L B (1976). Evanescent waves. *Journal of the Optical Society of America* **66**(8), 751–760.

Feng S and Winful H G (2001). Physical origin of the Gouy phase shift. *Optics Letters* **26**(8), 485–487.

Frączek E, Frączek W and Masajada J (2006). The new method of topological charge determination of optical vortices in the interference field of the optical vortex interferometer. *Optik–International Journal for Light and Electron Optics* **117**(9), 423–425.

Frączek E, Frączek W, Mroczka J et al. (2005). Experimental method for topological charge determination of optical vortices in a regular net. *Optical Engineering* **44**(2), 025601–025601.

Frączek W and Mroczka J (2008). Optical vortices as phase markers to wave-front deformation measurement. *Metrology and Measurement Systems* **15**(4), 433–440.

Freedman B, Bartal G, Segev M, Lifshitz R, Christodoulides D N and Fleischer J W (2006). Wave and defect dynamics in nonlinear photonic quasicrystals. *Nature* **440**(7088), 1166–1169.

Freedman B, Lifshitz R, Fleischer J W and Segev M (2007). Phason dynamics in nonlinear photonic quasicrystals. *Nature Materials* **6**(10), 776–781.

Freund I (1999). Critical point explosions in two-dimensional wave fields. *Optics Communications* **159**(1–3), 99–117.

Freund I and Shvartsman N (1994). Wave–field phase singularities: The sign principle. *Physical Review A* **50**, 5164–5172.

Gbur G and Visser T D (2003). Coherence vortices in partially coherent beams. *Optics Communications* **222**(1–6), 117–125.

Gerchberg R W and Saxton W O (1972). A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **35**, 237.

Golovach V N, Borhani M and Loss D (2010). Holonomic quantum computation with electron spins in quantum dots. *Physical Review A* **81**, 022315.

Gouy L G (1891). Sur la propagation anomale des ondes. *Anneles des Chimie et de Physique* **24**, 145–213.

Guidoni L, Triché C, Verkerk P and Grynberg G (1997). Quasiperiodic optical lattices. *Physical Review Letters* **79**, 3363–3366.

Gull E, Parcollet O and Millis A J (2013). Superconductivity and the pseudogap in the two–dimensional Hubbard model. *Physical Review Letters* **110**, 216405.

Gureyev T E, Roberts A and Nugent K A (1995). Phase retrieval with the transport-of-intensity equation: matrix solution with use of Zernike polynomials. *Journal of the optical society of America A* **12**(9), 1932–1941.

Hannay J H (1982). Intensity fluctuations beyond a one-dimensional random refracting screen in the short–wavelength limit. *Optica Acta* **29**(12), 1631–1649.

Harada K, Akashi T, Togawa Y, Matsuda T and Tonomura A (2005). Variable Interference Azimuth Angle in Double-Biprism Electron Interferometry. *Japanese Journal of Applied Physics* **44**(5L), L636.

Harada K, Tonomura A, Togawa Y, Akashi T and Matsuda T (2004). Double-biprism electron interferometry. *Applied Physics Letters* **84**(17), 3229–3231.

Hariharan P and Robinson P A (1996). The Gouy phase shift as a geometrical quantum effect. *Journal of Modern Optics* **43**(2), 219–221.

Hawkes P W and Kasper E (1996). *Principles of electron optics*, volume 3. Academic Press.

He H, Heckenberg N and Rubinsztein-Dunlop H (1995). Optical particle trapping with higher–order doughnut beams produced using high efficiency computer generated holograms. *Journal of Modern Optics* **42**(1), 217–223.

Heckenberg N R, McDuff R, Smith C P and White A G (1992). Generation of optical phase singularities by computer-generated holograms. *Optics Letters* **17**(3), 221–223.

Heflinger L, Wuerker R and Brooks R E (1966). Holographic interferometry. *Journal of Applied Physics* **37**(2), 642–649.

Hemo E, Spektor B and Shamir J (2011). Scattering of singular beams by subwavelength objects. *Applied Optics* **50**(1), 33–42.

Hirayama T, Tanji T and Tonomura A (1995). Direct visualization of electromagnetic microfields by interference of three electron waves. *Applied Physics Letters* **67**(9), 1185–1187.

Hirschfelder J O, Goebel C J and Bruch L W (1974). Quantized vortices around wavefunction nodes. II. *The Journal of Chemical Physics* **61**(12), 5456–5459.

Horn T, Slooten U V and Kleyn A (1989). Observation of elementary catastrophes in surface scattering. *Chemical Physics Letters* **156**(6), 623–627.

Jakeman E, Pike E R and Pusey P N (1976). Photon correlation study of stellar scintillation. *Nature* **263**, 215–217.

Kandpal H C, Raman S and Mehrotra R (2007). Observation of Gouy phase anomaly with an interferometer. *Optics and Lasers in Engineering* **45**(1), 249–251.

Kawaguchi Y, Nitta M and Ueda M (2008). Knots in a spinor Bose-Einstein condensate. *Physical Review Letters* **100**, 180403.

Keller J (1985). Semiclassical mechanics. *SIAM Review* **27**(4), 485–504.

Keller J B (1958). Corrected Bohr–Sommerfeld quantum conditions for nonseparable systems. *Annals of Physics* **4**(2), 180–188.

Kimura T, Kawamoto S, Yamada I, Azuma M, Takano M and Tokura Y (2003). Magnetocapacitance effect in multiferroic $BiMnO_3$. *Physical Review B* **67**, 180401.

Kingslake R (1925). The interferometer patterns due to the primary aberrations. *Transactions of the Optical Society* **27**(2), 94.

Kitchen M J, Paganin D, Lewis R A, Yagi N, Uesugi K and Mudie S T (2004). On the origin of speckle in x-ray phase contrast images of lung tissue. *Physics in Medicine and Biology* **49**(18), 4335.

Kurzynowski P and Borwińska M (2007). Generation of vortex–type markers in a one-wave setup. *Applied Optics* **46**(5), 676–679.

Kurzynowski P, Woźniak W A and Frączek E (2006). Optical vortices generation using the Wollaston prism. *Applied Optics* **45**(30), 7898–7903.

Leach J, Dennis M R, Courtial J and Padgett M J (2004). Laser beams: Knotted threads of darkness. *Nature* **432**(165).

Leach J, Dennis M R, Courtial J and Padgett M J (2005). Vortex knots in light. *New Journal of Physics* **7**(1), 55.

Lee R A (1983a). Generalized curvilinear diffraction gratings: II. Generalized ray equations and caustics. *Optica Acta: International Journal of Optics* **30**(3), 291–303.

Lee R A (1983b). Generalized curvilinear diffraction gratings: V. Diffraction catastrophes. *Optica Acta: International Journal of Optics* **30**(4), 449–464.

Levitin L V, Bennett R G, Casey A, Cowan B, Saunders J, Drung D, Schurig T and Parpia J M (2013). Phase diagram of the topological superfluid ^{3}He confined in a nanoscale slab geometry. *Science* **340**(6134), 841–844.

Lévy-Leblond J M (1987). A geometrical quantum phase effect. *Physics Letters A* **125**(9), 441–442.

Lindeberg T (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision* **30**(2), 117–156.

Lu Y C (1976). *Singularity theory and an introduction to catastrophe theory*. Springer–Verlag, New York.

Maeda K, Tachizawa T, Torii T and Maki T (1994). Stability of non-Abelian black holes and catastrophe theory. *Physical Review Letters* **72**, 450–453.

Mäkelä H, Zhang Y and Suominen K A (2003). Topological defects in spinor condensates. *Journal of Physics A* **36**(32), 8555.

Marasinghe M L, Paganin D M and Premaratne M (2011). Coherence-vortex lattice formed via Mie scattering of partially coherent light by several dielectric nanospheres. *Optics Letters* **36**(6), 936–938.

Marasinghe M L, Premaratne M and Paganin D M (2010). Coherence vortices in Mie scattering ofstatistically stationary partially coherentfields. *Optics Express* **18**(7), 6628–6641.

Mari E, Tamburini F, Swartzlander G A Jr, Bianchini A, Barbieri C, Romanato F and Thidé B (2012). Sub-Rayleigh optical vortex coronagraphy. *Optics Express* **20**(3), 2445–2451.

Marston P L and Trinh E H (1984). Hyperbolic umbilic diffraction catastrophe and rainbow scattering from spheroidal drops. *Nature* **312**(5994), 529–531.

Martinis J M, Nam S, Aumentado J, Lang K M and Urbina C (2003). Decoherence of a superconducting qubit due to bias noise. *Physical Review B* **67**, 094510.

Masajada J (2004). Small–angle rotations measurement using optical vortex interferometer. *Optics Communications* **239**(4–6), 373–381.

Masajada J (2005). The internal scanning method with optical vortex interferometer. In *Congress on Optics and Optoelectronics*, volume 59581. International Society for Optics and Photonics.

Masajada J and Dubik B (2001). Optical vortex generation by three plane wave interference. *Optics Communications* **198**(1–3), 21–27.

Masajada J, Leniec M and Augustyniak I (2011a). Optical vortex scanning inside the Gaussian beam. *Journal of Optics* **13**(3), 035714.

Masajada J, Leniec M and Augustyniak I (2011b). Optical vortex scanning inside the Gaussian beam. *Journal of Optics* **13**(3), 035714.

Masajada J, Leniec M, Drobczyński S, Thienpont H and Kress B (2009). Micro-step localization using double charge optical vortex interferometer. *Optics Express* **17**(18), 16144–16159.

Masajada J and Popiołek-Masajada A (2005). Phase reconstruction with optical vortex interferometer. In A Štrba, D Senderákoá and M Hrabovský, editors, *Proceedings of SPIE*, volume 5945, pages 1–8.

Masajada J, Popiołek-Masajada A and Leniec M (2007). Creation of vortex lattices by a wavefront division. *Optics Express* **15**(8), 5196–5207.

Masajada J, Popiołek-Masajada A and Wieliczka D M (2002). The interferometric system using optical vortices as phase markers. *Optics Communications* **207**(1–6), 85–93.

Mercier R P (1962). Diffraction by a screen causing large random phase fluctuations. *Mathematical Proceedings of the Cambridge Philosophical Society* **58**, 382–400.

Mermin N D (1979). The topological theroy of defects in ordered madia. *Reviews of Modern Physics* **51**, 591–648.

Metzger J J, Fleischmann R and Geisel T (2014). Statistics of extreme waves in random media. *Physical Review Letters* **112**, 203903.

Milnor J W (1963). *Morse theory*. Princeton University Press, New Jersey.

Misell D L (1973). An examination of an iterative method for the solution of the phase problem in optics and electron optics: II. Sources of error. *Journal of Physics D: Applied Physics* **6**(18), 2217.

Moodie A F, Etheridge J and Humphreys C J (1996). The symmetry of three–beam scattering equations: Inversion of three-beam diffraction patterns from centrosymmetric crystals. *Acta Crystallographica Section A* **52**(4), 596–605.

Mühlbauer S, Binz B, Jonietz F, Pfleiderer C, Rosch A, Neubauer A, Georgii R and Böni P (2009). Skyrmion lattice in a chiral magnet. *Science* **323**(5916), 915–919.

Nakashima P N H, Moodie A F and Etheridge J (2007). Structural phase and amplitude measurement from distances in convergent-beam electron diffraction patterns. *Acta Crystallographica Section A* **63**(5), 387–390.

Nicholls K W and Nye J F (1987). Three-beam model for studying dislocations in wave

pulses. *Journal of Physics A: Mathematical and General* **20**(14), 4673.

Nielsen M A and Chuang I L (2010). *Quantum computation and quantum information*. Cambridge University Press, Cambridge.

Nienhuis K and Nijboer B R A (1949). The diffraction theory of optical aberrations: Part III: General formulae for small aberrations; experimental verification of the theoretical results. *Physica* **14**(9), 590–608.

Niermann T, Verbeeck J and Lehmann M (2014). Creating arrays of electron vortices. *Ultramicroscopy* **136**, 165–170.

Nijboer B R A (1947). The diffraction theory of optical aberrations: Part II: Diffraction pattern in the presence of small aberrations. *Physica* **13**(10), 605–620.

Nye J F (1978). Optical caustics in the near field from liquid drops. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **361**(1704), 21–41.

Nye J F (1984). Rainbow scattering from spheroidal drops—an explanation of the hyperbolic umbilic foci. *Nature* **312**(5994), 531–532.

Nye J F (1999). *Natural focusing and fine structure of light: Caustics and wave dislocations*. Institute of Physics Publishing, Bristol.

Nye J F (2003a). Evolution from a Fraunhofer to a Pearcey diffraction pattern. *Journal of Optics A–Pure and Applied Optics* **5**, 495.

Nye J F (2003b). From Airy rings to the elliptic umbilic diffraction catastrophe. *Journal of Optics A–Pure and Applied Optics* **5**, 503.

Nye J F (2005). The relation between the spherical aberration of a lens and the spun cusp diffraction catastrophe. *Journal of Optics A: Pure and Applied Optics* **7**(3), 95.

Nye J F (2006). Evolution of the hyperbolic umbilic diffraction pattern from Airy rings. *Journal of Optics A: Pure and Applied Optics* **8**, 304.

Nye J F and Berry M V (1974). Dislocations in wave trains. *Proceedings of the royal society A* **336**, 165–190.

Nye J F and Hajnal J V (1987). The Wave Structure of Monochromatic Electromagnetic Radiation. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **409**(1836), 21–36.

Ohanian H C (1983). The caustics of gravitational 'lenses'. *The Astrophysical Journal* **271**, 551–555.

O'Holleran K, Dennis M R, Flossmann F and Padgett M J (2008). Fractality of light's darkness. *Physical Review Letters* **100**, 053902.

O'Holleran K, Dennis M R and Padgett M J (2009a). Topology of light's darkness. *Physical Review Letters* **102**, 143902.

O'Holleran K, Flossmann F, Dennis M R and Padgett M J (2009b). Methodology for imaging the 3D structure of singularities in scalar and vector optical fields. *Journal of Optics A: Pure and Applied Optics* **11**(9), 094020.

O'Holleran K, Padgett M J and Dennis M R (2006). Topology of optical vortex lines formed by the interference of three, four, and five plane waves. *Optics Express* **14**(7), 3039–3044.

Orlov Y I (1981). Caustics with anomalous phase shifts. *Radiophysics and Quantum Electronics* **24**(2), 15–159.

Padgett M (2008). On the focussing of light, as limited by the uncertainty principle. *Journal of Modern Optics* **55**(18), 3083–3089.

Padgett M J, O'Holleran K, King R P and Dennis M R (2011). Knotted and tangled threads of darkness in light beams. *Contemporary Physics* **52**(4), 265–279.

Paganin D M (2006). *Coherent X-Ray optics*. Clarendon Press, Oxford.

Pang X, Gbur G and Visser T D (2011). The Gouy phase of Airy beams. *Optics Letters* **36**(13), 2492–2494.

Pavlov K M, Paganin D M, Vine D J, Schmalz J A, Suzuki Y, Uesugi K, Takeuchi A, Yagi N, Kharchenko A, Blaj G, Jakubek J, Altissimo M and Clark J N (2011). Quantized hard-x-ray phase vortices nucleated by aberrated nanolenses. *Physical Review A* **83**, 013813.

Pearcey T (1946). XXXI. The structure of an electromagnetic field in the neighbourhood of a cusp of a caustic. *Philosophical Magazine* **37**(268), 311–317.

Peele A G, McMahon P J, Paterson D, Tran C Q, Mancuso A P, Nugent K A, Hayes J P, Harvey E, Lai B and McNulty I (2002). Observation of an x-ray vortex. *Optics Letters* **27**(20), 1752–1754.

Peregrine D H and Smith R (1979). Nonlinear effects upon waves near caustics. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **292**(1392), 341–370.

Petersen T C (2014). Private Communication.

Petersen T C, Paganin D M, Weyland M, Simula T P, Eastwood S A and Morgan M J (2013a). Measurement of the Gouy phase anomaly for electron waves. *Physical Review A* **88**, 043803.

Petersen T C, Paganin D M, Weyland M, Simula T P, Eastwood S A and Morgan M J (2014). Unifying interpretations of the Gouy phase anomaly for electron waves. *Physical Review A* **89**, 063801.

Petersen T C, Weyland M, Paganin D M, Simula T P, Eastwood S A and Morgan M J (2013b). Electron vortex production and control using aberration induced diffraction catastrophes. *Physical Review Letters* **110**, 033901.

Pieranski P, Yang B, Burtz L J, Camu A and Simonetti F (2013). Generation of umbilics by magnets and flows. *Liquid Crystals* **40**(12), 1593–1608.

Popiołek-Masajada A, Borwińska M and Dubik B (2007). Reconstruction of a plane wave's tilt and orientation using an optical vortex interferometer. *Optical Engineering* **46**(7), 073604–073604–8.

Popiołek-Masajada A and Frączek W (2011). Evaluation of a phase shifting method for vortex localization in optical vortex interferometry. *Optics & Laser Technology* **43**(7), 1219–1224.

Poston T and Stewart I (1996). *Catastrophe theory and its applications*. Dover publications inc., Mineola, New York.

Poston T and Stewart I N (1976). *Taylor expansions and catastrophes*. Pitman Publishing, London.

Ravikumar V, Rodrigues R P and Dravid V P (1997). Space–charge distribution across internal interfaces in electroceramics using electron holography: I, Pristine grain boundaries. *Journal of the American Ceramic Society* **80**(5), 1117–1130.

Rolland J P, Thompson K P, Lee K S, Tamkin J, Schmid T and Wolf E (2012). Observation of the Gouy phase anomaly in astigmatic beams. *Applied Optics* **51**(15), 2902–2908.

Röszler U K, Bogdanov A N and Pflederer C (2006). Spontaneous skyrmion ground states in magnetic metals. *Nature* **442**, 797–801.

Rothschild F, Bishop A I, Kitchen M J and Paganin D M (2014). Argand-plane vorticity singularities in complex scalar optical fields: An experimental study using optical speckle. *Optics Express* **22**(6), 6495–6510.

Rothschild F, Kitchen M J, Faulkner H M L and Paganin D M (2012). Duality between phase vortices and Argand-plane caustics. *Optics Communications* **285**(21–22), 4141–4151.

Roux F S (2003). Optical vortex density limitation. *Optics Communications* **223**(1–3), 31–37.

Rozas D, Law C T and G A Swartzlander J (1997). Propagation dynamics of optical vortices. *Journal of the Optical Society of America B* **14**(11), 3054–3065.

Ruben G, Morgan M J and Paganin D M (2010). Texture control in a pseudospin Bose-Einstein condensate. *Physical Review Letters* **105**, 220402.

Ruben G and Paganin D M (2007). Phase vortices from a Young's three-pinhole interferometer. *Physical Review E* **75**, 066613.

Sakdinawat A and Liu Y (2007). Soft-x-ray microscopy using spiral zone plates. *Optics Letters* **32**(18), 2635–2637.

Saunders P T (1980). *An introduction to catastrophe theory*. Cambridge University Press, Cambridge.

Schouten H F, Gbur G, Visser T D and Wolf E (2003). Phase singularities of the coherence functions in Young's interference pattern. *Optics Letters* **28**(12), 968–970.

Seibold G, Benfatto L, Castellani C and Lorenzana J (2012). Superfluid density and phase relaxation in superconductors with strong disorder. *Physical Review Letters* **108**, 207004.

Sethna J (2006). *Statistical mechanics: Entropy, order parameters and complexity*. Oxford University Press, Oxford.

Shechtman D, Blech I, Gratias D and Cahn J W (1984). Metallic phase with long-range orientational order and no translational symmetry. *Physical Review Letters* **53**, 1951–1953.

Shiloh R, Lereah Y, Lilach Y and Arie A (2014). Sculpturing the electron wave function using nanoscale phase masks. *Ultramicroscopy* **144**, 26–31.

Siegman A E (1986). *Lasers*. University Science Books, Sausalito, CA.

Siersma D (1974). *Classification and deformation of singularities*. Ph.D. thesis, University of Amsterdam.

Simon R and Mukunda N (1993). Bargmann invariant and the geometry of the Güoy effect. *Physical Review Letters* **70**, 880–883.

Simula T P and Paganin D M (2011). Coherence vortices in one spatial dimension. *Physical Review A* **84**, 052104.

Simula T P, Petersen T C and Paganin D M (2013). Diffraction catastrophes threaded by quantized vortex skeletons caused by atom-optical aberrations induced in trapped Bose-Einstein condensates. *Physical Review A* **88**, 043626.

Sobel I (1990). An Isotropic 3 x 3 image gradient operator. In H Freeman, editor, *Machine Vision for Three–Dimensional Scenes*. Academic Press, San Diego.

Solli D R, Ropers C, Koonath P and Jalali B (2007). Optical rogue waves. *Nature* **450**(7172), 1054–1057.

Sommerfeld A J W (1954). *Optics*. Academic Press.

Spektor B, Normatov A and Shamir J (2008). Singular beam microscopy. *Applied Optics* **47**(4), A78–A87.

Spektor B, Normatov A and Shamir J (2010). Singular beam scanning microscopy: preliminary experimental results. *Optical Engineering* **49**(4), 048001–048001–6.

Stewart I (1982). Catastrophe theory in physics. *Reports on Progress in Physics* **45**(2), 185.

Subbarao D (1995). Topological phase in Gaussian beam optics. *Optics Letters* **20**(21), 2162–2164.

Swartzlander G A (2001). Peering into darkness with a vortex spatial filter. *Optics Letters* **26**(8), 497–499.

Swartzlander G A Jr (2009). The optical vortex coronagraph. *Journal of Optics A: Pure and Applied Optics* **11**(9), 094022.

Szatkowski M, Popiołek-Masajada A and Masajada J (2014). Optical vortex in microscopy imaging. In A Forbes and T E Lizotte, editors, *Proc. SPIE 9194, Laser Beam Shaping XV*, volume 91941D, pages 1–6. International Society for Optics and Photonics.

Takeda M, Ina H and Kobayashi S (1982). Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *Journal of the Optical Society of America* **72**(1), 156–160.

Teague M R (1983). Deterministic phase retrieval: a Green's function solution. *Journal of the Optical Society of America* **73**(11), 1434–1441.

Terauchi M, Tsuda K, Kamimura O, Tanaka M, Kaneyama T and Honda T (1994). Observation of lattice fringes in convergent-beam electron diffraction patterns. *Ultramicroscopy* **54**(2–4), 268–275.

Thom R (1975). *Structural stability and morphogenesis*. Benjamim, Reading, Massachusetts.

Thom R (1977). Structural stability, catastrophe theory, and applied mathematics. *SIAM Review* **19**(2), 189–201.

Thom R (1983). *Mathematical models of morphogenesis*. Wiley, New York.

Tonomura A (1999). *Electron holography*. Springer, Berlin.

Trinkaus H and Drepper F (1977). On the analysis of diffraction catastrophes. *Journal of Physics A: Mathematical and General* **10**(1), L11.

Tyc T (2012). Gouy phase for full-aperture spherical and cylindrical waves. *Optics Letters* **37**(5), 924–926.

Tychinsky V P (2008). Superresolution and singularities in phase images. *Physics–Uspekhi* **51**(11), 1161.

Uchida M and Tonomura A (2010). Generation of electron beams carrying orbital angular momentum. *Nature* **464**(7289), 737–739.

Uzer T and Child M S (1982). Collisions and umbilic catastrophes: Direct determination of the control parameters for use in a uniform S matrix approximation. *Molecular Physics* **46**(6), 1371–1388.

Vaughan J and Willetts D (1979). Interference properties of a light beam having a helical wave surface. *Optics Communications* **30**(3), 263–267.

Verbeeck J, Tian H and Béché A (2012). A new way of producing electron vortex probes for STEM. *Ultramicroscopy* **113**, 83–87.

Vincent R, Vine W J, Midgley P A, Spellward P and Steeds J W (1993). Coherent overlapping LACBED patterns in 6H SiC. *Ultramicroscopy* **50**(3), 365–376.

Visser T D and Wolf E (2010). The origin of the Gouy phase anomaly and its generalization to astigmatic wavefields. *Optics Communications* **283**(18), 3371–3375.

Volkov V V and Zhu Y (2003). Deterministic phase unwrapping in the presence of noise. *Optics Letters* **28**, 2156–2158.

Vyas S and Senthilkumaran P (2007). Interferometric optical vortex array generator. *Applied Optics* **46**(15), 2893–2898.

Walther A (1995). *The ray and wave theory of lenses*. Cambridge University Press, Cambridge.

Wang J et al. (2003). Epitaxial BiFeO3 multiferroic thin film heterostructures. *Science* **299**(5613), 1719–1722.

Whewell W (1833). Essay towards a first approximation to a map of cotidal lines. *Philosophical Transactions of the Royal Society of London* **123**, 147–236.

Whewell W (1836). Researches on the tides. Sixth series. On the results of an extensive system of tide observations made on the coasts of Europe and America in June 1835. *Philosophical Transactions of the Royal Society of London* pages 289–341.

Winthrop J T and Worthington C R (1966). Convolution Formulation of Fresnel Diffraction. *Journal of the Optical Society of America* **56**(5), 588–591.

Woźniak W A and Banach M (2009). Measurements of linearly birefringent media parameters using the optical vortex interferometer with the Wollaston compensator. *Journal of Optics A: Pure and Applied Optics* **11**(9), 094024.

Wolf E (1951). The diffraction theory of aberrations. *Reports on Progress in Physics* **14**(1), 95.

Wolf E and Foley J T (1998). Do evanescent waves contribute to the far field? *Optics Letters* **23**(1), 16–18.

Wolter H (1950). Zur frage des lichtweges bei totalreflexion. *Zeitschrift Naturforschung Teil A* **5**, 276.

Wolter H (2009). Concerning the path of light upon total reflection. *Journal of Optics A: Pure and Applied Optics* **11**(9), 090401.

Xavier J, Vyas S, Senthilkumaran P, Denz C and Joseph J (2011). Sculptured 3D twister superlattices embedded with tunable vortex spirals. *Optics Letters* **36**(17), 3512–3514.

Xavier J, Vyas S, Senthilkumaran P and Joseph J (2012). Tailored complex 3D vortex lattice structures by perturbed multiples of three-plane waves. *Applied Optics* **51**(12), 1872–1878.

Xiang G Y, Li J, Yu B and Guo G C (2005). Remote preparation of mixed states via noisy entanglement. *Physical Review A* **72**, 012315.

Yang J and Winful H G (2006). Generalized eikonal treatment of the Gouy phase shift. *Optics Letters* **31**(1), 104–106.

Zeeman E C (1977). *Catastrophe theory: Selected papers, 1972–1977.* Addison-Wesley, Reading, Massachusetts.

Zernike F (1942). Phase contrast, a new method for the microscopic observation of transparent objects. *Physica* **9**(7), 686–698.

Zhan M, Li K, Wang P, Kong L, Wang X, Li R, Tu X, He L, Wang J and Lu B (2007). Cold atom interferometry. In *Journal of Physics: Conference Series*, volume 80, page 012047. IOP Publishing.

Zhang Y, Ran Y and Vishwanath A (2009). Topological insulators in three dimensions from spontaneous symmetry breaking. *Physical Review B* **79**, 245331.

Publications

C

One first–author and three co–authored peer reviewed papers have been published as a result of work reported in this thesis. A second first–author paper is currently under review. These publications are included here.

Phase measurement using an optical vortex lattice produced with a three-beam interferometer

Samuel A. Eastwood,* Alexis I. Bishop, Timothy C. Petersen, David M. Paganin, and Michael J. Morgan

School of Physics, Monash University, Clayton, Victoria 3800, Australia

Abstract: A new phase-measurement technique is proposed, which utilizes a three-beam interferometer. Three-wave interference in the interferometer generates a uniform lattice of optical vortices, which is distorted by the presence of an object inserted in one arm of the interferometer. The transverse displacement of the vortices is proportional to the phase shift in the object wave. Tracking the vortices permits the phase of the object to be reconstructed. We demonstrate the method experimentally using a simple lens and a more complex object, namely the wing of a common house fly. Since the technique is implemented in real space, it is capable of reconstructing the phase locally.

OCIS codes: (050.4865) Optical vortices; (100.5088) Phase retrieval; (120.3180) Interferometry; (260.6042) Singular optics.

References and links

1. L. O. Heflinger, R. F. Wuerker, and R. E. Brooks, "Holographic interferometry," J. Appl. Phys. **37**(2), 642–649 (1966).
2. M. Zhan, K. Li, P. Wang, L. Kong, X. Wang, R. Li, X. Tu, L. He, J. Wang, and B. Lu, "Cold atom interferometry," J. Phys.: Conf. Ser. **80**, 012047 (2007).
3. A. Tonomura, T. Matsuda, J. Endo, T. Arii, and K. Mihama, "Holographic interference electron microscopy for determining specimen magnetic structure and thickness distribution," Phys. Rev. B Condens. Matter **34**(5), 3397–3402 (1986).
4. U. Bonse and M. Hart, "An X-ray interferometer," Appl. Phys. Lett. **6**(8), 155–156 (1965).
5. M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," J. Opt. Soc. Am. **72**(1), 156–160 (1982).
6. P. Senthilkumaran, J. Masajada, and S. Sato, "Interferometry with vortices," Int. J. Opt. **2012**, 517591 (2012).
7. A. Popiołek-Masajada, M. Borwinska, and B. Dubik, "Reconstruction of a plane wave's tilt and orientation using an optical vortex interferometer," Opt. Eng. **46**(7), 073604 (2007).
8. J. Masajada, "Small-angle rotations measurement using optical vortex interferometer," Opt. Commun. **239**(4-6), 373–381 (2004).
9. M. Borwińska, A. Popiołek-Masajada, and P. Kurzynowski, "Measurements of birefringent media properties using optical vortex birefringence compensator," Appl. Opt. **46**(25), 6419–6426 (2007).
10. W. A. Woźniak and M. Banach, "Measurements of linearly birefringent media parameters using the optical vortex interferometer with the Wollaston compensator," J. Opt. A, Pure Appl. Opt. **11**(9), 094024 (2009).
11. W. Fraczek and J. Mroczka, "Optical vortices as phase markers to wave-front deformation measurement," Metrol. Meas. Syst. **15**, 433–440 (2008).
12. D. M. Paganin, *Coherent X-Ray Optics* (Clarendon Press, 2006).
13. K. W. Nicholls and J. F. Nye, "Three-beam model for studying dislocations in wave pulses," J. Phys. Math. Gen. **20**(14), 4673–4696 (1987).
14. J. Masajada and B. Dubik, "Optical vortex generation by three plane wave interference," Opt. Commun. **198**(1-3), 21–27 (2001).
15. K. O'Holleran, M. J. Padgett, and M. R. Dennis, "Topology of optical vortex lines formed by the interference of three, four, and five plane waves," Opt. Express **14**(7), 3039–3044 (2006).
16. G. Ruben and D. M. Paganin, "Phase vortices from a Young's three-pinhole interferometer," Phys. Rev. E **75**(6), 066613 (2007).
17. S. Vyas and P. Senthilkumaran, "Interferometric optical vortex array generator," Appl. Opt. **46**(15), 2893–2898 (2007).

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

18. J. F. Nye and M. V. Berry, "Dislocations in wave trains," Proc. R. Soc. Lond. A **336**(1605), 165–190 (1974).
19. M. W. Beijersbergen, R. P. C. Coerwinkel, M. Kristensen, and J. P. Woerdman, "Helical-wavefront laser beams produced with a spiral phaseplate," Opt. Commun. **112**(5-6), 321–327 (1994).
20. J. Verbeeck, H. Tian, and P. Schattschneider, "Production and application of electron vortex beams," Nature **467**(7313), 301–304 (2010).
21. A. Popiołek-Masajada and W. Frączek, "Evaluation of a phase shifting method for vortex localization in optical vortex interferometery," Opt. Laser Technol. **43**(7), 1219–1224 (2011).
22. E. Frączek, W. Fraczek, and J. Mroczka, "Experimental method for topological charge determination of optical vortices in a regular net," Opt. Eng. **44**(2), 025601 (2005).
23. V. V. Volkov and Y. Zhu, "Deterministic phase unwrapping in the presence of noise," Opt. Lett. **28**(22), 2156–2158 (2003).
24. L. J. Allen, H. M. Faulkner, K. A. Nugent, M. P. Oxley, and D. Paganin, "Phase retrieval from images in the presence of first-order vortices," Phys. Rev. E **63**(3), 037602 (2001).
25. I. Freund and N. Shvartsman, "Wave-field phase singularities: The sign principle," Phys. Rev. A **50**(6), 5164–5172 (1994).

1. Introduction

Interferometry is commonly used for wavefront determination in visible optics [1], cold atom optics [2], electron optics [3], and X-ray microscopy [4]. A two-beam interferogram is formed by interference between a wave that has passed through a specimen and a reference wave that does not interact with the specimen. Fourier-transform methods are commonly used to reconstruct the intensity and phase of the wave [5]. Although these methods are simple to implement, the spatial resolution of the reconstruction is limited by how isolated the spatial-frequency spectrum is from the central maximum. Transform methods also involve inverting a subsection of the Fourier transform of the interferogram; since a specific region of the desired real-space phase map cannot be localized in Fourier space, the phase is reconstructed non-locally. For example, poor fringe visibility in a subsection of the interferogram can introduce systematic errors in the recovered wave over the entire field of view. Further, a Gibbs-type phenomenon, namely 'ringing', can occur when processing subsections of interferograms that exhibit strong scattering or absorption, leading to poor contrast resolution.

The method developed in the present paper uses three-wave interference to generate a uniform lattice of optical vortices, which are used to recover the phase of the input wave-field. Perturbations to the input wave-field due to an object distort the vortex lattice. The resulting transverse displacements of the vortices are then used to determine the phase at the exit surface of the object [6]. The exit surface is the plane where the exit wave resides once it has propagated through the object. We show that the phase is proportional to the positions of each vortex in the lattice, thereby permitting the phase to be retrieved locally without introducing 'ringing' artefacts. Further, the method is also applicable to absorbing objects.

The utility of vortex lattices has been investigated in other contexts, such as wavefront tilt [7], small-angle rotations of wave-vectors [8], and to determine polarization parameters of birefringent media [9,10]. Reference [11] reports a method of tracking vortices in a vortex lattice, which plots the relative phase between each vortex in the lattice as a function of their transverse Cartesian coordinates (x, y). A plane is then fitted to this function and the distance between the positions of the vortices relative to the fitted plane is taken as the phase. However, in this method [11] the plane is fitted over the entire field of view and the phase solution at any point depends on all parts of the interferogram; consequently it is also non-local. In contrast, the method reported here reconstructs the phase locally, as it is based on an algebraic method that does not require integration.

The outline of the paper is as follows. In section 2 we detail the method and show how the position of vortices generated by three-beam interference can be used for phase determination. The experimental setup is described in section 3, with results presented in section 4. Section 5 provides a brief discussion of the technique and its potential applications.

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

2. Theoretical background

Consider the superposition of three coherent plane scalar waves of unit intensity. The first wave, ψ_{obj}, is directed along the z-axis and transmitted through a phase-amplitude object.

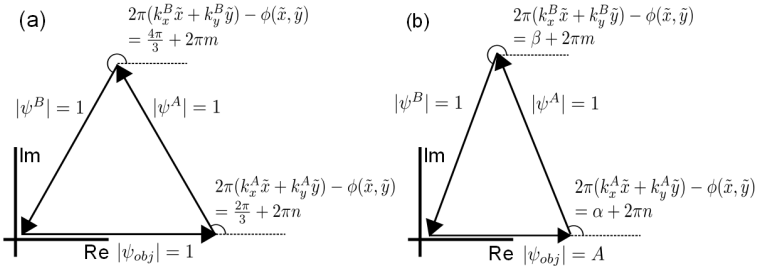

Fig. 1. (a) Phasor geometry for three waves of unit intensity at a vortex point. (b) Phasor geometry for a vortex when the exit wave is transmitted through an absorbing object.

The complex transmission function is $A(x,y)\exp[i\phi(x,y)]$, where $A(x,y)$ is a real function of (x,y) and $\phi(x,y)$ is the phase shift imparted to the wave by the object in the projection approximation [12]. Note that the trivial harmonic time dependence $\exp(-i\omega t)$, where ω is angular frequency and t is time, is suppressed throughout. Now consider a second and third plane wave, labeled $\psi^A(x,y)=\exp[i2\pi(k_x^A x+k_y^A y)]$ and $\psi^B(x,y)=\exp[i2\pi(k_x^B x+k_y^B y)]$, respectively, which are both tilted with respect to the z-axis. The superposition of the three wave-fields, $\Psi(x,y)$, at $z=0$ is given by

$$\begin{aligned}\Psi(x,y)&=\psi_{obj}(x,y)+\psi^A(x,y)+\psi^B(x,y)\\&=\exp[i\phi(x,y)]\times\\&\quad\{A(x,y)+\exp[i2\pi(k_x^A x+k_y^A y)-i\phi(x,y)]+\exp[i2\pi(k_x^B x+k_y^B y)-i\phi(x,y)]\}.\end{aligned}\tag{1}$$

It can be shown that Eq. (1) results in a lattice of phase vortices of both positive and negative topological charge, provided that all three wave vectors are not co-planar and $|A|<2$ [12–17]. At the vortex core P the phase of the wavefunction becomes undefined and the intensity vanishes [18]. Computing the line integral of the phase gradient over a smooth simply-connected closed path Γ enclosing P we obtain:

$$\oint_\Gamma \nabla\phi\cdot\mathbf{t}\,ds = 2\pi p,\tag{2}$$

where $\mathbf{t}$ is the unit tangent vector to the curve Γ, ds is the infinitesimal line element along Γ and p is a non-zero integer known as the topological charge of the vortex. The charge corresponds to the number of times the phase winds by 2π around the vortex core as the curve Γ is traversed; its sign represents the rotational direction of the phase winding, with anti-clockwise conventionally denoted as positive. Optical vortices may be engineered by the use of a spiral phase mask [19], a synthetic hologram [20], or the superposition of three or more coherent scalar waves [12–17]. At each vortex core the intensity must vanish, therefore at any vortex location the wavefunction must vanish.

Each term in Eq. (1) represents a phasor in the complex plane. The sum of the three plane waves will only vanish when each phasor in the braces in Eq. (1) forms a closed triangle as

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

shown in Fig. 1(a) (cf [12, 14].). To demonstrate how the phase is related to the position of each vortex, we consider two cases; first a completely transparent object (i.e. where A is unity) and secondly that of an absorbing object.

2.1. Non-absorbing object

For a non-absorbing object, each vortex occurs when the three phasors form an equilateral triangle as shown in Fig. 1(a), where $(\tilde{x}, \tilde{y})$ denotes the vortex location. Note that there are two possible phasor arrangements that can form an equilateral triangle. The second triangle is formed by complex conjugation of all three phasors, corresponding to reflection about the x-axis in Fig. 1(a). This dual arrangement corresponds to alternate signs of the topological charge of each vortex.

The arrangement of the phasors shown in Fig. 1(a) occurs when the arguments of the second and third phasor take the following values:

$$2\pi(k_x^A \tilde{x} + k_y^A \tilde{y}) - \phi(\tilde{x}, \tilde{y}) = \frac{2\pi}{3} + 2\pi n, \tag{3a}$$

$$2\pi(k_x^B \tilde{x} + k_y^B \tilde{y}) - \phi(\tilde{x}, \tilde{y}) = \frac{4\pi}{3} + 2\pi m, \tag{3b}$$

where n and m are integers. Substitution of Eqs. (3a) and (3b) into Eq. (1) shows that $\Psi(\tilde{x}, \tilde{y})$ vanishes for all n and m. Since the relative angles between phasors will remain constant at each vortex point $(\tilde{x}, \tilde{y})$, we may substitute Eq. (3a) into Eq. (3b) to obtain the relation:

$$4\pi(k_x^A \tilde{x} + k_y^A \tilde{y}) - 2\phi(\tilde{x}, \tilde{y}) = 2\pi(k_x^B \tilde{x} + k_y^B \tilde{y}) - \phi(\tilde{x}, \tilde{y}) + 2\pi(m - 2n). \tag{4}$$

Solving for the phase we find

$$\phi(\tilde{x}, \tilde{y}) = 2\pi(2k_x^A - k_x^B)\tilde{x} + 2\pi(2k_y^A - k_y^B)\tilde{y} + 2\pi l, \tag{5}$$

where l is an integer. Equation (5) shows that the phase shift in the exit wave-field depends directly on the coordinates of each vortex, up to an additive constant. Therefore, the phase of the object wave may be retrieved from the coordinates of each vortex in the distorted lattice along with knowledge of the transverse wave-vectors of each reference wave.

2.2. Absorbing object

Here we generalize the analysis of section 2.1, to the case of an absorbing object. The phasor geometry for an absorbing object is shown in Fig. 1(b), where the magnitude of the phasor of the exit wave can vary due to absorption. This means that Eqs. (3a) and (3b) will no longer be valid for every vortex point, since the relative angles between each phasor can vary. The angles that reference waves A and B make with the x-axis are denoted by α and β, respectively. Due to the symmetry of the phasor arrangement in Fig. 1(b) we note that $\beta = 2\pi - \alpha$ for any $|\psi_{obj}|$. Therefore we may write

$$2\pi(k_x^A \tilde{x} + k_y^A \tilde{y}) - \phi(\tilde{x}, \tilde{y}) = 2\pi(m + n + 1) - 2\pi(k_x^B \tilde{x} + k_y^B \tilde{y}) + \phi(\tilde{x}, \tilde{y}). \tag{6}$$

Equation (6) may be solved for the phase to give

$$\phi(\tilde{x}, \tilde{y}) = \pi(k_x^A + k_x^B)\tilde{x} + \pi(k_y^A + k_y^B)\tilde{y} + \pi s, \tag{7}$$

where s is an integer. Equations (5) and (7) are the two main results of this paper, and form the basis of our phase determination method.

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

Equation (7) is applicable for strongly absorbing objects, with the exception of the extreme case of total absorption by the object. The magnitude of the phasor of the object beam decreases with increasing absorption; unless the object is completely absorbing, the phasors will still sum to zero and a vortex will be formed. However, strongly absorbing objects will decrease the contrast resolution of the interference pattern. Moreover, low contrast can make it difficult to accurately localize a vortex, which requires the detection of minima in the intensity pattern. However, there are phase shifting methods that use an extra reference arm to form a fringe pattern, which can be exploited to increase the intensity gradient near a vortex core. Phase shifting the reference arm to generate a fringe pattern causes shifts in the intensity maxima, whilst the intensity minima due to vortices remain fixed at the same locations. By taking a series of interferograms, each with the fourth interferometer arm phase shifted, the absolute value of the difference between successive interferograms is calculated. These differences are summed, which increases the local intensity gradient near the vortex core; this compensates for the lack of contrast in the three-beam interferogram of a highly absorbing object [21]. Such an approach could be used to increase the precision with which vortices can be localized in those cases where the absorption is too large to accurately locate each vortex.

The phasor geometry used to derive Eq. (5) considered a single sign of topological charge. Therefore only the coordinates of one particular type of vortex are used to determine the phase using Eq. (5). This is not the case for Eq. (7), where there are two arrangements of the phasors that sum to zero; however, both result in the same equation. This is due to the symmetry in the coefficients of $\tilde{x}$ and $\tilde{y}$ in Eq. (7), which is not present in Eq. (5). Thus Eq. (7) allows us to utilize either vortex sign in the interference pattern to determine the phase. The presence of the third term in Eq. (7) shows that the phase is retrieved modulo π rather than modulo 2π. This is only a disadvantage for objects that have large phase gradients, in which case aliasing may occur and phase unwrapping is difficult.

The proposed method of phase determination proceeds as follows. First we localize all vortices in the lattice, resulting from a three-beam interference pattern, and determine their $(\tilde{x}, \tilde{y})$ coordinates by locating intensity zeros. Using either Eq. (5) or (7), we algebraically calculate ϕ for each vortex point $(\tilde{x}, \tilde{y})$ .The use of either equation is dependent on the object being imaged, i.e. for absorbing objects Eq. (7) is used. If Eq. (5) is utilized, an additional step is required in which we separate vortices by the sign of their topological charge. The wrapped phase is then recovered modulo 2π, if Eq. (5) is used, or modulo π in the case of Eq. (7). Once the phase is computed, interpolation between each vortex is performed to recover the phase over all (x, y) points in the image array.

3. Experiment

The experimental setup is shown in Fig. 2. The beam from a linearly-polarized Helium-Neon laser (Thorlabs 5 mW) is spatially filtered by focusing it through a 4.51 mm focal length aspheric lens and then passing it through a 20 μm pinhole. The filtered beam is re-collimated using a 100 mm focal length plano-convex lens and an iris, which is adjusted so that the beam is truncated at its first minimum. A neutral density filter located before the first focusing lens is used to attenuate the beam. The filtered beam is then passed through a polarizing beamsplitter cube, oriented to transmit the majority of the beam power; this ensures a pure polarization state for the resulting beam. The beam is then passed into a three-beam interferometer constructed from a pair of Mach-Zehnder interferometers that share a common arm. The path lengths through each arm of the interferometer are matched to maximize the coherence of the interference; 50:50 beamsplitter cubes are used to split and combine beams in each interferometer.

The common central arm of the combined interferometer (object beam) is used as the source of illumination for the object. The object is located close to the first beamsplitter cube

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

in the arm and is imaged onto a CCD camera via a system of two plano-convex lenses arranged to image the object plane onto the camera with a magnification of two. The lenses are arranged so that the first lens is located one focal length away from the object ($f = 75\,\text{mm}$) with the second lens ($f = 150\,\text{mm}$) spaced 225 mm away from the first lens and 150 mm away from the camera. The monochrome CCD camera (Prosilica GE1650) has 1600×1200 pixels and 12 bits per pixel.

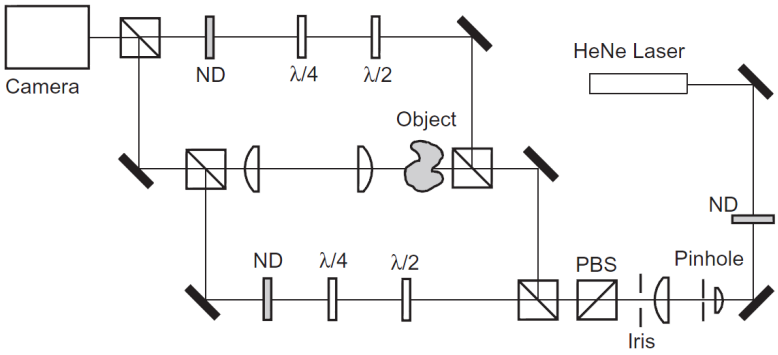

Fig. 2. Experimental setup of the three-beam vortex interferometer. A He-Ne laser (λ = 633nm) is spatially filtered by the pinhole and a single polarization direction is selected as it passes through the polarizing beam splitter. Two beamsplitters are used to create the three arms of the interferometer. The central beam is transmitted through an object after which two lenses are used to focus the exit surface of the object into the camera. Each reference arm contains λ/2 and λ/4 wave plates; which are used for phase stepping. Neutral density filters (ND) ensure that the two reference waves have the same intensity as the object wave.

The pixel size is 7.4 μm × 7.4 μm. The two remaining interferometer arms (reference beams) each contain a neutral density filter to match the intensity of the beam through each arm to that of the object beam. A $\lambda/2$ and $\lambda/4$ wave plate allows the phase of the beam exiting each arm to be varied in discrete steps. The phase is stepped by either $\lambda/2$ or $\lambda/4$ (depending on which wave plate is used) by changing the alignment of the wave plate so that the fast optical axis is either parallel or perpendicular to the polarization axis of the beam.

To form a vortex lattice, two reference beams are adjusted in angle relative to the object beam, such that the wave-vectors of the three beams are non-coplanar. The angles of the beams of the reference arm can be adjusted by tilting the final beamsplitter in the top arm of the interferometer and the final mirror in the bottom arm. Different geometries of the lattice are generated depending on the mutual angle of the three beams, with a hexagonal lattice being produced when the angle between any two beams is 120°. The period of the three-beam interferogram maxima was adjusted to be approximately 20 pixels. A phase shift of either reference beam causes a shift in the vortex in the direction of that beam's transverse wave-vector.

Our technique was experimentally tested on two objects: a Thorlabs spherical lens (part number LA1464) made from N-BK7 optical glass with a diameter of 25.4 mm, and the wing of a common house fly (Musca Domestica). For the spherical lens, in addition to the three-beam pattern, two-beam interference patterns of $\psi_{obj} + \psi^A$ and $\psi_{obj} + \psi^B$ were individually acquired by blocking each of the reference arms in turn. Both reference arms were then blocked to record an image of the Gaussian illumination. Wave-vectors of each arm were measured by locating the peak maximum in the power spectrum of the three-beam

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

interference pattern. The interferogram was smoothed, to reduce noise, by convolution with a Gaussian filter of approximately 60 μm full width half maximum (FWHM) and then flat-field corrected using a specimen-free illumination image. This was performed to increase the signal in the outer edges of the interferograms.

To determine the accuracy with which vortices can be localized we computed the normalized autocorrelation of a non-deformed section of our vortex lattice. Since the lattice is uniform, perfect localization gives complete autocorrelation between vortex points (i.e. the autocorrelation value is unity). However, small errors in localization decrease the autocorrelation for the vortex lattice, giving some width to the central maximum of the autocorrelation function. Using the half width of the maximum as a measure of the precision of vortex localization, it was found that the technique has a precision of 1.5 pixels. Based on Eq. (5) this translates to a measured phase accuracy of ± 0.2 radians.

4. Results

A section of the three-beam interferogram of the lens along with a simulated interferogram is reproduced in Fig. 3(a) and 3(b) respectively; these show excellent agreement is found between theory and experiment. To locate the vortices in the image, it is only necessary to locate the intensity zeros in the interference pattern. A zero in the intensity is a necessary but not sufficient condition for a vortex; a three-beam interference pattern of planar waves will always have a vortex at each intensity zero [12,14]. Note, in this context, that non-vortical zeros are not stable with respect to perturbation, whereas vortical zeros are stable with respect to perturbation. The vortices were separated depending on the sign of their topological charge using the methods outlined in Ref [22]. The phase of the lens was calculated using Eq. (5) for vortices of positive topological charge. Two-dimensional linear interpolation was then performed between each vortex point to map the phase over all Cartesian pixel coordinates within the field of view. This was performed by taking the real and imaginary parts of the wavefunction. The real and imaginary parts were then linearly interpolated, which allowed for the complete wrapped phase $\phi(x, y)$ to be recovered using

$$\phi(x, y) = \tan^{-1}\left\{\frac{\sin[\phi(\tilde{x}, \tilde{y})]}{\cos[\phi(\tilde{x}, \tilde{y})]}\right\}. \quad (8)$$

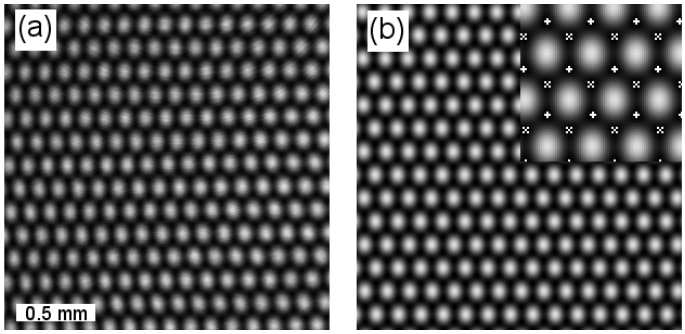

Fig. 3. Comparison of the experimental three-beam interference pattern with a numerical simulation. (a) Experimental three-beam interference pattern. (b) Simulated three-beam interference pattern. The plus signs and crosses in the top right insert indicate the locations of vortices of positive and negative topological charge, respectively.

The reconstruction of the lens in Fig. 4(a) shows a series of concentric circles where the phase has been wrapped by 2π, as would be expected for a spherical lens since its projection

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

is rotationally symmetric. To demonstrate our technique quantitatively, the wrapped lens phase was unwrapped using the methods outlined in Ref [23]. In this paper, Volkov and Zhu showed that the image may be represented in terms of the Fourier transforms of its x and y derivatives. The gradient of the wrapped phase suffers abrupt changes of 2π per pixel at each point where the phase undergoes wrapping. Thresholding the x and y phase gradients eliminates these discontinuities. The unwrapped phase is then recovered using Fourier transforms to reconstruct the phase from the thresholded x and y derivatives. Once the lens data had been unwrapped, the projected thickness was recovered by dividing the unwrapped phase by the factor δk , where δ is the difference of the refractive index of the lens from unity and k is the wavenumber. A circular arc was fitted to the profile centered on the recovered projected thickness in order to measure the radius of curvature of the lens. The profile of the lens and the fitted curve are shown in Fig. 4(b). The fitted curve returned a value of 519±1 mm, in good agreement with the quoted radius of curvature of the lens of 515.5 mm. This shows that our method is able to quantitatively reconstruct the object function.

The method was also applied to a three-beam interferogram of the wing of a common house fly. As this sample contained higher spatial frequency information than the lens, the sampling of the phase, and consequently, the number of vortices needed to be increased. This was achieved by using the $\lambda/2$ and $\lambda/4$ wave plates described in section 3 to perform phase shifting in each reference arm of the interferometer. Three-beam interferograms were recorded for the fly's wing, with 11 phase shifts of $\pi/4$ between images used to reconstruct the phase. Each interferogram was smoothed by convolution with a Gaussian filter of approximately 60 μm FWHM and the transverse wave-vectors and vortex points were found using the same method as for the spherical lens data. Vortices were not separated by sign for this object as Eq. (7) is applicable to vortices of either topological charge. For each of the 11 phase-stepped interferograms, the phase was recovered using Eq. (7). Before the phase of each image could be combined for interpolation, the relative offset between the phase of each image needed to be determined. Two-dimensional cross-correlation of a small portion of the vortex lattice was used to determine the offset between each phase step. The coordinates of the maximum in the cross-correlation data were substituted into Eq. (7), which gave the value of the phase difference between each vortex lattice. This phase difference was added to each reconstruction and linear interpolation was applied to every vortex point in all 11 phase stepped lattices. This results in a single phase image which incorporates information from all

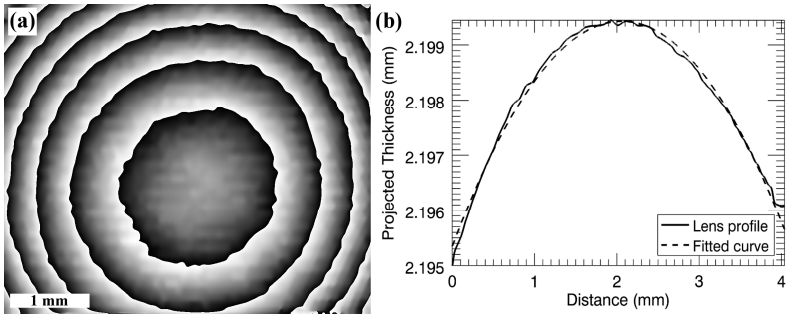

Fig. 4. Experimental results for phase reconstruction of a spherical lens using the three-beam vortex interferometer. (a) The unwrapped reconstructed phase of the lens. The greyscale in this image ranges from 0 (black) to 2π (white). (b) Unwrapped phase profile of the lens. The solid curve corresponds to the experimental data whilst the dashed curve is the fitted curve. The radius of curvature of the fitted profile is 519 ± 1 mm.

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

11 recorded vortex lattices. The recovered wrapped and unwrapped phase reconstruction of the fly's wing is shown in Figs. 5(a) and 5(b), respectively. The phase determined by the Takeda method [5] using a mask of 25 × 25 pixels applied to a two-beam interference pattern is given in Fig. 5(c) for comparison. The two-beam interference pattern was obtained by blocking one of the reference arms of the three-beam interferometer. This was also applied to a subsection of the three-beam interference pattern spatial-frequency spectrum, producing a consistent result to the phase obtained for the two-beam case. This is expected as a three-beam interference pattern implicitly contains all the information of a two-beam interference pattern. The phase recovered from the vortex lattice method was unwrapped using the same method [23] as for the lens data and is shown in Fig. 5(d). As Eq. (7) has been used to recover the phase of the fly wing, the phase is wrapped from $-\pi/2$ to $\pi/2$. The phase recovered from the Takeda method [5] has also been wrapped between these values, which facilitates comparison of the two results. Both images in Figs. 5 (a) and 5(b) show good quantitative agreement with each other. Minor differences are observed around the frame of the wing, where the absorption of the light is strongest. However, the vortex lattice phase determination technique is mostly unaffected by the absence of information in these regions. The local nature of our technique allows the phase reconstruction to simply ignore these strongly absorbing regions, which are then interpolated. For Fourier-based methods these regions result in larger phase gradients. This can be observed at the bottom edge of the wing where the phase rapidly increases, which is not observed in Fig. 5(b). Ringing artefacts in Fig. 5(c) are also observed, corresponding to spatial frequencies not present in the fly's wing. Our vortex lattice method is not susceptible to this type of artifact as the technique uses a local algebraic solution to recover the object's phase.

5. Discussion

A surprising aspect of the technique described in this paper is that vortices, which correspond to points at which the phase becomes undefined, are used to determine the phase shifts of an object. In many forms of phase retrieval vortices are often undesirable, since their presence can be detrimental to accurate phase retrieval [24]. The separation of vortex sublattices was performed using Eq. (5), which depends on a particular geometry of the phasors of the three waves; hence it applies to a particular topological sign of the sublattice. The ability to separate vortices based on their sign is a consequence of the vortex sign rule [25]. A vortex occurs at the intersection of lines, where the real and imaginary parts of the wavefunction vanish; the sign rule states that each time a line intersects with another, the sign of the topological charge alternates. If we assign any vortex to a particular charge, the sign rule allows us to determine the sign of every vortex present in the lattice. Hence we may separate the vortices in the lattice based on their topological sign. To reconstruct the phase without a sign ambiguity, using Eq. (5), the sign of k_x and k_y for each reference wave and the topological charge of each vortex must be determined. This is because the sign of the vortex determines the form of the equation used to calculate the phase, and the sign of k_x and k_y determines the value of the coefficients of $\tilde{x}$ and $\tilde{y}$. These steps can be bypassed by using *a priori* knowledge of the object, e.g., an object with refractive index greater than unity results in a positive sign in the phase shift, as a consequence of the retardation of the wave as it passes through the medium. If it is necessary to determine these parameters, the sign of k_x and k_y for each reference arm and the topological charge of the vortex sublattice can be experimentally measured using the method outlined in Ref [22].

The spatial resolution of the technique is determined by the spacing of the vortices in the lattice, which increase as the angle that each wave-vector makes with the *z*-axis increases; this allows the spatial resolution to be tuned. However, increasing the vortex density too much results in aliasing, which can lead to inaccurate vortex localization. Changing the relative

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

angles between the transverse wave-vectors can also be used to increase sampling, as this affects the geometry of the lattice, e.g., a hexagonal lattice geometry occurs when the angle between each wave-vector is 120°, whilst a rectangular geometry occurs when two wave-vectors are orthogonal. This could also be used as an alternative to phase stepping, obviating the need for wave plates; in this case the relative angles between wave-vectors can be changed for each image. The changing lattice would result in a shift of the vortices and therefore sample more of the object.

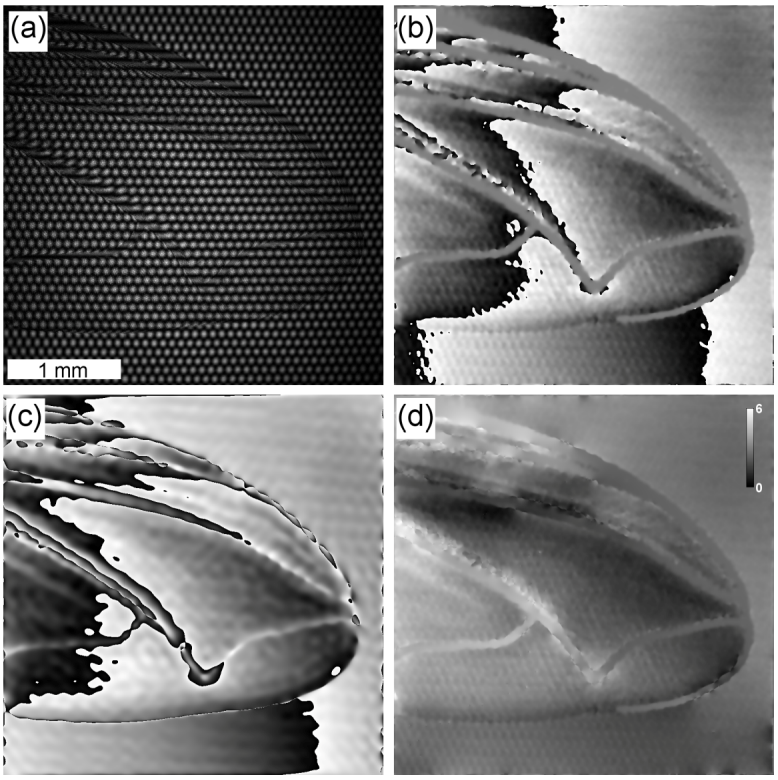

Fig. 5. Experimental results for the wing of a fly. (a) Experimental three-beam interferogram of the fly's wing. (b) Recovered phase of the fly wing using Eq. (7). (c) Recovered phase using the Takeda method [5]. (d) The unwrapped phase in (b). The greyscale in (b) and (c) is $[-\pi/2, \pi/2]$ from black to white.

An advantage of using vortex interferometry for phase retrieval is its robustness in the presence of noise. Because the method uses a real space calculation, the power spectrum of the noise has little or no effect on the reconstruction. In this case, noise only affects the vortex localization itself, since small variations due to noise may either lead to false detection of a vortex or inaccurate localization of the vortex points. However, the noise will have a significant effect on the interferogram, since shot noise is reduced for low intensity regions,

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

such as areas in the vicinity of the vortex core. These effects can be significant if the vortex lattice spacing is small and the intensity "blobs" in Fig. 3 are under-sampled.

6. Conclusion

This paper presents a new method of phase determination in interferometry. By interfering three plane waves, a uniform vortex lattice is generated. Distortions in this lattice due to phase shifts induced in the object wave have been shown to be related to the position of each vortex in the interference pattern (see Eqs. (5) and (7)). Localizing each vortex in the lattice allows for the phase to be algebraically calculated. The technique has been demonstrated on a spherical lens and a house fly's wing. Both objects show good agreement with existing Fourier methods. An advantage of the presented technique is its robustness to noise and its capacity to reconstruct the phase locally.

Acknowledgments

S.A. Eastwood gratefully acknowledges funding from an Australian Postgraduate Award. D.M. Paganin and M.J. Morgan acknowledge funding from an Australian Research Council Discovery Project Grant.

#165475 - $15.00 USD Received 29 Mar 2012; revised 27 Apr 2012; accepted 30 Apr 2012; published 8 Jun 2012

PRL 110, 033901 (2013) PHYSICAL REVIEW LETTERS week ending 18 JANUARY 2013

Electron Vortex Production and Control Using Aberration Induced Diffraction Catastrophes

T. C. Petersen,[1,3,*] M. Weyland,[2,3] D. M. Paganin,[1] T. P. Simula,[1] S. A. Eastwood,[1] and M. J. Morgan[1]

[1]*School of Physics, Monash University, Victoria 3800, Australia*
[2]*Department of Materials Engineering, Monash University, Victoria 3800, Australia*
[3]*Monash Centre for Electron Microscopy, Monash University, Victoria 3800, Australia*

(Received 21 August 2012; published 14 January 2013)

An aberration corrected electron microscope is used to create electron diffraction catastrophes, containing arrays of intensity zeros threading vortex cores. Vortices are ascribed to these arrays using catastrophe theory, scalar diffraction integrals, and experimentally retrieved phase maps. From measured wave function phases, obtained using focal-series phase retrieval, the orbital angular momentum density is mapped for highly astigmatic electron probes. We observe vortex rings and topological reconnections of nodal lines by tracking the vortex cores using the retrieved phases.

DOI: 10.1103/PhysRevLett.110.033901 PACS numbers: 42.30.Rx, 05.45.−a, 41.85.−p, 68.37.Og

Quantized vortices in propagating electron waves are of interest for electron wave- function phase mapping, since in-line holography approaches can fail in the presence of these wave front dislocations [1]. In this context, electron vortices were theoretically shown to arise from the transmission of fast electrons through an atomic lattice [2]. Recent experiments have shown how electron vortices may be produced and controlled. For example, cleaved steps of graphite films can act as approximate spiral phase masks, giving rise to dislocations in electron biprism interference fringes, thereby producing vortices [3]. It has been suggested that these screw-type phase singularities may be ubiquitous for specimens with heterogeneous thickness variations on the nanoscale [3]. However, our simulations of carbon phase objects with such thickness variations suggest that specimen geometry is important and that electron vortices are not readily produced. The coherent superposition of distorted plane waves is expected to give rise to vortices. By analogy with three-wave interference in light optics [4], we expect to experimentally observe electron vortices for Bragg diffraction from crystalline specimens. Indeed, in biprism interference patterns from crystalline specimens, one can find examples of forked dislocations [5], which are signatures of quantized phase vortices. Similarly, the hallmarks of three-wave electron interference (which is a classical method for vortex production [6]) are evident for three-beam electron diffraction from crystals [7–10]. Recently, controlled vortex beams were demonstrated using micron-scale forked masks in a transmission electron microscope (TEM) [11,12].

Electron vortex beams open up new avenues for revealing specimen properties on the nanoscale. For example, Verbeeck *et al.* [11] exploited the orbital angular momentum (OAM) in a vortex beam to create a dichroic effect for $2p_{1/2} \rightarrow 3d$ and $2p_{3/2} \rightarrow 3d$ inelastic transitions in Fe. Related experiments for subnanometer spatial resolution have since been explored using spiral masks [13] and forked masks placed in the illumination aperture of the TEM [14]. It seems plausible to associate electron beams that contain vortices with probes that can promote the exchange of OAM. However, Berry [15] has cautioned against a direct association, demonstrating that OAM is not precisely connected with the presence of vortices in general. Nevertheless, the OAM density can vary significantly across an electron beam containing vortices. For light optics, astigmatism aberrations can impart significant OAM onto Gaussian beams [16].

Quantized phase vortices have also been studied intensively in visible light optics [17–19], x-ray optics [20], for microwaves [21], and acoustics [22]. For optical wave fields it is known that diffraction catastrophes [23] give rise to caustics. Diffraction catastrophes are generic to optical wave fields and are stable with respect to perturbations; i.e. they persist upon continuously varying aberrations, such as free space propagation, maintaining recognizable shapes such as linear "folds" or pointed "cusps" [23,24]. Diffraction catastrophes and phase discontinuities generally form in the focal volumes of lenses with aberrations [25,26]. Berry *et al.* [17] demonstrated the elliptic umbilic catastrophe using a triangular lens formed by water. Further, it was shown that the associated diffraction detail can be accurately approximated using superposition of plane waves, which create a lattice of phase vortices. The generation of vortex lattices by the superposition of three plane waves [4,6], or distorted spherical waves [27], has been studied in wave optics and, more recently, in Bose-Einstein condensates [28,29]. The analysis of umbilic and cusp catastrophes by Berry *et al.* [17] demonstrates that vortical wave fields can be formed by lenses with aberrations.

Optical caustics formed by primary aberrations have been widely recognized since the pioneering work of Nijboer and Nienhuis [30–32]. Caustics arising from lens aberrations in the TEM have also been characterized theoretically using geometric optics [33]. Primary aberrations can induce diffraction catastrophes; in particular, the umbilic foci for astigmatism and coma have been shown to be hyperbolic

[34]. Caustics are routinely observed when TEM illumination apertures are removed to include electron trajectories deflected through large angles. These rays do not produce significant interference effects in the probe intensity, since the removal of probe-forming apertures degrades the spatial coherence, which would seem to exclude the possibility of observing phase dislocations arising from diffraction catastrophes. However, partial coherence can be improved if small illumination apertures are used. Diffraction catastrophes can then be explored by imposing significant probe forming aberrations to disturb the point focus and create severely distorted coherent electron probes.

We have conducted experiments to produce and manipulate electron vortices in the specimen plane of a TEM, employing lens aberrations to create diffraction catastrophes. We used a Titan^3 80–300 TEM (FEI) that provides dual aberration correction (CEOS GmbH) of both the illumination and imaging lenses. Operating at an acceleration voltage of 300 kV, imaging lens aberrations were corrected to third order and the microscope was set up in the bright field imaging mode to minimize the semiangle subtended by the field emission gun source. A 10 μm condenser aperture was selected; the resulting intensity distribution was a small yet parallel probe. By adjusting the condenser lenses, the illumination was focused in the specimen plane to produce a far-field diffraction pattern of the circular condenser aperture in the form of Airy rings. The circular symmetry of these rings was then broken by increasing twofold condenser astigmatism, producing a sharp (subnanometer) line focus streaking along one direction. Through-focus propagation of the resulting distorted electron probe, using the imaging lens, revealed the presence of four umbilic foci, which outlined the caustic of the probe. The electron probe cross section and decorating diffraction detail maintained form as the probe was imaged throughout the focal series. This observed persistence of form and stability under perturbations is a key aspect of a diffraction catastrophe.

Figure 1(a) shows the logarithm of an astigmatic experimental electron probe imaged near one of the line foci. Figure 1(b) shows the same probe imaged several hundred nanometers further along the optic axis. Since the electron optical configuration was chosen to optimize the spatial coherence, the diffraction pattern was quite dim. Accordingly, the acquisition times were 100 s per probe pattern. Comparing Figs. 1(a) and 1(b), the stability of the umbilic features with respect to smooth variation of the electron wave (changing focus) implies the formation of a diffraction catastrophe. Electron trajectories make very small angles with the optic axis and we can assume paraxial imaging conditions. If we further ignore chromatic aberrations or the spread of energies in the electron source, we can utilize the diffraction theory of aberrations based upon the Huygens-Fresnel principle [35]. To this end, we consider the condenser lens aberrations as modifying the optical path lengths of spherical waves originating from a circular aperture $A(x, y)$. Fresnel propagation from the

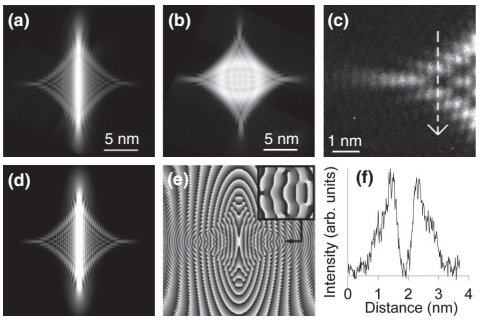

FIG. 1. (a) Experimental intensity (logarithm) near a line focus caused by condenser astigmatism. (b) Experimental intensity (logarithm) at a defocus half way between two line foci. (c) Enlarged experimental intensity for the leftmost umbilic focus in (a). (d) Logarithm of the caustic intensity computed from Eq. (1). (e) Corresponding phase computed from Eq. (1), exhibiting an array of phase vortices. (f) Five-pixel wide line plot measured from (c) showing interior intensity zeros threading vortex cores.

aperture plane describes the electron wave at a distance z along the optic axis, downstream from the aperture:

$$\Psi(x, y, z) = S(x, y, R_2)F^{-1}\{F[A(x, y)S(x, y, R_1)e^{(2\pi i/\lambda)C(x^2+3y^2)}] \times e^{-i\pi\lambda z(q_x^2+q_y^2)}\}, \quad (1)$$

where q_x and q_y are the Fourier coordinates conjugate to the Cartesian aperture-plane coordinates x and y, $\Psi(x, y, z)$ is the scalar wave function and λ is the wavelength. The symbols F and F^{-1} in Eq. (1) denote forward and inverse Fourier transforms, respectively. Primary astigmatism in the aperture plane is parametrized by the coefficient C using the polynomial form given by Kingslake [36]. The function $S(x, y, R_1)$ models the electron source as a simple spherical wave in the paraxial approximation with radius R_1, which numerically serves to condense the wave within an image array of fixed size [37]. The spherical wave $S(x, y, R_2)$ removes residual scaling of $\Psi(x, y, z)$ after propagation. Adjustment of the free parameters in Eq. (1), namely, C and R_1, produced excellent agreement between the experimental results in Fig. 1(a) and the theoretical predictions in Fig. 1(d).

The intensity logarithm and phase calculated from the diffraction integral in Eq. (1) are shown in Figs. 1(d) and 1(e), respectively. Between the umbilic foci in Fig. 1(d), spontaneously nucleated phase vortices decorate the outer edges of the caustic and some are within the interior, where intensity zeros occur in Figs. 1(a), 1(c), and 1(d). Several minima within the umbilic foci in Fig. 1(a) contain intensity zeros, which is consistent with the creation of electron vortices. However, none of the minima inside the caustic in Fig. 1(b) correspond to intensity zeros, in agreement with diffraction integral calculations. Figure 1(f) shows a five-pixel wide line plot of the intensity shown in Fig. 1(c). This

line plot confirms the presence of intensity zeros near the line focus, inside the caustic. The observation of intensity zeros is a necessary condition for the existence of vortices.

The Titan^3 80–300 TEM (FEI) provides sensitive control over lens aberrations; however, strong excitation of corrector lenses leads to parasitic aberrations and it is difficult to apply pure aberrations. To model probe cross sections accurately, it is generally necessary to characterize and account for a multitude of probe forming and imaging lens aberrations [38]; yet in Eq. (1) we have only included primary astigmatism. The detailed agreement between experiment and theory is a consequence of catastrophe theory, in that the umbilic foci represent structurally stable forms of the electron wave. To further validate this observation, astigmatism was minimized and the probe corrector was adjusted to induce primary coma, since the form of this aberration function produces the hyperbolic umbilic catastrophe [34]. Three basic examples of caustics with varying coma, collected using exposure times of 5 s per diffraction pattern are provided in the Supplemental Material [39], showing excellent agreement with earlier observations of caustics reported in coherent light optics [32,40]. To explore the coma induced diffraction catastrophe, the coma caustic was further enlarged by increasing the size of the condenser aperture to 150 μm, which diminished the contrast of the pattern. To compensate for the reduced coherence, the first condenser lens was excited to the maximum nominal setting after which the fringe contrast increased significantly. Additional coma caustics were then recorded using multiple frames (ten) and shorter exposures (3 s per frame) to offset residual beam drift.

Figure 2 compares experiment with both catastrophe theory for the hyperbolic umbilic and the diffraction integral for the coma aberration. The experimentally measured intensity in Fig. 2(c) is an average of 50 images, each exposed for 1 s and then postaligned with respect to each other. The horizontal line in Fig. 2(c) marks a single-pixel wide line plot, which is displayed in the Supplemental Material [39], showing significant fringe visibility and an array of intensity zeros. For Fig. 2(d), 10 images were acquired for 3 s, postaligned and averaged. Arrays of intensity zeros were also observed at this defocus setting, which was nominally 100 nm from that shown in Fig. 2(c). The logarithm of the intensity in Fig. 2(d) compares well to that of the hyperbolic umbilic catastrophe in Fig. 2(a) and the diffraction integral intensity logarithm in Fig. 2(e). The hyperbolic umbilic diffraction catastrophe is described by

$$\Psi(x, y, z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(s_1^3+s_2^3-xs_1-ys_2-zs_1s_2)}ds_1ds_2 \quad (2)$$

where s_1 and s_2 are the relevant state variables [34] and the wave function $\Psi(x, y, z)$ is computed at focal depth z and plotted in the plane spanned by (x, y). Equation (2) was evaluated numerically with the coordinates x, y normalized by the number of pixels, to span ± 30 dimensionless units over 512×512 pixels. The integration variables s_1, s_2 were truncated to ± 3.5 and incremented in steps of 0.014

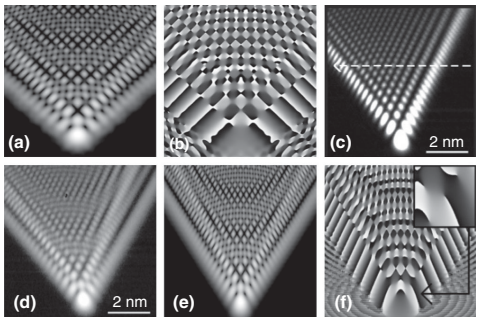

FIG. 2. (a) Logarithm of the intensity calculated using diffraction catastrophe theory. (b) Phase of the diffraction catastrophe showing a distorted lattice of vortices and antivortices. (c) Experimental intensity of coma caustic near the diffraction focus. (d) Logarithm of the experimental intensity nominally 100 nm away from the diffraction focus. (e) Logarithm of the coma caustic intensity computed from the diffraction integral in Eq. (3). (f) Phase of the coma caustic from the diffraction integral in Eq. (3).

dimensionless units. The wave function $\Psi(x, y, z)$ was then cropped to 256×256 pixels to approximately match the field of view of the experimental data. The phase of $\Psi(x, y, z)$, determined from Eq. (2), is shown in Fig. 2(b), where arrays of phase vortices decorate the interior and outer portions of the coma caustic. Similar vortices are evident in Fig. 2(f), which was computed using the diffraction integral:

$$\Psi(x, y, z) = S(x, y, R_2)F^{-1}\{F[A(x, y)S(x, y, R_1)e^{(2\pi i/\lambda)By(x^2+y^2)}] \times e^{-i\pi\lambda z(q_x^2+q_y^2)}\}, \quad (3)$$

where all symbols in Eq. (3) are identical to those in Eq. (1) except for the coefficient B, which parametrizes the degree of coma [36]. This parameter was varied until the caustic intensity pattern visually matched the experimental data.

Electron phase maps were experimentally determined from a through-focus series of images using the Gerchberg-Saxton-Misell phase retrieval algorithm [41–45] (for details, see the Supplemental Material [39]). Objective lens defocus was calibrated using power spectra of experimental images using the DIFFTOOLS suite of scripts [46]. Figure 3(a) shows the logarithm of the retrieved intensity for the first image in the through-focus series for the astigmatism caustic. Comparing Fig. 3(a) and 1(b), it is evident that most of the diffraction detail was captured by the retrieval algorithm, although the intensity minima are not as distinct. The retrieved phase map in Fig. 3(b) has a similar form to that in Fig. 1(e). Propagation of the reconstructed wave over many focal planes correctly predicted a pair of mutually orthogonal sharp line foci, as well as preservation of the caustic shape and diffraction detail with varying focus, all of which were observed in our experiments. Transverse current densities were computed

from the retrieved waves (see Supplemental Material [39]). The phase and current density vectors were rendered using software scripts based upon the algorithms of Mitchell and Schaffer [47].

Figure 3(c) shows the retrieved intensity logarithm for the first image in a through-focus series for the coma caustic. Vortices are seen to decorate the exterior of the coma caustic, with several inside the caustic shown in Fig. 3(d). Again, the intensity minima are not as distinct when compared to the experimental input. Nonetheless, the reconstructed wave reproduced much of the detail observed in the experimental images over the entire focal range. Figures 3(e) and 3(f) show the experimentally measured z component of the OAM density, determined for the retrieved wave function about a central cross section of the astigmatism caustic, displayed in SI units of 1.0×10^{-16} kg s^{-1} per electron. The root mean square value of the OAM density in Fig. 3(f) is $0.15\hbar$ nm^{-2}. The OAM density in Figs. 3(e) and 3(f) varies significantly across the beam, which is analogous to the optical interferometry measurements of Courtial *et al.* [16] for elliptical Gaussian beams. The small boxes shown in Figs. 3(e) and 3(f) have side lengths of 0.7 nm, within which the integrated OAM per electron is $42\hbar$ and $35\hbar$, respectively. Within both boxes, the corresponding average probability densities are greater than 75% of the maximum intensity over the entire field of view.

Using the reconstructed experimental wave functions, vortices were tracked throughout a propagated focal series comprising 2048^3 points and detected by measuring points of nonzero circulation in two-dimensional phase maps. Arrays of tracked vortices were represented as small spheres [48] and are shown in the Supplemental Material [39]. The vortex tracks exhibited fine detail, such as Kelvin waves along nodal lines and vortex loops. The observed nodal line instabilities were possibly due to wave perturbations from residual aberrations in the experiment, since it is known that additional plane waves can warp the structure of nodal lines in three-wave vortex lattices [49]. In accordance with known results in light optics [19], we also observed vortex loops in diffraction integral calculations, using Eqs. (1) and (3), which portrayed similar structures to nodal lines computed from the experimentally retrieved wave functions. In light optics, entire arrays of nodal loops can be formed from the superposition of four plane waves [19]. At the other extreme, tangled vortices and nodal loops can exist in self-similar speckle fields created by a ground glass screen, where the vortex lines possess a fractal structure [50]. In the context of catastrophe optics, nodal loops have been identified and studied in the hyperbolic umbilic diffraction catastrophe [51]. For our retrieved wave functions, we examined particular nodal lines in detail by cropping out all tracks except those in the subregions marked in the phase maps of Figs. 3(b) and 3(d). The electron vortex loop from the coma caustic in Fig. 4(a) shows nodal line excitations and Crow-like instabilities [52], apparently initiating dissociation into several vortex loops.

The colored plane in Fig. 4(a) shows the phase windings around adjacent sides of the loop, which highlight vortices of opposite topological charge. Figure 4(b) shows three electron vortices represented as intensity iso surfaces, which are adjacent to the umbilic focus of the astigmatism caustic shown in Fig. 3(b).

In conclusion, we have induced electron diffraction catastrophes, thereby creating distorted lattices of spontaneously nucleated electron vortices, using an aberration corrected electron microscope. We have demonstrated that phenomena, such as Crow-type nodal-line instabilities and nonlinear effects, such as pair creation or annihilation, can be measured for matter waves that obey a linear wave equation. Our singular electron optics observations closely

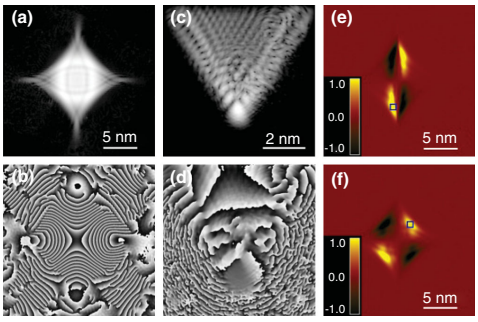

FIG. 3 (color). (a) Logarithm of the reconstructed intensity for the first image in the through focus series of astigmatic images. (b) Corresponding retrieved phase with vortices surrounding the caustic. (c) Logarithm of the reconstructed intensity for the first image in the through focus series of five coma images. (d) Corresponding retrieved phase. (e) The z component of the orbital angular momentum density for the retrieved astigmatism caustic wave function near one line focus, and (f) a defocus between the two line foci.

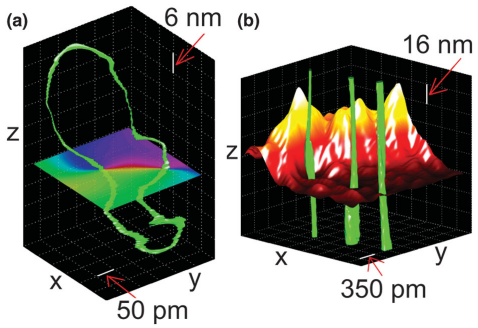

FIG. 4 (color). (a) Electron vortex loop observed in the interior of the coma caustic in Fig. 3, where the color-coded phase map shows a vortex-antivortex pair, which nucleates and annihilates at specific points along the optic z axis. (b) Vortex lines puncturing the image plane for the astigmatism caustic of Fig. 3(b); the color-coded part represents the intensity.

parallel experimental and theoretical findings in light optics and thereby raise the possibility of creating topologically knotted electron waves [53]. Using experimentally retrieved electron wave functions, we have also mapped the OAM density per electron for astigmatic wave fields.

T. C. P. thanks the Monash University electron scattering group for useful comments. The authors acknowledge use of facilities within the Monash Centre for Electron Microscopy. This research used equipment funded by the Australian Research Council Grant No. LE0454166. Funding from the Australian Research Council Discovery Project Grant No. DP1092745 is also gratefully acknowledged.

*Corresponding author.
timothy.petersen@monash.edu

[1] L. J. Allen, H. M. L. Faulkner, K. A. Nugent, M. P. Oxley, and D. Paganin, Phys. Rev. E **63**, 037602 (2001).
[2] L. J. Allen, H. M. L. Faulkner, M. P. Oxley, and D. Paganin, Ultramicroscopy **88**, 85 (2001).
[3] M. Uchida and A. Tonomura, Nature (London) **464**, 737 (2010).
[4] J. Masajada and B. Dubik, Opt. Commun. **198**, 21 (2001).
[5] V. Ravikumar, R. P. Rodrigues, and V. P. Dravid, J. Am. Ceram. Soc. **80**, 1117 (1997).
[6] K. W. Nicholls and J. F. Nye, J. Phys. A **20**, 4673 (1987).
[7] R. Vincent, W. J. Vine, P. A. Midgley, P. Spellward, and J. W. Steeds, Ultramicroscopy **50**, 365 (1993).
[8] M. Terauchi, K. Tsuda, O. Kamimura, M. Tanaka, T. Kaneyama, and T. Honda, Ultramicroscopy **54**, 268 (1994).
[9] A. F. Moodie, J. Etheridge, and C. J. Humphreys, Acta Crystallogr. Sect. A **52**, 596 (1996).
[10] P. N. H. Nakashima, A. F. Moodie, and J. Etheridge, Acta Crystallogr. Sect. A **63**, 387 (2007).
[11] J. Verbeeck, H. Tian, and P. Schattschneider, Nature (London) **467**, 301 (2010).
[12] B. J. McMorran, A. Agrawal, I. M. Anderson, A. A. Herzing, H. J. Lezec, J. J. McClelland, and J. Unguris, Science **331**, 192 (2011).
[13] J. Verbeeck, H. Tian, and A. Beche, Ultramicroscopy **113**, 83 (2012).
[14] J. Verbeeck, P. Schattschneider, S. Lazar, M. Stoger-Pollach, S. Loffler, A. Steiger-Thirsfeld, and G. Van Tendeloo, Appl. Phys. Lett. **99**, 203109 (2011).
[15] M. V. Berry, J. Opt. A **11**, 094001 (2009).
[16] J. Courtial, K. Dholakia, L. Allen, and M. J. Padgett, Opt. Commun. **144**, 210 (1997).
[17] M. V. Berry, J. F. Nye, and F. J. Wright, Phil. Trans. R. Soc. A **291**, 453 (1979).
[18] M. S. Soskin and M. V. Vasnetsov, in *Progress in Optics*, edited by E. Wolf (Elsevier, Amsterdam, 2001), p. 219.
[19] M. R. Dennis, K. O'Holleran, and M. J. Padgett, in *Progress in Optics*, edited by E. Wolf (Elsevier, Amsterdam, 2009), p. 293.
[20] A. G. Peele, K. A. Nugent, A. P. Mancuso, D. Paterson, I. McNulty, and J. P. Hayes, J. Opt. Soc. Am. A **21**, 1575 (2004).
[21] G. W. Farnell, Can. J. Phys. **36**, 935 (1958).
[22] B. T. Hefner and P. L. Marston, J. Acoust. Soc. Am. **106**, 3313 (1999).
[23] M. V. Berry, Adv. Phys. **25**, 1 (1976).
[24] J. F. Nye, *Natural Focusing and Fine Structure of Light: Caustics and Wave Dislocations* (Institute of Physics Publishing, Bristol, 1999).
[25] A. Boivin, J. Dow, and E. Wolf, J. Opt. Soc. Am. **57**, 1171 (1967).
[26] K. M. Pavlov *et al.*, Phys. Rev. A **83**, 013813 (2011).
[27] G. Ruben and D. M. Paganin, Phys. Rev. E **75**, 066613 (2007).
[28] D. R. Scherer, C. N. Weiler, T. W. Neely, and B. P. Anderson, Phys. Rev. Lett. **98**, 110402 (2007).
[29] G. Ruben, D. M. Paganin, and M. J. Morgan, Phys. Rev. A **78**, 013631 (2008).
[30] B. R. A. Nijboer, Physica (Amsterdam) **13**, 605 (1947).
[31] K. Nienhuis and B. R. A. Nijboer, Physica (Amsterdam) **14**, 590 (1949).
[32] E. Wolf, Rep. Prog. Phys. **14**, 95 (1951).
[33] P. W. Hawkes and E. Kasper, *Principles of Electron Optics* (Academic Press, London, 1996), Vol. 3.
[34] M. V. Berry and C. Upstill, in *Progress in Optics*, edited by E. Wolf (Elsevier, New York, 1980), p. 257.
[35] M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, Cambridge, England, 1999), Seventh (Expanded) Edition.
[36] R. Kingslake, Trans. Opt. Soc. **27**, 94 (1925).
[37] D. M. Paganin, *Coherent X-ray Optics* (Oxford University Press, New York, 2006).
[38] C. Dwyer, S. Lazar, L. Y. Chang, and J. Etheridge, Acta Crystallogr. Sect. A **68**, 196 (2012).
[39] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.110.033901 for examples of three measured coma caustics, coma intensity zeros, explicit details of phase retrieval parameters, retrieved electron vortices and transverse current densities, and plots of vortex tracks in experimentally retrieved wave functions.
[40] R. Kingslake, Proc. Phys. Soc. London **61**, 147 (1948).
[41] R. W. Gerchberg and W. O. Saxton, Optik (Stuttgart) **35**, 237 (1972).
[42] D. L. Misell, J. Phys. D **6**, 2217 (1973).
[43] L. J. Allen and M. P. Oxley, Opt. Commun. **199**, 65 (2001).
[44] L. J. Allen, W. McBride, N. L. O'Leary, and M. P. Oxley, Ultramicroscopy **100**, 91 (2004).
[45] C. T. Koch, Ultramicroscopy **108**, 141 (2008).
[46] D. R. G. Mitchell, Microsc. Res. Tech. **71**, 588 (2008).
[47] D. R. G. Mitchell and B. Schaffer, Ultramicroscopy **103**, 319 (2005).
[48] R. A. Sayle and E. J. Milnerwhite, Trends Biochem. Sci. **20**, 374 (1995).
[49] K. O'Holleran, M. J. Padgett, and M. R. Dennis, Opt. Express **14**, 3039 (2006).
[50] K. O'Holleran, M. R. Dennis, F. Flossmann, and M. J. Padgett, Phys. Rev. Lett. **100**, 053902 (2008).
[51] J. F. Nye, J. Opt. A **8**, 304 (2006).
[52] S. Crow, AIAA J. **8**, 2172 (1970).
[53] M. R. Dennis, R. P. King, B. Jack, K. O'Holleran, and M. J. Padgett, Nat. Phys. **6**, 118 (2010).

PHYSICAL REVIEW A **88**, 043803 (2013)

Measurement of the Gouy phase anomaly for electron waves

T. C. Petersen,[1,2,*] D. M. Paganin,[1] M. Weyland,[2,3] T. P. Simula,[1] S. A. Eastwood,[1] and M. J. Morgan[1]

[1]*School of Physics, Monash University, Victoria 3800, Australia*

[2]*Monash Centre for Electron Microscopy, Monash University, Victoria 3800, Australia*

[3]*Department of Materials Engineering, Monash University, Victoria 3800, Australia*

(Received 24 May 2013; published 4 October 2013)

We measure the Gouy phase anomaly for matter waves using in-line holography to retrieve the full complex field of an astigmatic electron wave function. Sequential phase shifts of $\pi/2$ rad are observed for electron trajectories along the optic axis that pass through each line-focus caustic of subnanometer transverse width. Our observations demonstrate that anomalous phase shifts of matter waves in the vicinity of caustics can be robustly measured using phase retrieval, extending the current scope of singular electron optics.

DOI: 10.1103/PhysRevA.88.043803 PACS number(s): 42.25.−p, 41.85.−p, 42.30.Rx, 42.50.Xa

I. INTRODUCTION

The Gouy phase anomaly [1], which describes the additional phase shift accumulated by a wave packet upon focusing, has been of fundamental interest in light optics for more than a century, and the diverse literature on this phenomenon continues to grow. Understanding, measuring, and ultimately exploiting the Gouy phase in a variety of experimental contexts is crucial for the development of particular optical technologies. For example, the rotation of the Poynting vector in Laguerre-Gauss beams is proportional to the Gouy phase [2]. Applications that exploit the phase anomaly include visible-wavelength super-resolution [3], subwavelength terahertz (THz) imaging [4], ultrashort laser pulses [5], single nanoparticle interferometry [6], and astigmatic mode conversion [7]. Use of the Gouy phase in this latter situation has been demonstrated for vortical electron beams [8]. In light optics, the rotation of topological defects in Laguerre-Gauss beams [9] has been used as a means to study Gouy phase effects. Such rotations were recently measured and were demonstrated for matter waves in a transmission electron microscope (TEM) [10].

Gouy's original observations were made using mirrors and white-light interferometry [1,11]. Visible-light lasers have since been used to measure the effect [12]. Measurements of the Gouy phase for cylindrically focused waves were reported in visible-light optics [13] where the anomaly was generalized for astigmatic wave fields [14]. Specifically, Visser and Wolf [14] derived the following expression for the on-axis Gouy phase anomaly $\delta(u)$ for a scalar wave diffracted from an aperture of size a, focused by an astigmatic lens of focal length f with coefficient of astigmatism A_0,

$$\delta(u) = \arg\left[\int_0^1 e^{i[(kA_0-u)\rho^2/2]}\rho J_0(kA_0\rho^2/2)d\rho\right] - \frac{\pi}{2}, \tag{1}$$

where J_0 is a zeroth-order Bessel function of the first kind and the integration variable ρ is dimensionless. The parameter u is proportional to the distance z along the optic axis according to $u = 2\pi(a/f)^2 z/\lambda$, where λ is the wavelength and k is the wave number. Equation (1) describes two sequential Gouy phase shifts of $\pi/2$ rad along the optic axis, associated with a pair of mutually orthogonal line foci, separated by $\Delta u = 2kA_0$ (see Fig. 1).

Refracting objects can act as natural lenses and inherently can give rise to Gouy phase shifts. For example, near the focal point of a light-scattering microsphere, the phase anomaly has been retrieved using a modified Hartman wave-front sensor [15]. The Gouy phase has also been measured in other contexts, such as the local expansion of adjacent intensity minima in standing microwaves [16], time-resolved THz pulses [17], and the cylindrical focusing of phonon-polariton wave packets in Raman scattering [18].

Recently, experiments were proposed for measuring the Gouy phase in matter waves, such as coherent atomic beams, using cylindrical focusing of Rydberg atoms [19]. Inspired by this proposal, we have measured the Gouy phase for astigmatic electron matter waves using phase retrieval [20,21]. Aberration correction lenses in a TEM were used to induce astigmatic pairs of line foci with transverse cross sections narrower than 1 nm, separated by more than 1 μm along the longitudinal optic axis. Through-focal series of images were used to retrieve the wave-function phase; the Gouy anomaly through each line focus was measured by propagating the retrieved electron wave function.

II. EXPERIMENT

Figure 1 shows a schematic of the experiment where electron matter waves diffracted by a circular aperture are focused by an astigmatic lens, which produces a caustic volume containing a pair of mutually orthogonal line foci at different points along the optic axis in the vicinity of the backfocal plane of the aberrated lens.

With a small 10-μm condenser aperture, a thin disordered carbon specimen was used to correct the aberrations in the imaging lens on a Titan^3 80-300 (FEI) aberration-corrected (CEOS GmbH) TEM, operating at 300 kV. Without fully correcting aberrations in the probe-forming lenses, the condenser system was configured to produce a small yet parallel probe, and the first condenser lens was adjusted to about half the maximum nominal setting. Airy rings were observed upon convergence of the electron probe in the imaging plane after which the probe-corrector stigmator coils were then grossly

*Corresponding author: timothy.petersen@monash.edu

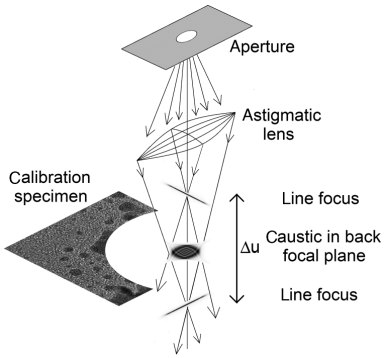

FIG. 1. Schematic of electron matter waves diffracted by a circular aperture and focused by a lens with astigmatism. The Gouy phase anomaly describes the variation in the electron wave phase along the longitudinal (vertical) optic axis, through the transverse center of the aberrated focal volume as compared to the linear variation of phase predicted by geometric optics. A thin specimen is used to correct imaging lens aberrations with a parallel probe, prior to the imposition of astigmatism in the illumination and focusing of the probe in the specimen plane. The parameter u is proportional to the distance along the optic axis as described in the paragraph containing Eq. (1).

excited to produce an astigmatic line focus of subnanometer width. Defocusing of the imaging objective lens revealed another mutually orthogonal line focus about 1.5 μm farther along the optic axis. A micron-sized hole in the specimen was found so that images of electron probe cross sections could be observed in the absence of scattering. Using 100 s acquisition times, a focal series of 12 images was collected at points along the optic axis between the two line foci, using nominal defocus increments of 60 nm.

The electron phase was retrieved using the Gerchberg-Saxton-Misell algorithm [20,21], which propagates a paraxial wave between focal planes and replaces wave intensities with those measured in experiment. Starting with an initial phase estimate, iterations of this process can retrieve the wave function with transverse probability distributions that are consistent with experimental observations. Although alternative approaches exist, this algorithm was employed to reliably handle intensity zeros surrounding caustic cross sections as well as possible phase vortices [22,23], which arise in the focal volumes of lenses with aberrations [24,25]. To increase the signal-to-noise ratio, all images were down-sampled from 2048×2048 pixels to 256×256 pixels. For improved robustness to noise, propagated waves were averaged in a fixed plane with each iteration using the approach of Allen *et al.* [23], but excluding the effects of partial coherence. Despite these measures and the long exposure times used to acquire the experimental intensities, the propagated intensities failed to converge towards experiment and the phase-retrieval algorithm stagnated. Some features, such as the caustic shape and blurred line foci, were retrieved; yet after thousands of iterations, the consistency of the wave function intensities with experiment was not satisfactory, even on a qualitative level. To deal with this, we approximated the caustic using a diffraction integral to compute the astigmatic wave function, which was then used to seed the initial phase in the retrieval iterations. Furthermore, the retrieval algorithm was modified to exclude the fitting of intensities far from the caustic where the experimental data contained only noise and were devoid of phase information. This was achieved in a systematic manner by iteratively replacing only those pixels where the experimental intensity was above one standard deviation of the intensity, measured in a region far from the caustic. Satisfactory convergence of propagated intensities towards the experimental input was then achieved after 10^3 iterations. To improve the spatial resolution of diffraction detail within the caustic, the retrieved wave function was then resampled to 1024×1024 pixels and was used to seed a further 10^3 iterations with the experimental data down-sampled to the same number of pixels.

III. ANALYSIS

In principle, any in-line holographic method can be used to reconstruct a desired monochromatic scalar wave function, provided that the solution reproduces the experimental data when propagated between focal planes. Allen *et al.* [23] define a sum-squared error (SSE) to characterize the convergence for iterative phase retrieval, which has the form $\mathrm{SSE}_j = \Sigma\Sigma(\sqrt{I_{\mathrm{exp}}} - \sqrt{I_{\mathrm{rec}}})^2/\Sigma\Sigma I_{\mathrm{exp}}$, where the double sums are over all pixels in the jth experimental intensity I_{exp} and reconstructed intensity I_{rec}. The average of SSE_j over all N images in the focal series defines an averaged sum-square error: $\mathrm{SSE}_{\mathrm{av}} = \Sigma \mathrm{SSE}_j/N$. Over the entire field of view for all 12 images in the focal series, the caustic reconstruction converged to $\mathrm{SSE}_{\mathrm{av}} = 5.6 \times 10^{-2}$. For all pixels with intensity above the threshold used for iterative replacement, the error metric was measured to be $\mathrm{SSE}_{\mathrm{av}} = 1.4 \times 10^{-2}$.

Figure 2 shows the phase-retrieval results for the astigmatic caustic. Figure 2(a) shows the retrieved phase of the propagated wave function at a focal distance in-between the two line

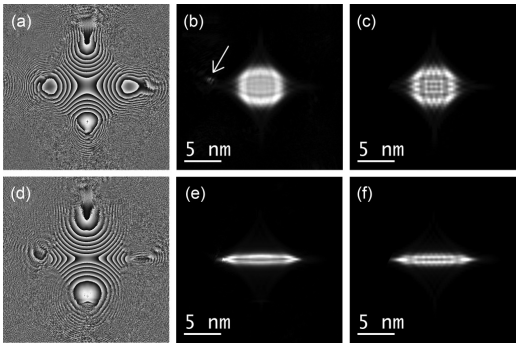

FIG. 2. (a) Retrieved electron phase map at a focal distance in-between the two astigmatic focal lines. (b) Corresponding propagated intensity. (c) Experimental intensity distribution at the same focal distance. (d) Retrieved electron phase map near one of the astigmatic focal lines. (e) Corresponding propagated intensity. (f) Experimental intensity distribution at the same focal distance. The arrow in (b) highlights an outlying discrepancy in the retrieved probability density, discussed in the main text.

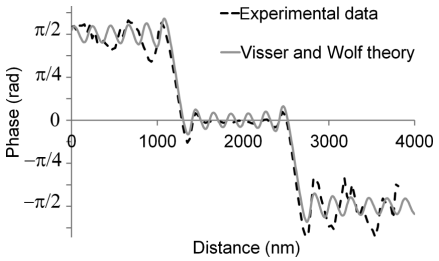

FIG. 3. Experimentally retrieved on-axis phase (dashed curve) for a focused astigmatic matter-wave field compared with the Gouy anomaly theory predicted by Visser and Wolf [14]. Two sequential phase shifts of $\pi/2$ rad occur at each line focus due to the Gouy anomaly. Parameters required for the theory given by Eq. (1) were measured from the experimentally retrieved intensity distributions alone.

foci with the corresponding retrieved intensity in Fig. 2(b), which compares favorably with the experimental data in Fig. 2(c). Figure 2(d) shows the retrieved phase near one of the line foci with the associated retrieved intensity in Fig. 2(e) compared to the experimental data in Fig. 2(f). Close inspection of Fig. 2(f) indicates that the left cusp is clipped since this region was outside the field of view of the CCD used to capture the unprocessed experimental data. This explains the outlying discrepancy labeled in Fig. 2(b), which arose because the intensity in this region was unconstrained and the clipping of the experimental data is an unphysical artifact. Figures 2(e) and 2(f) also show good agreement, thereby demonstrating satisfactory convergence of the iterative phase-retrieval algorithm, which is consistent with the small value of $\mathrm{SSE}_{\mathrm{av}}$.

Using the retrieved wave function, the centers of the phase maps were tracked by propagating the electron wave in steps of 30 nm along the optic axis. The positions of the sharp line foci in the transverse intensity distributions were used to accurately determine the optic axis position, along which the retrieved on-axis phase was then plotted, as shown by the dashed profile in Fig. 3. The solid-line profile in Fig. 3 was calculated from Eq. (1). Apart from the electron wavelength, three additional parameters were required for the computation, namely, the effective aperture size to focal length ratio a/f, the coefficient of astigmatism A_0, and the position of the line foci. All of these parameters were robustly determined from the intensities of the experimentally retrieved wave function by computing the Heisenberg uncertainties $\Delta x(z)$ and $\Delta y(z)$. Away from the line foci, $\Delta x(z)$ and $\Delta y(z)$ behave asymptotically as $T|z|$, where T is a constant representing the gradient of the uncertainty as a function of the distance z. Each measured value of T was averaged to determine A_0 using Eq. (1) with $a/f \cong 2T$ [14]. Both phase profiles contain an arbitrary vertical offset, which we have systematically chosen so that the on-axis variations approach zero in-between the two line foci.

The experimental on-axis phase profile in Fig. 3 follows the theory of Visser and Wolf [14] closely; in particular, the slopes and horizontal positions of the rapid phase variations near each line focus match. Some differences are evident between the undulations in Fig. 3, which are sensitive to the effects of diffraction. These discrepancies could be ascribed to systematic errors in the phase retrieval. However, we expect differences on account of the fact that large excitation of the TEM stigmator coils does not produce pure twofold astigmatism but, rather, an astigmatic beam with a pair of line foci, which is also perturbed by coma and higher-order aberrations.

In addition to the scalar diffraction theory of focused paraxial waves, a variety of interpretations exist for the Gouy anomaly. One persistent theme is the idea that the Gouy effect arises from fluctuations in the transverse momentum, induced by variations in the uncertainty of the beam at different focal points along the optic axis [26–28]. For Gaussian beams, the variation in the standard deviation is characterized by the evolution of the beam waist, and the transverse intensity distribution maintains the same shape at different points along the optic axis [11]. Accordingly, the Gouy phase evolves along the optic axis, varying most rapidly near the focus and more abruptly for smaller beam waists. Borghi *et al.* [29] have theoretically demonstrated a "universal form" of this Gouy phase variation for a wide class of such shape-invariant beams. The caustic shapes and standard deviations of other beams, such as those formed by astigmatic lenses, are not shape invariant and are nontrivial with interference phenomena more readily described by catastrophe optics [30–32]. We performed another experiment to demonstrate the three-dimensional nature of this standard deviation, organized by the diffraction detail within an astigmatic volume. Similar illumination conditions were chosen, with the exception of the field-emission gun lens, the strength of which was halved, thereby increasing the probe intensity to reduce the Poisson noise in recorded images. Thirty-two images were then acquired using 100-s exposures and nominal defocus increments of 80 nm. These images were stacked to create a tomogram of the matter-wave probability density. Figure 4 shows a false-color isosurface of

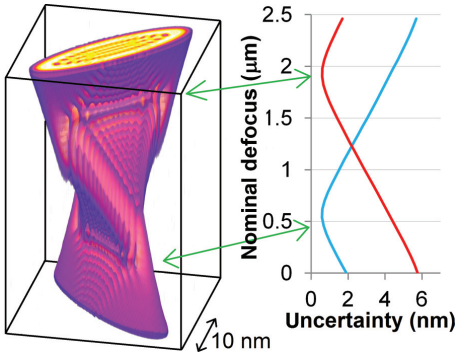

FIG. 4. (Color) Electron probability density measured from a through-focus series of 32 images, which were stacked along the vertical direction and were interpolated. The uncertainties along two transverse directions, orthogonal to each line focus, were calculated from each image. The paraxial electron wave evolves slowly along the optic axis as evidenced by the 10-nm transverse scale bar, which is to be contrasted with the micron-scale focal range along the vertical axis.

the tomogram, which was rendered using the VOLUME VIEWER plug-in for IMAGEJ [33,34]. The accompanying Heisenberg uncertainties were measured along directions orthogonal to the line foci and fall to a minimum at each line focus. These plots and other image-processing operations were performed using the scientific data analysis package DIFFTOOLS [35].

Diffraction effects, due to the finite electron wavelength and the uncertainty principle, prevent the uncertainties from reaching zero in Fig. 4 where the Gouy anomaly varies most rapidly. The entire tomogram of the electron probability density can be viewed as a richly detailed three-dimensional in-line hologram. Several transverse sections of such a volume, which encode the holographic information, enabled the complex field of the matter wave to be retrieved to yield the Gouy anomalies shown in Fig. 3.

IV. CONCLUSION

To summarize, we have measured the Gouy phase anomaly for astigmatic matter waves using electron wave-function phases inferred from experimental intensities. Successive Gouy phase shifts of $\pi/2$ rad were observed for fast electrons traveling along the optic axis, passing through two subnanometer lines of focus. These longitudinal phase variations were compared with wave optics theory, and consistency with experiment was demonstrated. With singular electron optics in its infancy, opportunities exist to extend these findings and to explore quantized phase changes of matter waves in more general settings, such as discrete phase changes for rays that touch caustics, which are characterized by Maslov indices [36,37].

ACKNOWLEDGMENTS

S. Kennedy and D. Jesson from the Monash University School of Physics are thanked for their useful input and stimulation of ideas. The authors acknowledge use of the facilities within the Monash Centre for Electron Microscopy. This research used equipment funded by the Australian Research Council Grant No. LE0454166. Funding from the Australian Research Council Discovery Project Grant No. DP1092745 is also gratefully acknowledged. T. C. Petersen is grateful for financial support from J. Etheridge of the Monash Centre for Electron Microscopy.

[1] L. G. Gouy, Ann. Chim. Phys. **24**, 145 (1891).
[2] M. J. Padgett and L. Allen, Opt. Commun. **121**, 36 (1995).
[3] A. Whiting, A. Abouraddy, B. Saleh, M. Teich, and J. Fourkas, Opt. Express **11**, 1714 (2003).
[4] M. Yi, K. Lee, J. D. Song, and J. Ahn, Appl. Phys. Lett. **100**, 161110 (2012).
[5] F. Lindner, G. G. Paulus, H. Walther, A. Baltuška, E. Goulielmakis, M. Lezius, and F. Krausz, Phys. Rev. Lett. **92**, 113001 (2004).
[6] J. Hwang and W. E. Moerner, Opt. Commun. **280**, 487 (2007).
[7] M. W. Beijersbergen, L. Allen, H. Vanderveen, and J. P. Woerdman, Opt. Commun. **96**, 123 (1993).
[8] P. Schattschneider, M. Stöger-Pollach, and J. Verbeeck, Phys. Rev. Lett. **109**, 084801 (2012).
[9] J. Hamazaki, Y. Mineta, K. Oka, and R. Morita, Opt. Express **14**, 8382 (2006).
[10] G. Guzzinati, P. Schattschneider, K. Y. Bliokh, F. Nori, and J. Verbeeck, Phys. Rev. Lett. **110**, 093601 (2013).
[11] A. E. Siegman, *Lasers* (University Science Books, Saulsalito, CA, 1986).
[12] H. C. Kandpal, S. Raman, and R. Mehrotra, Opt. Lasers Eng. **45**, 249 (2007).
[13] J. P. Rolland, K. P. Thompson, K. S. Lee, J. Tamkin, T. Schmid, and E. Wolf, Appl. Opt. **51**, 2902 (2012).
[14] T. D. Visser and E. Wolf, Opt. Commun. **283**, 3371 (2010).
[15] P. Bon, B. Rolly, N. Bonod, J. Wenger, B. Stout, S. Monneret, and H. Rigneault, Opt. Lett. **37**, 3531 (2012).
[16] C. R. Carpenter, Am. J. Phys. **27**, 98 (1959).
[17] A. B. Ruffin, J. V. Rudd, J. F. Whitaker, S. Feng, and H. G. Winful, Phys. Rev. Lett. **83**, 3410 (1999).
[18] T. Feurer, N. S. Stoyanov, D. W. Ward, and K. A. Nelson, Phys. Rev. Lett. **88**, 257402 (2002).
[19] I. G. da Paz, P. L. Saldanha, M. C. Nemes, and J. G. P. de Faria, New J. Phys. **13**, 125005 (2011).
[20] R. W. Gerchberg and W. O. Saxton, Optik (Jena) **35**, 237 (1972).
[21] D. L. Misell, J. Phys. D: Appl. Phys. **6**, 2200 (1973).
[22] L. J. Allen, H. M. L. Faulkner, K. A. Nugent, M. P. Oxley, and D. Paganin, Phys. Rev. E **63**, 037602 (2001).
[23] L. J. Allen, W. McBride, N. L. O'Leary, and M. P. Oxley, Ultramicroscopy **100**, 91 (2004).
[24] A. Boivin, J. Dow, and E. Wolf, J. Opt. Soc. Am. **57**, 1171 (1967).
[25] K. M. Pavlov, D. M. Paganin, D. J. Vine, J. A. Schmalz, Y. Suzuki, K. Uesugi, A. Takeuchi, N. Yagi, A. Kharchenko, G. Blaj, J. Jakubek, M. Altissimo, and J. N. Clark, Phys. Rev. A **83**, 013813 (2011).
[26] R. W. Boyd, J. Opt. Soc. Am. **70**, 877 (1980).
[27] P. Hariharan and P. A. Robinson, J. Mod. Opt. **43**, 219 (1996).
[28] S. M. Feng and H. G. Winful, Opt. Lett. **26**, 485 (2001).
[29] R. Borghi, M. Santarsiero, and R. Simon, J. Opt. Soc. Am. A **21**, 572 (2004).
[30] T. C. Petersen, M. Weyland, D. M. Paganin, T. P. Simula, S. A. Eastwood, and M. J. Morgan, Phys. Rev. Lett. **110**, 033901 (2013).
[31] M. V. Berry and C. Upstill, in *Progress in Optics*, edited by E. Wolf (Elsevier, New York, 1980), Vol. 18, p. 257.
[32] R. Borghi, Phys. Rev. E **85**, 046704 (2012).
[33] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, Nat. Methods **9**, 671 (2012).
[34] K. U. Barthel, http://rsbweb.nih.gov/ij/plugins/volume-viewer.html, Berlin, Germany (2012).
[35] D. R. G. Mitchell, Microsc. Res. Tech. **71**, 588 (2008).
[36] J. B. Keller, SIAM Rev. **27**, 485 (1985).
[37] Y. I. Orlov, Radiophys. Quantum Electron. **24**, 154 (1981).

PHYSICAL REVIEW A **89**, 063801 (2014)

Unifying interpretations of the Gouy phase anomaly for electron waves

T. C. Petersen,[1,2,*] D. M. Paganin,[1] M. Weyland,[2,3] T. P. Simula,[1] S. A. Eastwood,[1] and M. J. Morgan[1]
[1]*School of Physics, Monash University, Victoria 3800, Australia*
[2]*Monash Centre for Electron Microscopy, Monash University, Victoria 3800, Australia*
[3]*Department of Materials Engineering, Monash University, Victoria 3800, Australia*
(Received 13 October 2013; published 3 June 2014)

We provide a unified description of the Gouy phase anomaly, highlighting the pivotal role of local fluctuations in transverse momentum and quantal phase changes associated with electron trajectories that touch caustics. Using arguments based on wave optics, at the interface between classical and quantum mechanics, we derive the magnitude of the quantized Gouy phase changes for scalar waves. Our analysis unifies disparate descriptions of the Gouy anomaly within a single theoretical framework. In particular, we find that the phase anomaly connects Maslov indices, quantal Berry phases, Heisenberg momentum fluctuations, and confinement-induced Lévy-Leblond phase shifts.

DOI: 10.1103/PhysRevA.89.063801 PACS number(s): 42.25.−p, 03.65.Vf, 42.50.Xa, 68.37.Lp

I. INTRODUCTION

For over 100 years, Gouy's quantized phase changes incurred by focused rays [1] have resisted intuitive interpretation with many accurate, yet contrasting, theoretical descriptions detailed in the literature. The phase anomaly has been derived from diffraction integrals describing aberrations in scalar light optics [2,3]. For converging spherical waves diffracted by a circular aperture, the total phase excursion along the optic axis differs from that predicted in ray optics by π rad [2]. For converging astigmatic waves, a phase anomaly of $\pi/2$ rad is incurred for each line focus [3]. By focusing annular-shaped and collimated laser light, Gouy phase variations have been found to occur in the focal region where diffraction is suppressed in such beams [4]. Gouy phase variations have also been defined and analytically derived for nondiffracting Airy beams [5]. Using the vector model of electromagnetic waves, focused light in high numerical aperture systems has been studied where three components of the electric field exhibit different phase anomalies [6]. Nonparaxial and radially polarized beams have been investigated where noninteger and even irrational Gouy phases were identified as arising from discontinuities in the associated pupil function [7]. Substantial polarization variations in the focal volume of beams with radial polarization have been attributed to the different Gouy phases of the electric-field components [8]. For partially coherent waves, the spectral degree of coherence also exhibits the Gouy phase [9]. Similarly, for partially coherent x rays, an on-axis anomaly occurs in the phase of the cross-spectral density, which measures the evolution of coherence in terms of correlation functions [10].

Analytical formulations in light optics have been developed for Gaussian beams [11] where the Gouy phase follows an inverse tangent function of a scaled distance along the optic axis. For beams that propagate in free space without changing shape, this characteristic inverse tangent phase variation has been derived from wave optics as a "universal form" [12]. From a quantum-mechanical point of view, a phase anomaly arises for Gaussian beams because the effective longitudinal momentum is reduced due to transverse confinement of the wave, which increases the spread of the transverse momentum [13]. Indeed, the Gouy phase has been associated with the uncertainty principle to describe changes in an effective propagation constant [14,15]. Gaussian beams have also been used to describe the Gouy anomaly [16] in terms of Berry's geometric phase [17,18].

Certain optical devices can collapse rays from all spatial directions towards a common focal point, from which rays then diverge [19]. For such wide-angle lenses, a Gouy phase anomaly protects against an unphysical singularity since the superposition of converging and diverging spherical waves must differ in phase by π rad at the focal point in order to satisfy the wave equation [19]. Such phase shifts more generally arise on passing through a caustic where singularities occur in geometric optics as rays overlap [20,21] and are characterized by Maslov indices [21,22].

Visible light images of high-order Gaussian beams partially obstructed by an aperture have been observed experimentally to undergo rotations upon propagation, which can be quantified directly by the Gouy effect [23], revealing the anomaly without explicit interferometry. Recently, such Gouy rotations of matter waves were measured in a transmission electron microscope (TEM) by studying electron vortex probes, which were created with micron-sized holographic masks and were cut by a knife edge [24]. The Gouy phase for astigmatic matter waves has also been used to induce mode conversion of electron vortex beams, which was experimentally demonstrated in the TEM [25]. In visible light optics, astigmatic phase anomalies of $\pi/2$ rad recently were measured using a modified Mertz-Sagnac interferometer, quantitatively confirming theoretical predictions in wave optics [3]. Using in-line electron holography, we recently measured the Gouy phase for astigmatic electron waves and compared the results to the same theory [26]. Our measurements motivated the current paper, which was prompted by a desire to ascribe our experimental observations to the rich variety of physical interpretations detailed in the literature. For fast electrons obeying the Schrödinger equation, we sought to unify these different points of view to better understand the Gouy effect and ultimately identify new ways in which to measure and exploit the phenomenon.

In this paper, we provide a theoretical analysis of the Gouy effect to describe quantized phase changes within the context of focused scalar waves. We begin with a geometric

*timothy.petersen@monash.edu.au

construction to visualize the standard deviation of the transverse probability density of a focused beam and its evolution along the optic axis. Using this model, we show that the intuitive notion of the Gouy phase as being induced by second moments of transverse momentum, measured over the entire transverse plane, is incorrect. Examination of local transverse momentum fluctuations in a neighborhood containing the optic axis leads to quantitative derivations of the Gouy phase in terms of paraxial waves. Upon further scrutiny, we are able to connect several disparate interpretations of the Gouy phase phenomenon within a single theoretical construction. Specifically, for focused paraxial waves, we relate the anomaly to statistical fluctuations, a geometric phase, semiclassical phase changes, and statistical confinement. Although Gaussian beams are used for certain numerical evaluations, we show that the quantized on-axis phase variation is independent of the explicit form of the probability density.

II. THEORY

Figure 1 shows the transverse spread of rays in the vicinity of a line focus. The solid curves delineate the standard deviation in the transverse position $\Delta x(z)$, which varies along the longitudinal optic axis z. The gray dashed curves mark the asymptotes of the standard deviation $\Delta x(z) = \pm T|z|$, where $T = \tan(\theta)$ is a positive real constant. The curved portion of the focal region represents the effects of diffraction in which the dashed boundary, governed by geometric optics, is broadened to account for a finite wavelength, thereby avoiding an unphysical singularity for the line focus at $z = 0$.

For the distribution of ray trajectories defining the angular range θ, $T = \tan\theta$, and the mean transverse position $\langle x(z)\rangle$ is zero (enforcing symmetry about the z axis). The dashed lines define a quadratic variance $\langle x^2(z)\rangle = T^2z^2$. The angular brackets refer to averages over the entire transverse plane such that, for a given function $f(x)$ and probability density $P(x)$, $\langle f(x)\rangle = \int_{-\infty}^{+\infty} f(x)P(x)dx$. One simple way to account for nonzero variance at the line focus is to add a small positive constant ε so that $\langle x^2(z)\rangle = T^2z^2 + \varepsilon^2$. The expression for the standard deviation $\Delta x(z)$ in the transverse position is now given by

$$\Delta x(z) = \sqrt{T^2z^2 + \varepsilon^2}. \tag{1}$$

The spread in the transverse momentum p_x is constrained by the Heisenberg uncertainty principle $\Delta x\Delta p_x \geq \hbar/2$. In terms of the transverse component k_x of the wave vector $\mathbf{k} = (k_x, k_z)$, de Broglie's relation gives $\Delta k_x \geq 1/(2\,\Delta x)$. Under paraxial conditions, we have $k_x = k\theta$ so that the angular spread $\Delta\theta$ defines the spread in k_x and Eq. (1) becomes

$$k\Delta\theta = \Delta k_x \geqslant \frac{1}{2\sqrt{T^2z^2 + \varepsilon^2}}. \tag{2}$$

Along the optic axis, the z component k_z of $\mathbf{k}$ varies statistically on account of the variation in k_x and constant energy in the waves where $|\mathbf{k}|^2 = k^2 = k_x^2 + k_z^2$. Averaged over the transverse axis x, we can define the propagation constant $\langle k_z\rangle$ in terms of the dispersion in transverse momentum as

$$\langle k_z\rangle = \sqrt{k^2 - \langle k_x^2\rangle} \cong k - \frac{1}{2k}\langle k_x^2\rangle = k - \frac{1}{2k}(\Delta k_x)^2. \tag{3}$$

The phase accumulates along the optic axis in accordance with variations in k_z. The phase predicted by ray optics is the wave number k times the distance along the optic axis. Upon subtracting this contribution, we can define the average accumulated phase $\langle\phi(z)\rangle$ by

$$\begin{aligned}\langle\phi(z)\rangle &\overset{\text{def.}}{=} \int_{z_{\min}}^{z}\left\langle\frac{\partial\phi}{\partial z'}\right\rangle dz' - k\int_{z_{\min}}^{z} dz' \\ &= k\int_{z_{\min}}^{z} dz' - \frac{1}{2k}\int_{z_{\min}}^{z}\langle k_x^2\rangle dz' - k\int_{z_{\min}}^{z} dz' + C,\end{aligned} \tag{4}$$

where C is an arbitrary constant and $z_{\min}$ is the leftmost point on the optic axis.

Equations (1), (3), and (4) can be combined, whence,

$$\langle\phi(z)\rangle \geqslant -\frac{1}{8k}\int_{z_{\min}}^{z}\frac{1}{T^2z'^2 + \varepsilon^2}dz', \tag{5}$$

where the constant C has been omitted by fixing the average phase to be zero at the focus.

Taking the limit $z_{\min} \to -\infty$, the definite integral in Eq. (5) evaluates to give

$$\langle\phi(z)\rangle \geqslant -\frac{1}{8k\varepsilon T}\tan^{-1}(Tz/\varepsilon). \tag{6}$$

At the origin, $z = 0$, the standard deviation $\Delta x(z = 0)$ in the transverse position is defined to be ε. Assuming that the angular spread $\Delta\theta$ is approximately equal to the semiangle ($\theta \ll 1$), Eq. (2) provides bounds on the constants in Eq. (6), i.e.,

$$k\theta \approx k\ \tan\ \theta = kT \geqslant \frac{1}{2\varepsilon} \to \varepsilon kT \geqslant \frac{1}{2}. \tag{7}$$

Equation (6) then becomes

$$\langle\phi(z)\rangle \geqslant -\frac{1}{4}\tan^{-1}(Tz/\varepsilon). \tag{8}$$

In three dimensions, the same arguments produce an additional, but identical, term on the right-hand side of Eq. (5), which doubles the right-hand side of Eq. (8). On account of

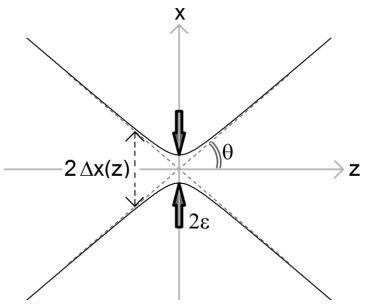

FIG. 1. Longitudinal variation in the standard deviation $\Delta x(z)$ along the optic axis near a focal point. The parameter ε accounts for the effects of diffraction and the minimum transverse spread in the electron beam. The minimum spread ε is proportional to the beam waist for Gaussian beams. For large z, the standard deviation $\Delta x(z)$ behaves asymptotically as $T|z|$, where $T = \tan\theta$.

this inverse tangent variation in z, it is tempting to directly associate $\langle\phi(z)\rangle$ with the Gouy phase anomaly. However, this bound for $\langle\phi(z)\rangle$ was derived from the minimum product in the uncertainty principle, which occurs at $z = 0$ in Fig. 1. The explicit value of $\langle\phi(z)\rangle$ for all z can be computed using, say, the Gaussian beam solution of the paraxial wave equation either directly or with expectation values of the transverse momentum through Eq. (4). In either case, we find $\langle\phi(z)\rangle = -z/(2z_R)$ in two transverse dimensions where $z_R = \pi\omega_0^2/\lambda$ is the Rayleigh range for a Gaussian beam with minimum beam waist ω_0 and wavelength λ [27]. Consequently, the transverse average of the phase is not equal to the Gouy phase anomaly. Generally, the linear variation in $\langle\phi(z)\rangle$ is expected from the paraxial approximation, which requires $\langle k_z^2\rangle = 0$. Hence, we must have $\langle k_z\rangle =$ constant so that, by analogy with Ehrenfest's theorem, the momentum (or energy) remains unchanged with respect to z in the absence of any forces.

We have shown that the Gouy phase ϕ_G is not equal to the integrated transverse average of the longitudinal phase gradient $\langle\phi(z)\rangle$. Using the same theoretical framework, we can inquire as to the longitudinal variation in the phase about the optic axis in the presence of statistical fluctuations. Within the paraxial approximation, one has the following three-dimensional form of Eq. (3):

$$\langle k_z\rangle = k - \frac{1}{2k}\langle k_x^2 + k_y^2\rangle. \tag{9}$$

Writing out the expectation value in full, we have

$$\langle k_z\rangle = k\int_{\mathbb{R}^2} A^2 dx\, dy + \frac{1}{2k}\int_{\mathbb{R}^2} Ae^{-i\phi}\nabla_\perp^2 Ae^{i\phi} dx\, dy, \tag{10}$$

where A is real and represents the amplitude of a paraxial electron wave $\xi = Ae^{i\phi}$.

The two-dimensional Laplacian operator $\nabla_\perp^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$ creates a mixture of first- and second-order partial derivatives in the amplitude A and phase ϕ. For a symmetric beam, all first-order partial derivatives are zero on the optic axis, so the integrand in Eq. (10) can be written as

$$Ae^{-i\phi}\nabla_\perp^2 Ae^{i\phi} = A\nabla_\perp^2 A + iA^2\nabla_\perp^2\phi. \tag{11}$$

The effective propagation constant can be complex with the imaginary part $\mathrm{Im}(\tilde{k}_z)$ describing the amplitude variation and the real part $\mathrm{Re}(\tilde{k}_z)$ giving the longitudinal phase derivative. With this in mind, we now consider the expectation values in Eq. (9) over an infinitesimal patch of the transverse plane and equate the integrands to obtain

$$\begin{aligned} A^2\tilde{k}_z &= A^2\frac{\partial\phi(z)}{\partial z} - iA^2\mathrm{Im}(\tilde{k}_z) = kA^2 + \frac{A^2}{2k}\frac{1}{\xi}\nabla_\perp^2\xi \\ &= kA^2 + \frac{1}{2k}A\nabla_\perp^2 A + i\frac{1}{2k}I\nabla_\perp^2\phi, \end{aligned} \tag{12}$$

where $\tilde{k}_z$ denotes a complex variable and Eq. (11) has been used to produce the final expression. It should be noted that the longitudinal phase gradient in Eq. (12) is consistent with the local momentum [28] $\mathrm{Re}(\xi^*\partial\xi/\partial z)/A^2$ of the paraxial envelope ξ for which we also have $\mathrm{Im}(\xi^*\partial\xi/\partial z)/A^2 = \ln(\partial A/\partial z)$ since transverse derivatives of ξ are zero on the optic axis for our symmetric beam. The real part of Eq. (12) gives the paraxial eikonal equation [29], and the imaginary part gives the transport of intensity equation [29] with $I = |A|^2$. Equating the real terms, we have

$$\frac{\partial\phi(z)}{\partial z} = k + \frac{1}{2k}\frac{1}{A}\nabla_\perp^2 A = k + \frac{1}{2k}\nabla_\perp^2\ln(A), \tag{13}$$

where the natural logarithm has been introduced by discarding first-order on-axis partial derivatives in amplitude A due to symmetry of the beam.

By choosing a Gaussian function to model an arbitrary amplitude distribution, we can show that Eq. (13) predicts the Gouy phase anomaly. Replacing the probability density $I = |A|^2$ with that of a paraxial Gaussian beam [11] with the same standard deviation, we obtain

$$\begin{aligned} \phi_G(z) &= \frac{\partial\phi(z)}{\partial z} - k \\ &= \frac{1}{2k}\nabla_\perp^2\ln\left\{\exp\left[-\frac{1}{4}(x^2+y^2)/(T^2z^2+\varepsilon^2)\right]\right\} \\ &= -\frac{1}{2k(T^2z^2+\varepsilon^2)}. \end{aligned} \tag{14}$$

Finally, integrating the longitudinal phase changes along z, we obtain the Gouy phase anomaly,

$$\begin{aligned} \int\left\{\frac{\partial\phi(z)}{\partial z} - k\right\}dz &= -\int\frac{1}{2k(T^2z^2+\varepsilon^2)}dz \\ &= -\frac{1}{2k\varepsilon T}\tan^{-1}\left(\frac{Tz}{\varepsilon}\right) = -\tan^{-1}\left(\frac{Tz}{\varepsilon}\right) \\ &= \phi_G(z), \end{aligned} \tag{15}$$

where $\varepsilon kT = \frac{1}{2}$ for a Gaussian distribution.

Therefore, we see that transverse local momentum fluctuations arising from the uncertainty principle produce the Gouy phase anomaly $\phi(x = 0, y = 0, z) = \phi_G(z)$. Using similar considerations, we can also interpret the effect in terms of the Lévy-Leblond [30] longitudinal phase shift due to statistical confinement in the transverse dimensions (see the Appendix).

Within the same theoretical framework, we can associate the phase anomaly with the quantized caustic phase changes described by Keller [21] as well as the Berry phase [17]. We begin with Keller's phase anomaly for rays that touch caustics, which are governed by Maslov indices [21]. Keller substitutes a Wentzel-Kramers-Brillouin wave form into the Schrödinger equation in terms of a phase with a classical action (in units of $\hbar$) and with an amplitude that can be complex and multivalued. The substitution leads to semiclassical corrections for asymptotic solutions with integer or half-integer quantum numbers, applicable for separable and nonseparable systems. Following Keller, we choose the classical action of a ray with momentum directed along the optic axis to be $S = kz$ and factor the wave function ψ as a product of this phase contribution with a complex amplitude ξ, which we will refer to as an "envelope" such that $\psi = \xi\exp(ikz)$. By demanding ψ to be single valued, the phase of ξ is found to change by $\pi/2$ rad along a trajectory touching a caustic [21]. For a closed curve trajectory which touches a caustic m times, the following circulation integral describes Keller's associated phase anomaly ϕ_K:

$$\phi_K = -i\oint_\Gamma \nabla\ln\xi\cdot d\mathbf{l} = -2\pi\frac{m}{4}, \tag{16}$$

where m is the Maslov index, $\mathbf{l}$ is a vector along the trajectory Γ, and ∇ is the three-dimensional gradient operator. We can associate Keller's anomaly ϕ_K with the Gouy phase ϕ_G by noting that the envelope ξ, by construction, obeys the paraxial equation $(\nabla_\perp^2 + 2ik\,\partial/\partial z)\xi = 0$ for large wave number k. If we integrate along the entire optic axis and assume paraxial conditions $|\partial^2\xi/\partial z^2| \ll k|\partial\xi/\partial z|$ [31], Eq. (16) can be rewritten as

$$\phi_K = -i \lim_{R\to\infty} \int_{-R}^{R} \frac{\partial \ln \xi}{\partial z} dz \equiv -i \oint_\Gamma \frac{\partial \ln \xi}{\partial \tilde{z}} d\tilde{z} = \frac{1}{2k} \oint_\Gamma \frac{1}{\xi} \nabla_\perp^2 \xi \; d\tilde{z}, \tag{17}$$

where the limit represents the Cauchy principal value at infinity [32] and we have extended the integral over the Argand plane for complex $\tilde{z}$.

Equation (17) is formally the same as Eq. (12) for the Gouy phase anomaly, deduced from fluctuations in transverse momentum since the imaginary part of ξ contributes nothing to the circulation due to conservation of probability density. However, the exchange of the definite integral for the closed line integral requires some explanation. We integrate along the real (optic) axis with a semicircle of radius R in the upper half of the complex plane joining both ends of the optic axis. It is assumed that the complex integral over the semicircle tends to zero as $R \to \infty$. As an example, consider the paraxial Gaussian beam once again, utilizing Eq. (11),

$$\phi_K = \frac{1}{2k} \oint \nabla_\perp^2 \ln\, A \; d\tilde{z} = \lim_{R\to\infty} \left\{ -\frac{1}{2k} \int_{-R}^{R} \frac{dz}{T^2 z^2 + \varepsilon^2} - \frac{1}{2k} \int_{C_R} \frac{d\tilde{z}}{T^2 \tilde{z}^2 + \varepsilon^2} \right\}, \tag{18}$$

where $\tilde{z}$ is complex, z is the real distance along the optic axis, and C_R is a semicircular contour in the upper half of the complex plane. The semicircle contour integral explicitly tends to zero [32], and Eq. (18) is then $2\pi i$ times the sum of the residues for the enclosed poles. The integrand contains two simple poles at $\tilde{z} = \pm i\varepsilon/T$. The pole in the upper half plane contributes to give the residue $1/(4\tilde{z}) = 1/(4i\varepsilon kT)$. Using $\varepsilon kT = 1/2$, for a Gaussian beam, the contour integral for the Keller anomaly is then a phase change of $\phi_K = -\pi$ rad. Half of this phase change occurs if the number of transverse dimensions is reduced to 1 since the Laplacian contribution reduces by a factor of 2. For the Maslov index m in Eq. (16), the trajectory touches the caustic twice ($m = 2$) for a pair of astigmatic line foci in two transverse dimensions and only once ($m = 1$) for one transverse dimension, which is consistent with the evaluation of the Gaussian beam using the residue theorem.

We can both simplify and generalize the analysis further if we reduce the essence of the Gouy anomaly to the mere existence of a quantized phase change without concern for the explicit functional form of this phase variation, which is accumulated as the wave field evolves along the optic axis. Equations (17) and (implicitly) (15) are proportional to the logarithmic derivative of the paraxial wave $\xi'(z)/\xi(z) = \partial_z \ln[\xi(z)]$ where the apostrophe and ∂_z refer to the longitudinal derivative with respect to z. Using the same contour of integration, we can evaluate the logarithmic derivative using Cauchy's argument principle [32] and can determine the scale of the phase shift without recourse to the explicit details of the Gaussian beam. Suppose, then, that the transverse probability distribution is Gaussian but the scale of the Gouy phase and phase variation in Eq. (15) is not specified such that $\phi_G(z) = \alpha \tan^{-1}(\beta z)$ with α and β unknown at present. For a given standard deviation $\Delta x(z) = \Delta y(z) = \sigma(z)$, conservation of probability requires that the beam amplitude varies in proportion to $1/\sigma(z)$ in two transverse spatial dimensions. Accordingly, the natural logarithm of $\xi(z)$ takes the following form on the optic axis:

$$\ln[\delta\xi(\beta z)] = \ln\left[\frac{\delta}{\sigma(\beta z)} e^{i\alpha\phi_G(\beta z)}\right] = \frac{1}{2}\ln\left[\frac{1}{\beta^2 T^2 z^2 + \varepsilon^2}\right] + \alpha \ln\left[\frac{1 + i\beta z}{1 + (\beta z)^2}\right] + \ln[\delta], \tag{19}$$

where α, β, and δ are all constants of proportionality, the factor of $\frac{1}{2}$ arises from the square root in Eq. (1), and the term multiplied by α derives from the inverse tangent form of $\phi_G(z)$. Cauchy's argument principle states that the closed contour integral of the logarithmic derivative in the complex plane is equal to $2\pi i(N - P)$ with the number of zeros N and poles P weighted by their respective multiplicities and orders [32]. Both N and P are trivially zero for the δ term in Eq. (19) (as is the derivative of this term). Similarly, N and P are both unity for the second term on the right in the upper half of the complex plane, so $\phi_G(z)$ does not contribute to the closed contour integral. The remaining term has no zeros and one simple pole in the upper half of the complex plane, hence, the closed contour integral evaluates to give a total Gouy phase shift of $\phi_G = -\frac{1}{2}i(2\pi i)(0 - 1) = -\pi$ rad where the prefactor of $-i$ in Eq. (17) has been included. In one transverse dimension, the beam amplitude varies as $[\sigma(z)]^{-1/2}$ in which case, the same analysis gives $\phi_G = -\pi/2$ rad for the total Gouy shift. Hence, the argument principle fixes the scale of the Gouy anomaly. These considerations stem from the fact that the wave-function envelope $\xi(z)$ is an analytic function, contains no zeros or poles along the closed contour, and there are only a finite number of poles within it since $\xi(\tilde{z})$ is meromorphic in the Argand plane for complex $\tilde{z}$. Moreover, the total Gouy phase shift is invariant to distortions of $\xi(\tilde{z})$ that preserve the topology in the complex plane. In other words, a phase change of π (or $\pi/2$) is guaranteed as long as no additional zeros or poles are introduced upon deforming $\xi(\tilde{z})$ to some non-Gaussian or nonsymmetric form. The argument principle cannot be used, for example, to describe on-axis phase changes for higher-order Gaussian modes. Such beams can contain intensity zeros and phase vortices on the optic axis. Shifting the closed contour slightly off axis serves little purpose for then, $\xi(\tilde{z})$ ceases to be meromorphic [32] in the complex plane. These facts are not surprising, given that we are trying to describe the phase change along the optic axis at points where the phase is not defined in these particular circumstances.

Berry [17] explained that, for quantum states which evolve slowly (adiabatically) and, therefore, remain in the same eigenstate at later times, there is an associated phase factor $\gamma(t)$ which also varies in time. Upon changing a parameter in the Hamiltonian cyclically, the wave function will return to

the same state at some later time; however, the accumulated phase will not generally match the initial value of $\gamma(t)$. In such cases, a measurable geometric phase arises. Let us associate slow adiabatic evolution with the paraxial equation for the envelope ξ where the fast oscillation of e^{ikz} has been factored out of the wave function ψ. In this context, evolution refers to propagation along the optic z axis since the paraxial equation is formally identical to the time-dependent Schrödinger equation in two spatial dimensions if z replaces the time parameter t. Following Berry [17], we can determine the evolution of the phase factor $\gamma(z)$ by substituting a formal solution for $\xi(z)$ [33], i.e.,

$$\xi(z) = e^{(iz/2k)\nabla_\perp^2}\xi(z=0), \tag{20}$$

into the paraxial wave equation,

$$\left(\nabla_\perp^2 + 2ik\frac{\partial}{\partial z}\right)e^{i\gamma(z)}e^{(iz/2k)\nabla_\perp^2}\xi(z=0) = 0, \tag{21}$$

which implies that

$$2ke^{i\gamma(z)}e^{(iz/2k)\nabla_\perp^2}\left(i\frac{\partial\xi}{\partial z} - \frac{\partial\gamma}{\partial z}\xi\right) = 0. \tag{22}$$

From Eq. (22), we deduce that

$$\begin{aligned}\phi_B &= \int_{-\infty}^{+\infty}\frac{\partial}{\partial z}\gamma(z)dz = -i\int_{-\infty}^{+\infty}\frac{1}{\xi}\frac{\partial\xi}{\partial z}dz \\ &= \frac{1}{2k}\int_{-\infty}^{+\infty}\frac{1}{\xi}\nabla_\perp^2\xi\, dz,\end{aligned} \tag{23}$$

where ϕ_B denotes the integration of phase changes described by $\gamma(z)$.

These integrals can be written as circulations if we integrate over the complex plane in the same manner as for the Keller phase anomaly since the results are formally identical. Note that $\gamma(z)$ is complex but ϕ_B is real if the integral in Eq. (23) extends over the entire z axis. Indeed, upon expressing Eq. (23) as a closed contour integral for complex z, we can express the phase in terms of a surface integral and thereby can emphasize the geometric aspect of the anomaly. Similar to Subbarao's construction based upon a Gaussian beam waist [16], we avoid Stokes' theorem and, instead, employ Green's theorem to compute areas which are then stereographically projected onto the Riemann sphere to define a solid angle over which the area subtends. However, our parameter space is the complex defocus, rather than the real and imaginary parts of the complex beam waist employed by Subbarao [16].

For complex defocus $\tilde{z}$, we can decompose an integration contour Γ in the complex plane as a succession of line segments Γ_{Re} and Γ_{Im}, which are parallel to either the real or the imaginary axis, respectively. Writing $\tilde{z} = z_R + iz_I$, Eq. (23) becomes

$$\begin{aligned}&-i\int_\Gamma\frac{\partial\ln(\xi)}{\partial\tilde{z}}dz \\ &= -i\left\{\int_{\Gamma_{\mathrm{Re}}}\frac{\partial\ln(\xi)}{\partial z_R}dz_R + \int_{\Gamma_{\mathrm{Im}}}\frac{\partial\ln(\xi)}{\partial z_I}dz_I\right\} \\ &= -i\int_\Gamma\left\{\frac{\partial\ln(\xi)}{\partial z_R} - i\frac{\partial\ln(\xi)}{\partial z_I}\right\}d\tilde{z}.\end{aligned} \tag{24}$$

For closed contour integrals, Green's theorem can be used on both the real and the imaginary parts of Eq. (24). For a differentiable complex function $g(z_R,z_I)$, Green's theorem is given by (p. 99 of Ref. [32])

$$\oint g\, d\tilde{z} = i\oiint\left\{\frac{\partial g}{\partial z_R} + i\frac{\partial g}{\partial z_I}\right\}dz_R dz_I. \tag{25}$$

Substitution of Eq. (24) into Eq. (25) gives

$$\begin{aligned}&\oiint_S\left[\frac{\partial^2\ln(\xi)}{\partial z_R^2} + i\frac{\partial^2\ln(\xi)}{\partial z_R\partial z_I}\right]dz_R dz_I \\ &\quad -i\oiint_S\left[\frac{\partial^2\ln(\xi)}{\partial z_R\partial z_I} + i\frac{\partial^2\ln(\xi)}{\partial z_I^2}\right]dz_R dz_I \\ &= \oiint_S\left[\frac{\partial^2\ln(\xi)}{\partial z_R^2} + \frac{\partial^2\ln(\xi)}{\partial z_I^2}\right]dz_R dz_I,\end{aligned} \tag{26}$$

so that Eq. (23) can be expressed as

$$\phi_B = -i\oint\frac{\partial\ln(\xi)}{\partial\tilde{z}}d\tilde{z} = \oiint\nabla^2_{z_R,z_I}\ln(\xi)dz_R dz_I, \tag{27}$$

where the two-dimensional Laplacian is defined by the double derivatives with respect to z_R and z_I in Eq. (26). To reveal the geometric aspect of Eq. (27), consider again our archetypal Gaussian beam on the optic axis in two transverse dimensions [11], and evaluate the Laplacian,

$$\begin{aligned}\nabla^2_{\tilde{z}}\ln(\xi) &= \nabla^2_{x,y}\ln[(1+|\tilde{z}|^2)^{-1/2}e^{i\,\tan^{-1}(\tilde{z})}] \\ &= -\frac{2}{\left(1+z_R^2+z_I^2\right)^2}.\end{aligned} \tag{28}$$

Mapping points from the Riemann sphere to the complex plane (p. 17 of Ref. [32]) allows the areal element dA on the unit sphere to be written as

$$dA = \frac{4}{\left(1+z_R^2+z_I^2\right)^2}dz_R dz_I. \tag{29}$$

Combining Eqs. (27)–(29), we have

$$\phi_G = \phi_B = -i\oint\frac{\partial\ln(\psi)}{\partial\tilde{z}}d\tilde{z} = -\frac{1}{2}\oiint dA = -\frac{\Omega}{2}. \tag{30}$$

In one transverse dimension, the square root in the Gaussian beam [27] changes to give the expression,

$$\phi_G = \phi_B = -i\oint\frac{\partial\ln(\psi)}{\partial\tilde{z}}d\tilde{z} = -\frac{1}{4}\oiint dA = -\frac{\Omega}{4}, \tag{31}$$

where Ω is the solid angle subtended of the areal patch on the Riemann sphere.

As with our derivation based upon the argument principle, we find that the phase changes in Eqs. (30) and (31) are invariant to an arbitrary change in the Gaussian beam phase scale and to that in the complex focus $\tilde{z}$. Similarly, we are free to distort the Gaussian envelope wave by multiplying ξ with any function that has an as analytic logarithmic derivative, without changing the value of the quantized phase, by virtue of Cauchy's integral theorem. In other words, when we integrate over half the Cartesian plane or contour integrate along the entire real (optic) axis, the magnitude of the Gouy phase anomaly is explicitly $\pi/2$ rad in one transverse dimension and π rad in two transverse dimensions.

III. CONCLUSION

Within the context of focused matter waves, our theoretical description of the Gouy anomaly unites Heisenberg's uncertainty principle, Berry's geometric phase [17], and Maslov indices for rays that touch caustics [20,21]. If we denote the phase change from transverse fluctuations in momentum as a "Heisenberg phase," then, for focused paraxial electron beams, the Gouy, Heisenberg, Keller, Berry, and Lévy-Leblond phases all are seen to be manifestations of the same phenomenon. Using a geometric construction based upon the standard deviation of a focused electron beam, these anomalous phase shifts were united by the longitudinal logarithmic derivative of a paraxial wave. The simplicity of this view suggests feasible new experiments for accurately measuring these phase anomalies for a variety of caustics. For matter waves that maintain extrema on the optic axis, Eq. (12) shows that phase anomalies can be measured directly from transverse probability densities without recourse to any form of interferometry or phase retrieval, which is possible since transverse sections of caustics are Gabor holograms [34]. Indeed, experiments utilizing Eq. (12) would constitute successive weak measurements of the local momentum [28] on the optic axis.

ACKNOWLEDGEMENTS

We thank S. Kennedy and D. Jesson from the Monash University School of Physics for their input and stimulating discussions. Funding from the Australian Research Council Discovery Project Grant No. DP1092745 is acknowledged. T.C.P. is also grateful for financial support from the Monash Centre for Electron Microscopy.

APPENDIX: THE LÉVY-LEBLOND PHASE SHIFT AND STATISTICAL CONFINEMENT

In this section, we describe the Gouy phase anomaly by considering the effects of geometric confinement, which supports the interpretation given by Hariharan and Robinson [13]. Far from any obstructions, an electron wave is free and propagates along the z axis with energy E given by

$$E = \frac{h^2k^2}{2m}, \tag{A1}$$

where the wave number $k = 1/\lambda$ and m is the electron mass.

Confinement along the transverse axis within a distance L produces a change in the wave-vector magnitude from k to k' [30] since

$$E = \frac{h^2k'^2}{2m} + \frac{h^2n^2}{8mL^2}, \tag{A2}$$

where n is an integer greater than zero. Using conservation of energy, we can equate Eqs. (A1) and (A2) to obtain

$$E = \frac{h^2k^2}{2m} = \frac{h^2k'^2}{2m} + \frac{h^2n^2}{8mL^2} \rightarrow k^2 = (k-\Delta k)^2 + \frac{n^2}{4L^2}, \tag{A3}$$

where $h\,\Delta k$ is the change in the longitudinal momentum k_z.

Assuming small changes in k, $(\Delta k)^2 \ll k\,\Delta k$, we have

$$\Delta k = \frac{n^2}{8kL^2}. \tag{A4}$$

The integer n^2 must be replaced by $n^2 + m^2$ for waves propagating in two transverse dimensions, which doubles the change in momentum when both n and m are unity. Hence, the essence of the Lévy-Leblond [30] phase shift is a geometric effect which reduces the momentum upon confinement of the wave in the transverse coordinate, orthogonal to the longitudinal propagation direction.

In the absence of boundaries or cavities, focused beams conserve kinetic energy, and we expect no change in the average longitudinal momentum. However, we can describe the Gouy anomaly in terms of the Lévy-Leblond phase shift if we identify the standard deviation curve in Fig. 1 with statistical confinement of the wave, which varies continuously along the z axis. To examine local phase variations, we must consider an expectation value of the energy or momentum in the vicinity of the optic axis instead of Eq. (A3). Since the paraxial wave equation is formally equivalent to the time-dependent Schrödinger equation with two spatial variables, we can replace the expectation value of the energy with $\langle \partial/\partial z \rangle$, integrated over an infinitesimal patch of the transverse plane containing the optic axis, which is equivalent to Eq. (12) in the paper. Hence, from the perspective of the Lévy-Leblond effect, focusing produces statistical confinement, which, in turn, induces local momentum fluctuations that give rise to longitudinal phase variations in the optic axis. Naively, this can be inferred from Eq. (A4) if the confinement parameter L is made proportional to the uncertainty of the beam. Integration along the optic axis then gives an inverse tangent function which is proportional to the Gouy phase.

[1] L. G. Gouy, Ann. Chim. Phys. **24**, 145 (1891).
[2] E. H. Linfoot and E. Wolf, Proc. Phys. Soc. London, Sect. B **69**, 823 (1956).
[3] T. D. Visser and E. Wolf, Opt. Commun. **283**, 3371 (2010).
[4] M.-S. Kim, T. Scharf, A. da Costa Assafrao, C. Rockstuhl, S. F. Pereira, H. P. Urbach, and H. P. Herzig, Opt. Express **20**, 28929 (2012).
[5] X. Y. Pang, G. Gbur, and T. D. Visser, Opt. Lett. **36**, 2492 (2011).
[6] X. Y. Pang, T. D. Visser, and E. Wolf, Opt. Commun. **284**, 5517 (2011).
[7] E. Y. S. Yew and C. J. R. Sheppard, Opt. Lett. **33**, 1363 (2008).
[8] X. Y. Pang and T. D. Visser, Opt. Express **21**, 8331 (2013).
[9] X. Pang, D. G. Fischer, and T. D. Visser, J. Opt. Soc. Am. A **29**, 989 (2012).
[10] D. Pelliccia and D. M. Paganin, Phys. Rev. A **86**, 015802 (2012).
[11] R. W. Boyd, J. Opt. Soc. Am. **70**, 877 (1980).
[12] R. Borghi, M. Santarsiero, and R. Simon, J. Opt. Soc. Am. A **21**, 572 (2004).
[13] P. Hariharan and P. A. Robinson, J. Mod. Opt. **43**, 219 (1996).
[14] S. M. Feng and H. G. Winful, Opt. Lett. **26**, 485 (2001).
[15] J. Yang and H. G. Winful, Opt. Lett. **31**, 104 (2006).
[16] D. Subbarao, Opt. Lett. **20**, 2162 (1995).

[17] M. V. Berry, Proc. R. Soc. London, Ser. A **392**, 45 (1984).
[18] R. Simon and N. Mukunda, Phys. Rev. Lett. **70**, 880 (1993).
[19] T. Tyc, Opt. Lett. **37**, 924 (2012).
[20] J. B. Keller, Ann. Phys. **4**, 180 (1958).
[21] J. B. Keller, SIAM Rev. **27**, 485 (1985).
[22] Y. I. Orlov, Radiophys. Quantum Electron. **24**, 154 (1981).
[23] J. Hamazaki, Y. Mineta, K. Oka, and R. Morita, Opt. Express **14**, 8382 (2006).
[24] G. Guzzinati, P. Schattschneider, K. Y. Bliokh, F. Nori, and J. Verbeeck, Phys. Rev. Lett. **110**, 093601 (2013).
[25] P. Schattschneider, M. Stöger-Pollach, and J. Verbeeck, Phys. Rev. Lett. **109**, 084801 (2012).
[26] T. C. Petersen, D. M. Paganin, M. Weyland, T. P. Simula, S. A. Eastwood, and M. J. Morgan, Phys. Rev. A **88**, 043803 (2013).
[27] A. E. Siegman, *Lasers* (University Science, Sausalito, CA, 1986).
[28] M. V. Berry, J. Opt. A: Pure Appl. Opt. **11**, 094001 (2009).
[29] T. E. Gureyev, A. Roberts, and K. A. Nugent, J. Opt. Soc. Am. A **12**, 1932 (1995).
[30] J.-M. Lévy-Leblond, Phys. Lett. A **125**, 441 (1987).
[31] L. J. Allen and M. P. Oxley, Opt. Commun. **199**, 65 (2001).
[32] M. J. Ablowitz and A. S. Fokasz, *Complex Variables, Introduction and Applications* (Cambridge University Press, Cambridge, U.K., 1997).
[33] M. R. Teague, J. Opt. Soc. Am. **73**, 1434 (1983).
[34] D. Gabor, Proc. R. Soc. London, Ser. A **197**, 454 (1949).

Bloch Sphere Catastrophes

Samuel A. Eastwood,* David M. Paganin, Timothy C. Petersen, and Michael J. Morgan

School of Physics, Monash University, Victoria 3800, Australia

(Dated: June 4, 2014)

Abstract

Caustics are optical phenomena which occur when a family of rays creates an envelope of divergent intensity. Here we show that caustic surfaces also appear when a real or complex field is mapped to its order parameter manifold. We study these structures in the context of spin-1/2 fields, where the order parameter manifold is the Bloch sphere. These generic structures are a manifestation of catastrophe theory and are stable with respect to perturbations. The corresponding field configurations are also stable and represent a new type of topological defect. Equations governing the conditions for their existence and unfolding are derived.

PACS numbers: 05.45.-a, 04.20.Gz, 61.72.-y, 02.40.Pc

INTRODUCTION

Order parameters are pivotal to the study and classification of topological defects and phase transitions [1]. Many important areas of physics can be understood using this construct, e.g., crystal growth [2], quantum computing [3], nematic liquid crystals [4], magnetic textures, such as Skyrmions [5–7], multiferroics [8–10], superconductors [11, 12] and particle cosmology [13], to name but a few. The order parameter may also be viewed as a function that maps points in the physical space to the order parameter space. The topological properties of the order parameter space are important in the study of defects, since they govern the stability of defects in the field [14].

Here we demonstrate the existence of a new type of singularity formed when a field is mapped to its order–parameter space; these defects are governed by the framework of catastrophe theory [15–17]. Thom's theorem [18] tells us that catastrophes are stable with respect to perturbation; consequently features of the field that map to a catastrophe on the order–parameter space manifold must also be stable to perturbations. This singularity represents a new type of topological defect, which we refer to as an *order–parameter catastrophe defect* (see Fig. 1).

Such catastrophes possess their own topology regardless of the order parameter manifold on which they exist. Their stability is intrinsic to the catastrophe defect, rather than depending on the topology of the order parameter manifold. This implies that local regions of the field, which map to the order parameter catastrophe defect, transcend the topology of the order parameter space, due to the intrinsic topology of the catastrophe itself. Catastrophe theory may be applied to any system whose order parameter space is "tattooed" with such catastrophes.

Catastrophes themselves are very generic structures, appearing naturally in many physical systems [15, 19–29]. Hence the concepts presented here will be applicable to a broad range of phenomena. Although the fundamental idea is very general, we choose to contextualize the concept of order–parameter catastrophes by applying it to a 2D spin–1/2 field, for which the corresponding order parameter space is the Bloch sphere.

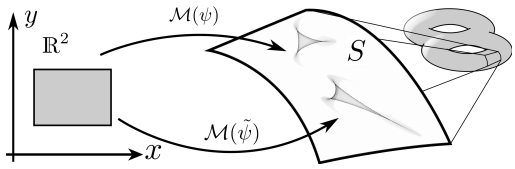

FIG. 1. Schematic showing how the field ψ in an $\mathbb{R}^2$ patch is mapped to its order parameter manifold, leading to a catastrophe on S. Here S is a local patch of the manifold, which may possess a non-trivial global topology. Due to the intrinsic stability of the catastrophe, a mapping of the patch of the perturbed field, $\tilde{\psi}$, would yield a deformed catastrophe. The features of the field that are mapped to the catastrophe persist in the presence of perturbations.

BLOCH SPHERE MAPPING

Consider the general wave–function of a two–component spinor

$$|\Psi(\mathbf{x},t)\rangle = \psi_0(\mathbf{x},t)\,|0\rangle + \psi_1(\mathbf{x},t)\,|1\rangle\,, \tag{1}$$

where $\mathbf{x}$ is the spatial variable and t the time coordinate; ψ_0 and ψ_1 are the probability amplitudes of the spin up state $|0\rangle$ and spin down state $|1\rangle$, respectively. Without loss of generality we may assume $|\psi_0|^2 + |\psi_1|^2 = 1$, and consider only pure spin states that map to the surface of the Bloch sphere. A single spin state is represented graphically as a unit position vector of the Bloch sphere, denoted by $|\mathscr{B}\rangle$. It is parametrized by the spherical polar angles (ϕ,θ), as

$$\begin{aligned} |\mathscr{B}\rangle = e^{i\chi}[&e^{-i\phi(x,y)/2}\cos\left(\theta(x,y)/2\right)|0\rangle \\ &+e^{i\phi(x,y)/2}\sin\left(\theta(x,y)/2\right)|1\rangle], \end{aligned} \tag{2}$$

where χ is the global phase of the Bloch vector [30–32]. The global phase factor $\exp(i\chi)$ is an unobservable of the system.

Here we are concerned with the mapping of the Bloch vector, located at every point in coordinate space, to the Bloch sphere. Consider the mapping of an (x,y)–patch of Bloch vectors, corresponding to a spin–1/2 field, to the surface of the Bloch sphere. If the patch contains a texture defect it will wrap the surface of the Bloch sphere an integer number of times [2, 33–35]. If the field is perturbed locally, the texture structure will not be destroyed, since it is topologically protected. Instead, the mapping is altered in such a way that a local

(x, y)–patch may be sheared, dilated or rotated by the mapping, whilst still maintaining the condition that the surface of the Bloch sphere remains wrapped. To accommodate transforms of this type, small regions on the Bloch sphere must fold back onto themselves, resulting in a fold catastrophe. This highlights the ubiquity of order–parameter catastrophes; a continuous perturbation of a texture defect can yield a caustic "tattoo" on the Bloch sphere (see Fig. 1). In the general case, when a texture defect exists in a spin field, Bloch–sphere catastrophes will also be present. However, the converse is not necessarily true.

CONDITIONS FOR A CATASTROPHE.

The explicit form of the Bloch sphere map is determined by the inverse of Eq. (2), given by

$$\phi(x, y) = \arg\left(\psi_1/\psi_0\right), \tag{3}$$

$$\theta(x, y) = \arccos\left(\eta\right). \tag{4}$$

The quantity $\eta = \psi_0^*\psi_0 - \psi_1^*\psi_1$ is the spin asymmetry, which represents the difference in probability density of the spin up and down components. The Bloch sphere image is a 2D histogram of ϕ and θ, each of which are calculated for every (x, y) point of the spin field. The mapping takes a patch of the (x, y)–plane and calculates the complex functions $\psi_0(x, y)$ and $\psi_1(x, y)$. Together these functions completely define the spin–1/2 field over the Euclidean patch in $\mathbb{R}^2$ These states are then mapped to the corresponding (ϕ, θ)–patch on the Bloch sphere. The density of the patch on the Bloch sphere at a particular point is proportional to the number of times the spin–1/2 field takes on the values of that particular state.

A randomly varying spinor wavefunction was generated by low–pass filtering an image of white noise for each spin component, producing the field shown in Fig. 2(a). By applying Eqs. (3) and (4) the field was mapped to the Bloch sphere to produce Fig. 2(b). The presence of several phase vortices can be seen in the Bloch sphere coordinate $\phi(x, y)$, shown in real space in Fig. 2(c). These screw–type topological defects come from the vortices in the phase of both components in Eq. (1), which is preserved when $\phi(x, y)$ is calculated.

A Bloch sphere singularity corresponds to a many–to–one mapping. This occurs when the mapping becomes singular, hence the Jacobian determinant of the mapping must vanish.

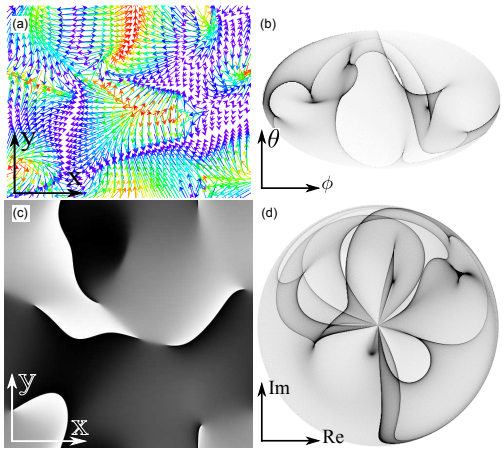

FIG. 2. (a) Bloch vectors of a random spin–1/2 field distributed in real space with the color corresponding to their z–component. (b) Mollweide projection of the Bloch sphere map. Here the density of the plot is proportional to the number of spin states in the region of the spin–1/2 field shown in (a). (c) The Bloch sphere coordinate ϕ displayed as a 2D function of (x, y). (d) The Γ map, defined in the main text, of the spin field in (a). This map can be thought of as a projection mapping the Bloch sphere to the Argand plane.

The Jacobian is defined by

$$J(x,y) = \begin{pmatrix} \frac{\partial\phi}{\partial x} & \frac{\partial\phi}{\partial y} \\ \frac{\partial\theta}{\partial x} & \frac{\partial\theta}{\partial y} \end{pmatrix}. \tag{5}$$

Setting $\det[J(x,y)] = 0$ gives the conditions for a particular spin state to map to a catastrophe, i.e.

$$\left[\left(|\psi_0|^2\mathbf{j}_2 - |\psi_1|^2\mathbf{j}_1\right) \times \nabla\eta\right] \cdot \hat{\boldsymbol{z}} = 0, \tag{6}$$

where $\mathbf{j}_1$ and $\mathbf{j}_2$ are the probability current densities of ψ_0 and ψ_1, respectively, and $\hat{\boldsymbol{z}}$ is a unit vector in the z direction. It can be shown that, up to a scaling factor, Eq. (6) is equivalent to the z component of the curl of the probability current density of a single complex scalar wave-function of the form

$$\Gamma(x,y) = \sqrt{\eta}\exp(i\phi). \tag{7}$$

Equation (7) is a projection that maps the Bloch sphere to the Argand plane; this "Γ map" is shown in Fig. 2(d). Zeros of Γ are mapped to the equator of the Bloch sphere, while the

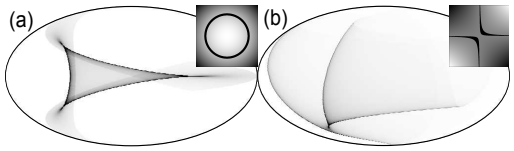

FIG. 3. Mollweide plots of (a) elliptic, and (b) hyperbolic umbilic Bloch sphere catastrophes. The insets of each figure show the Jacobian determinant of the mapping, with their zeros highlighted in black. When the zeros of the Jacobian form an ellipse, the map of that region forms an elliptic umbilic catastrophe. Similarly, when the zeros form a hyperbolic curve, the field maps to the hyperbolic umbilic catastrophe.

entire equator is mapped to the the origin of the Argand plane. Each hemisphere is projected over a unit disc, with the poles mapping to the disc's circumference at $|\Gamma| = 1$. However, each hemisphere of the Bloch sphere is rotated relative to the other by $\pi/2$ radians; this is a direct result of the abrupt phase change Γ acquires when η changes sign.

The construction of Γ allows us to express the equation governing order–parameter space defects of the spin–1/2 field in terms of the vorticity of Γ. The current density of Γ has the form $\eta\nabla\phi$, which may be interpreted as an equation describing the flow of spin asymmetry. Catastrophes are mapped to the Bloch sphere along points where the vorticity of Γ vanishes.

RELATION TO THE CANONICAL FORM

Catastrophes are associated with an instability of a system. Any potential function which governs the behavior of a system will consist of state variables, in this case (x, y), and a set of parameters called the control variables. Degenerate critical points of a potential lead to instabilities of the system because a bifurcation may occur at these points when there is a small change in the control variables. Catastrophes exist in the space of the control variables of the potential function. The dimensionality of the control space is determined by the number of control parameters of the potential function, i.e. the co–dimension. The *bifurcation set* defines a caustic surface embedded in control space; this corresponds to the set of values of the control variables at which bifurcations of the critical points in the potential function occur.

Thom's theorem describes a number of so-called *elementary catastrophes*, these being all possible topologically distinct catastrophes that occur when the co-dimension is less than or equal to four [18]. Every elementary catastrophe has an associated function known as the canonical form [15–18]. Therefore if a local patch of a spin–1/2 field maps to a particular elementary catastrophe on the Bloch sphere, the field itself must have a relation to the canonical form of the catastrophe. The Bloch sphere catastrophe is embedded in the coordinates (ϕ, θ) of the sphere. However, the catastrophe bifurcation set lies in the control space. In essence, the Bloch sphere catastrophe is a slice though the control space of the bifurcation set, which has also been subject to a continuous deformation, rotation and translation.

In catastrophe theory, a function has a degenerate critical point when the Hessian determinant vanishes. Points of the spin field where the Jacobian determinant vanishes map to a catastrophe. Therefore the Hessian of the canonical form of the catastrophe, $V(x, y)$, is locally equivalent to the Jacobian matrix of the mapping, i.e.

$$\begin{pmatrix} \frac{\partial\tilde{\phi}}{\partial x} & \frac{\partial\tilde{\phi}}{\partial y} \\ \frac{\partial\tilde{\theta}}{\partial x} & \frac{\partial\tilde{\theta}}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 V}{\partial x^2} & \frac{\partial^2 V}{\partial y \partial x} \\ \frac{\partial^2 V}{\partial x \partial y} & \frac{\partial V}{\partial y^2} \end{pmatrix}, \tag{8}$$

where $\tilde{\phi}$ and $\tilde{\theta}$ are the transformed coordinates of the Bloch sphere, which account for the local translation and rotation of the catastrophe from its canonical form. To account for smooth deformations of the catastrophe of the Bloch sphere mapping, the coefficients of the transformations of ϕ and θ must be functions of the state space coordinates (x, y). Since the transformation matrix has an inverse, this allows us to express the Bloch sphere coordinates in terms of the transformed coordinates as

$$\begin{pmatrix} \phi \\ \theta \end{pmatrix} = \begin{pmatrix} a(x,y) & b(x,y) \\ c(x,y) & d(x,y), \end{pmatrix} \begin{pmatrix} \tilde{\phi} \\ \tilde{\theta} \end{pmatrix}, \tag{9}$$

where the inverse transformation matrix elements a, b, c and d are arbitrary functions of (x, y), which encapsulate the local rotations, translations or deformations of the coordinate system.

Given Eqs. (8) and (9) we find that $\phi(x, y)$ and $\theta(x, y)$ may be expressed in terms of the partial derivatives of the canonical form of the catastrophe:

$$\phi(x, y) = a(x, y)V_x(x, y) + b(x, y)V_y(x, y) \tag{10}$$

$$\theta(x, y) = c(x, y)V_x(x, y) + d(x, y)V_y(x, y) \tag{11}$$

where $V_x \equiv \partial V/\partial x$ and $V_y \equiv \partial V/\partial y$. Equations (10) and (11) show that the Bloch sphere catastrophe is indeed a form of gradient mapping of the canonical form of the catastrophe. If a small patch on the spin field in state space were to map to a catastrophe, the field in that patch of state space must relate to the canonical form of the catastrophe. The field in these local regions has the functional form:

$$\begin{aligned} |\Psi\rangle = e^{-\frac{i}{2}[(aV_x+bV_y)}\cos\left[(cV_x+dV_y)/2\right]|0\rangle + \\ e^{\frac{i}{2}(aV_x+bV_y)}\sin\left[(cV_x+dV_y)/2\right]|1\rangle\,. \end{aligned} \tag{12}$$

The Bloch sphere coordinates (ϕ, θ) are deformed gradient maps of the canonical form of the catastrophe; the structure of the zeros of the Jacobian determinant determines the type of elementary catastrophe to which the field will map. In essence, the lines of zeros of the Jacobian determinant determine where to fold a patch in $\mathbb{R}^2$, as it is mapped to the Bloch sphere. The way in which the patch is folded yields a particular type of catastrophe. Examples of elliptic and hyperbolic umbilic catastrophes are shown in Fig. 3, along with the geometry of each catastrophe's Jacobian determinant given in the insets. The canonical form of the elliptic umbilic catastrophe is $x^3 - xy^2$, where the null set for the Hessian determinant of this form corresponds to the equation of an ellipse, i.e. the zeros of the Jacobian determinant map to an elliptic umbilic catastrophe. Similarly the Hessian determinant of the hyperbolic umbilic catastrophe is the equation of a hyperbola, given by its canonical form of $x^3 + y^3$. We have shown that the spin field is related to the canonical form of the catastrophe by a gradient map, therefore perturbations to the spin–1/2 field behave as a particular unfolding of the canonical form.

STABILITY AND UNFOLDING OF BLOCH SPHERE CATASTROPHES.

A Bloch sphere catastrophe is a particular cross–section of the full bifurcation set surface. Catastrophes may undergo a process known as *unfolding*. This occurs when the potential function of any given catastrophe is subject to a perturbation in such a way that the slice of the bifurcation set that is viewed on the Bloch sphere changes. In this picture we see that the topological stability of a Bloch sphere catastrophe defect arises because every catastrophe already possesses a topology that is fixed by its bifurcation set. A perturbation to the spin field only serves to unfold the Bloch sphere catastrophe, whence we observe different slices through the bifurcation set of the catastrophe.

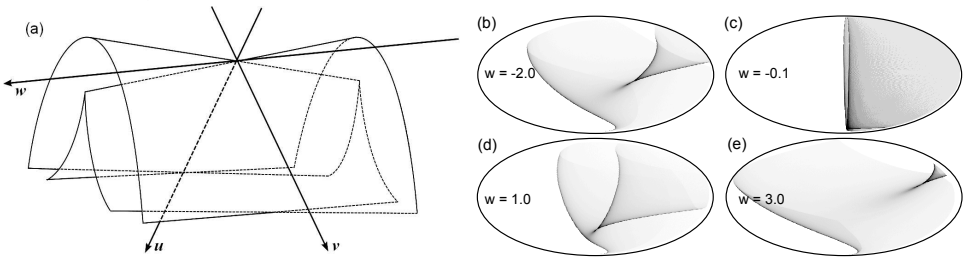

FIG. 4. (a) Bifurcation set of the hyperbolic umbilic catastrophe. (b)-(e) Bloch sphere projections showing a hyperbolic umbilic catastrophe unfolding through various values of w.

To demonstrate this consider the full bifurcation set of the hyperbolic umbilic catastrophe shown in Fig. 4(a). By inputting the known canonical form and using Eq. (12) we can map an isolated hyperbolic umbilic catastrophe to the Bloch sphere and control its unfolding. This is demonstrated in Figs. 4(b)–(e). This method of spin–engineering by way of using a catastrophe's canonical form could be used to generate any Bloch sphere mapping at a given unfolding. The hyperbolic umbilic catastrophes that appear on the Bloch sphere are intersections of the Bloch sphere surface with the catastrophe's 3–dimensional bifurcation set. In this example the cross–sections are the $u - v$ plane. As the control parameter w is varied the $u-v$ plane shifts along the w axis, which unfolds the Bloch sphere catastrophe. In general, a hyperbolic umbilic Bloch sphere catastrophe will express itself as cross–sections of the bifurcation set through some arbitrary plane, in which case the unfolding of the catastrophe will depend on all three control parameters u, v and w.

DISCUSSION

Order–parameter catastrophe defects are a generic form of topological defect that are found in fields. Catastrophes are very general structures that almost always arise in non–trivial mappings to an order parameter manifold. Equations (10) and (11) show that the order–parameter coordinates are proportional to the first partial derivatives of a catastrophe's canonical form, which are always simple polynomial functions. This means that both ϕ and θ will also have a local polynomial form for regions that map to a catastrophe. If we consider any sufficiently small local expansion about a point of an arbitrary spin field,

the patch only need be diffeomorphic to the canonical form of one of the elementary catastrophes. This criterion is satisfied for the creation of a catastrophe and demonstrates that a mapping of a generic field to its order parameter space will result in caustic tattoos on its manifold. This is a general result that can be found in other types of mappings. For example, catastrophes manifest themselves as *Argand–plane caustics*, which arise when a complex scalar field is mapped to the Argand–plane [36, 37]. In that context, the condition that a point maps to a caustic, $\det[J(x, y)] = 0$, is equivalent to the optical vorticity vanishing. By making a similar comparison in Eq. (8) for a complex field it is then apparent that Argand–plane catastrophes also rise from gradient mappings.

Whilst we have used the specific example of 2D spin–1/2 fields and the Bloch sphere, it is important to emphasize that all the concepts presented here are very general. We need only ask what is the mapping of any field to its order parameter space to be able to view the field through the new perspective of order parameter catastrophe defects. This concept is not restricted to 2–dimensions; the 2D surface of the Bloch sphere allows us to view 1D or 2D catastrophes, such as folds or cusps, or 2D cross–sections of higher order catastrophes, such as the elliptic or hyperbolic umbilic catastrophes (Fig. 3). For mappings to order parameter spaces of higher dimensionality, the full bifurcation sets or higher dimensional cross–sections of catastrophes with higher co–dimension would also become observable.

In summary, order parameter catastrophe defects represent a new form of topological defect in fields. Their stability is attributed to the intrinsic topology of the catastrophes themselves rather than that of the order parameter space, as it is the case for standard topological defects and textures.

The authors thank Dr. Kavan Modi for insightful discussions. S.A.E gratefully acknowledges funding from an Australian Postgraduate Award. D.M.P and M.J.M acknowledge funding from the Australian Research Council.

---

* samuel.eastwood@monash.edu

[1] J. Sethna, *Statistical Mechanics: Entropy, Order Parameters and Complexity* (Oxford University Press, Oxford, 2006).

[2] N. D. Mermin, Rev. Mod. Phys. **51**, 591 (1979).

[3] V. N. Golovach, M. Borhani, and D. Loss, Phys. Rev. A **81**, 022315 (2010).

[4] P. Pieranski, B. Yang, L. J. Burtz, A. Camu, and F. Simonetti, Liq. Cryst. **40**, 1593 (2013).

[5] U. K. Röszler, A. N. Bogdanov, and C. Pflederer, Nature **442**, 797 (2006).

[6] S. Mühlbauer, B. Binz, F. Jonietz, C. Pfleiderer, A. Rosch, A. Neubauer, R. Georgii, and P. Böni, Science **323**, 915 (2009).

[7] D. Ezwana, Phys. Rev. Lett. **105**, 197202 (2010).

[8] J. Wang *et al.*, Science **299**, 1719 (2003).

[9] T. Kimura, S. Kawamoto, I. Yamada, M. Azuma, M. Takano, and Y. Tokura, Phys. Rev. B **67**, 180401 (2003).

[10] Y. H. Chu *et al.*, Nature Mat. **7**, 478 (2008).

[11] E. Gull, O. Parcollet, and A. J. Millis, Phys. Rev. Lett. **110**, 216405 (2013).

[12] G. Seibold, L. Benfatto, C. Castellani, and J. Lorenzana, Phys. Rev. Lett. **108**, 207004 (2012).

[13] M. Fairbairn and R. Hogan, J. High Energy Phys. **2013**, 1 (2013).

[14] G. Ruben, M. J. Morgan, and D. M. Paganin, Phys. Rev. Lett. **105**, 220402 (2010).

[15] V. I. Arnol'd, Russ. Math. Surv+ **30**, 1 (1975).

[16] P. T. Saunders, *An Introduction To Catastrophe Theory* (Cambridge University Press, Cambridge, 1980).

[17] T. Poston and I. Steward, *Catastrophe Theory and Its Applications* (Dover Publications Inc., New York, 1996).

[18] R. Thom, *Mathematical Models of Morphogenesis* (Wiley, New York, 1983).

[19] J. Hannay, Opt. Acta **29**, 1631 (1982).

[20] M. V. Berry, Adv. Phys. **25**, 1 (1976).

[21] M. V. Berry and C. Upstill, Prog. Opt. **18**, 259 (1980).

[22] J. Nye, Proc. R. Soc. A **361**, 21 (1978).

[23] J. F. Nye, J. Opt. A-Pure Appl. Op. **5**, 495 (2003).

[24] J. F. Nye, J. Opt. A-Pure Appl. Op. **5**, 503 (2003).

[25] J. F. Nye, J. Opt. A-Pure Appl. Op. **8**, 304 (2006).

[26] T. C. Petersen, M. Weyland, D. M. Paganin, T. P. Simula, S. A. Eastwood, and M. J. Morgan, Phys. Rev. Lett. **110**, 033901 (2013).

[27] T. P. Simula, T. C. Petersen, and D. M. Paganin, Phys. Rev. A **88**, 043626 (2013).

[28] R. Thom, SIAM Rev. **19**, 189 (1977).

[29] K. Maeda, T. Tachizawa, T. Torii, and T. Maki, Phys. Rev. Lett. **72**, 450 (1994).

[30] J. M. Martinis, S. Nam, J. Aumentado, K. M. Lang, and C. Urbina, Phys. Rev. B **67**, 094510 (2003).

[31] G.-Y. Xiang, J. Li, B. Yu, and G.-C. Guo, Phys. Rev. A **72**, 012315 (2005).

[32] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[33] Y. Zhang, Y. Ran, and A. Vishwanath, Phys. Rev. B **79**, 245331 (2009).

[34] Y. Kawaguchi, M. Nitta, and M. Ueda, Phys. Rev. Lett. **100**, 180403 (2008).

[35] H. Mäkelä, Y. Zhang, and K.-A. Suominen, J. Phys. A **36**, 8555 (2003).

[36] F. Rothschild, M. J. Kitchen, H. M. L. Faulkner, and D. M. Paganin, Optics Commun. **285**, 4141 (2012).

[37] F. Rothschild, A. I. Bishop, M. J. Kitchen, and D. M. Paganin, Opt. Express **22**, 6495 (2014).