

E. WOLF, PROGRESS IN OPTICS XVIII © NORTH-HOLLAND 1980

IV

**CATASTROPHE OPTICS:
MORPHOLOGIES OF CAUSTICS AND THEIR DIFFRACTION
PATTERNS**

BY

M. V. BERRY and C. UPSTILL

*H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue,
Bristol, BS8 1TL, U.K.*

CONTENTS

	PAGE
§ 1. INTRODUCTION	259
§ 2. CATASTROPHE THEORY AND STABLE CAUSTICS. .	262
§ 3. DIFFRACTION CATASTROPHES	277
§ 4. THE GEOMETRY OF WAVEFRONTS AND THEIR NORMALS	297
§ 5. LIQUID DROPLET LENSES	305
§ 6. CAUSTIC NETWORKS	315
§ 7. STATISTICAL CATASTROPHE OPTICS.	326
§ 8. CONCLUDING REMARKS.	336
ACKNOWLEDGMENTS	337
APPENDIX 1: WAVEFRONTS	337
APPENDIX 2: PRIMARY ABERRATIONS	339
REFERENCES	343

§ 1. Introduction

As light propagates, it is common for focusing to occur. In the most elementary picture, all the rays pass through a focal point. It is exceptional, however, for this to correspond to reality: what happens typically is that the rays envelop a caustic surface in space. The light intensity is greatest near caustics, and they dominate optical images whenever the wavelength is small enough for geometrical optics to form the basis of a valid approximation to the wavefield.

In spite of their importance, caustics have been awkward to fit into optical theory, for two main reasons. Firstly, their forms are very intricate and until recently were only classified as perturbations from a perfect focus, i.e. as lens aberrations. Secondly, the simplest short-wave theory, in which the rays of geometrical optics are endowed with amplitude and phase and allowed to interfere where they cross, fails on caustics by predicting unphysical intensity divergences. In other words, the theory goes wrong in precisely the most important places – where the light is brightest!

This unsatisfactory state of affairs has been changed dramatically by the injection into optics of a new branch of mathematics called *catastrophe theory*, invented by THOM [1972] and developed most notably by ARNOL'D [1975]. It is the purpose of this review to show how the classification of caustics as catastrophes is leading us to understand a wide variety of optical phenomena. The resulting *catastrophe optics* is unfamiliar and unorthodox, cutting across traditional categories within the subject. In large measure this new point of view arises from one concept which plays a crucial role in the mathematics, namely *structural stability*. We shall show that this enables us to classify those caustics whose topology survives perturbation. Such stable caustics abound in nature in circumstances where there is no special symmetry, for example when caustics are formed as a result of refraction or reflection of light by a wavy water surface. The perfect point focus so sought after in optical technology is in this mathematical sense highly unstable, a fact borne out

by the difficulty of producing it. Thus catastrophe optics is complementary to traditional optics, where the emphasis is on solving ray or wave equations in highly symmetric situations.

At the outset we wish to dispel some possible misconceptions. Firstly, the fact that catastrophe theory is rooted in topological (i.e. qualitative) concepts does not make it imprecise; we shall show that it is perfectly possible to make quantitative statements in catastrophe optics. Secondly, the fact that the fundamental theorem is a classification does not make its application to optics purely taxonomic; we shall show how to use the theory to make surprising and testable predictions. And thirdly, although some invocations of catastrophe theory in science have been criticised (ZAHLER and SUSSMAN [1977]), these criticisms do not apply (and were not intended to apply) to catastrophe optics, which has a firm mathematical foundation.

We begin in § 2 by explaining how the physics of stable caustics, which is concerned with envelopes of families of rays, is isomorphic to the mathematics of catastrophe theory, which is concerned with stable singularities of gradient maps. The precise correspondence occurs because rays obey the Hamilton–Fermat principle of stationary time. The stable caustics have the same structure as the catastrophes, and we illustrate them by simple examples.

It is helpful to think of the catastrophes as elemental “atomic forms”. The corresponding caustics have a delicate “subatomic” structure of diffraction fringes, and they can link up into elaborate “macroscopic” networks whose details are often not resolved. Catastrophe optics thus embodies a hierarchy of scales. Our subsequent examples will range from the topology of the zeroes of diffraction patterns on scales of a few wavelengths to the topology of caustic networks on scales of metres. The catastrophes themselves occupy a sort of mesoscale, organising finer and grosser structures.

In § 3 we step down to the microscale, to explore the *diffraction catastrophes* that both clothe and underlie caustics. Each structurally stable caustic has its characteristic diffraction pattern, whose wave function has an integral representation in terms of the standard polynomial describing the corresponding catastrophe. We shall explain how these integrals arise as short wave asymptotic solutions of the wave equation. The diffraction catastrophes constitute a new hierarchy of functions, different from the special functions of analysis. Several are understood in detail; theory and experiment agree down to the finest scales. Important

aspects of diffraction near caustics are embodied in a series of universal scaling laws, giving exponents in power laws governing the increase in intensity and the decrease in diffraction fringe dimensions as the wavelength tends to zero.

Often, waves propagate paraxially in a uniform medium, from an initial wavefront which is a gentle deformation of a plane. In § 4 we consider paraxiality by describing the relevant differential geometry of wavefronts and their normals. Usually, textbooks confine themselves to second order properties of surfaces; we shall find it necessary to discuss umbilic points (where both principal curvatures are equal), whose classification involves third order properties.

In § 5 we apply catastrophe optics to irregular liquid droplet lenses. The existence of these droplets is a consequence of surface tension, but gravity affects their shape and can act as a parameter changing the topology of the caustics they produce by refraction (or reflection).

In § 6 we step up to the macroscale, to explore the caustics produced by an extended irregular wavefront such as emerges from the wavy surface of sunlit water or from randomly undulating glass (as used for bathroom windows). The gross structure of these caustic patterns is unintelligible until it is realised that they are made up of poorly resolved catastrophes. Once this is appreciated, it is possible to design and carry out experiments to reveal the complete caustic structure predicted theoretically.

The newest application of catastrophe optics is to random short waves, whose statistical properties are determined by the random caustic structure. In § 7 we show how some of these statistics can be calculated. The deepest results concern the wavelength dependence of the intensity moments of twinkling starlight, which involve the scaling laws of § 3 and depend on the whole infinite hierarchy of catastrophes.

In Appendix 1 we give a brief account of the complementary subject of singularities of wavefronts. These can be of two very different sorts: wavefront dislocations, which are singularities of the phase of the wave function, constituting structurally stable features of diffraction catastrophes, and Legendre singularities, which are singularities of the (fictitious) surfaces normal to the rays of geometrical optics.

In Appendix 2 we give a brief account of the connection between the catastrophe classification of caustics and the traditional classification of lens aberrations.

We have tried to make this review self-contained, so far as is possible. Readers will find catastrophe theory expounded in the excellent text by

POSTON and STEWART [1978] which complements the original, often rather speculative work of THOM [1972]; more mathematical treatments are given by ARNOLD [1975] and ZEEMAN [1977]. The relevant asymptotics of wave equations were discussed before catastrophe theory by MASLOV [1965] and LUDWIG [1966], and in terms of catastrophe theory by DUISTERMAAT [1974] and GUILLEMIN and STERNBERG [1977]. Catastrophe theory and some physical applications have been reviewed by GOLUBITSKY [1978] and POSTON [1980]. More elementary accounts of catastrophe optics are given by BERRY [1976, 1978a, 1979a].

§ 2. Catastrophe Theory and Stable Caustics

2.1. CAUSTICS ARE SINGULARITIES OF GRADIENT MAPS

In optics what matters is not so much individual rays as *families of rays* filling regions of space. The singularities of ray families are the caustics that this review will be concerned with. To prepare for the application of catastrophe theory we need to recall some standard properties of ray families. Let the rays (and waves) propagate in an inhomogeneous medium with refractive index $n = n(\mathcal{R})$ at the point \mathcal{R} . For simplicity of exposition we consider n independent of direction and time and consider only monochromatic light whose waves have frequency ω . This will be sufficient for later examples, and in any case the generalisation of catastrophe optics to anisotropic, time-varying, dispersive media involves no essentially new ideas.

A ray through the point \mathcal{R} is specified by its direction. It will prove convenient to express this in terms of the wave vector \mathbf{k} whose length k is given by the dispersion relation

$$k = n\omega/c \equiv \kappa n \quad (2.1)$$

where c is the speed of light, and κ the wave number, in vacuo. If τ measures arc length along the ray, the unit vector $\hat{\mathbf{k}}$ is

$$\hat{\mathbf{k}} \equiv \mathbf{k}/k = d\mathcal{R}/d\tau. \quad (2.2)$$

Of crucial importance to catastrophe optics is the fact that several rays in a family may pass through the point \mathcal{R} . These can be labelled by a suffix μ ; the wave vector on the μ th path will be denoted by \mathbf{k}_μ . The \mathbf{k}_μ make up vector fields over \mathcal{R} . By virtue of their origin in a wave equation

(§ 3) these vector fields must be derivable from scalar fields $\mathcal{S}_\mu = \mathcal{S}_\mu(\mathcal{R})$, defined (up to constants) by

$$\mathbf{k}_\mu = \kappa \nabla \mathcal{S}_\mu \quad (2.3)$$

(this implies that the \mathbf{k}_μ are irrotational). \mathcal{S}_μ are the branches of a multivalued action function \mathcal{S} which from (2.1) satisfies the Hamilton-Jacobi equation

$$|\nabla \mathcal{S}|^2 = n^2 \quad (2.4)$$

(LANCZOS [1966]). The action surfaces $\mathcal{S} = \text{constant}$ are the wavefronts of geometrical optics (Appendix 1); from (2.3) and (2.2) they are orthogonal to the rays, which therefore form a normal congruence (BORN and WOLF [1975] chap. 3).

The importance of the multivaluedness of \mathcal{S} lies in the fact that it is precisely on a caustic that two or more of its branches join. To see this, consider a point \mathcal{R} lying on a caustic of the ray family (Fig. 2.1a). By definition, a caustic is an envelope of the family, so as \mathcal{R} is approached (from the bright side of the caustic) two or more rays become parallel and hence by (2.2) their \mathbf{k}_μ coincide. By (2.3) this shows that \mathcal{R} is a branch point of \mathcal{S} . Also shown in Fig. 2.1 are the wavefronts, whose branches touch the caustic in cusps (Appendix 1). Fig. 2.1b shows a more complicated case where the caustic itself is cusped. Outside the caustic there is one ray through each point, and \mathcal{S} is single valued. Inside the caustic there are three rays and \mathcal{S} is three valued. Each arm of the caustic curve is an envelope where rays touch in pairs, while at the cusp point itself three rays touch.

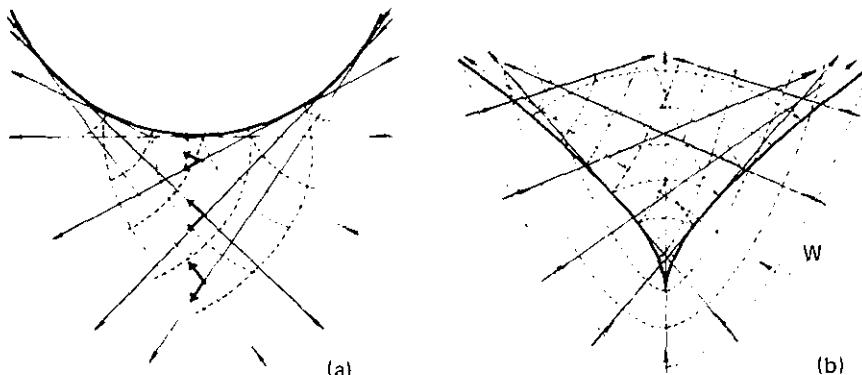


Fig. 2.1. Rays (full lines), wavefronts (dashed lines) and caustics (bold lines). (a) Smooth caustic curve at each of whose points two rays touch; wave vectors \mathbf{k} are shown by bold arrows; (b) cusped caustic formed by propagation of initial wavefront W .

Sometimes infinitely many rays pass through each point in some region. Such behaviour is very important in mechanics, and corresponds to stochastic motion (BERRY [1978b]); but it is not commonly encountered in optics and moreover does not involve catastrophe theory. Therefore we shall not consider this case here.

In catastrophe theory the multivaluedness of \mathcal{S} is tamed by embedding this function in a single-valued function with extra variables, in the following manner. To find the action \mathcal{S} for a ray family, eq. (2.4) must be solved under some boundary condition. A general boundary condition is to specify \mathcal{S} on a surface Σ (not necessarily a wavefront). If spatial position on Σ is denoted by \mathcal{R}_0 , this boundary condition is

$$\mathcal{S} = \mathcal{S}(\mathcal{R}_0) = \text{given on } \Sigma. \quad (2.5)$$

The corresponding solution of (2.3) is obtained in a convenient form from the following version of Fermat's principle of extremal time (BORN and WOLF [1975] chap. 3, LANCZOS [1966]).

Consider all paths P consistent with the given conditions, i.e. starting on Σ and passing through \mathcal{R} (Fig. 2.2), and specify each P by position as a function of arc length, $\mathcal{R}(\tau)$. For each P form the path integral

$$\begin{aligned} \phi(\mathcal{R}) &= \mathcal{S}(\mathcal{R}_0) + \frac{1}{\kappa} \int_{\mathcal{R}_0}^{\mathcal{R}} \mathbf{k}(\mathcal{R}(\tau)) \cdot d\mathcal{R}, \\ \text{i.e.} \quad \phi(\mathcal{R}) &= \mathcal{S}(\mathcal{R}_0) + \int_0^{\tau(\mathcal{R})} n(\mathcal{R}(\tau)) d\tau, \end{aligned} \quad (2.6)$$

which is simply the optical distance along P from Σ to \mathcal{R} , i.e. the time multiplied by c . Then the branches of the action \mathcal{S} are the extreme values of ϕ in the space of paths, i.e. the values of ϕ for paths P satisfying

$$\delta \left[\mathcal{S}(\mathcal{R}_0) + \int_0^{\tau(\mathcal{R})} n(\mathcal{R}(\tau)) d\tau \right] = 0, \quad (2.7)$$

where δ denotes first order path variations. The extreme paths are the rays through \mathcal{R} . If the ray family has no caustics between Σ and \mathcal{R} , \mathcal{S} is single valued; otherwise \mathcal{S} is multivalued.

We remark in passing that the Euler–Lagrange equations from (2.7) yield the ray equations

$$\frac{d}{d\tau} \left(n(\mathcal{R}(\tau)) \frac{d\mathcal{R}(\tau)}{d\tau} \right) = \nabla n(\mathcal{R}(\tau)). \quad (2.8)$$

These show how refractive index variations alter the ray direction $d\mathcal{R}/d\tau$; the equations reduce to Snell's law if the medium is stratified.

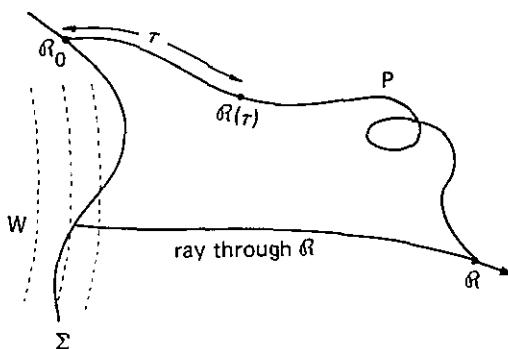


Fig. 2.2. A path P from the point \mathcal{R}_0 on an initial surface Σ , together with a ray from Σ to \mathcal{R} , in a medium of refractive index $n(\mathcal{R})$. Wavefronts W near Σ are shown as dashed lines.

Now, although \mathcal{S} may be multivalued (and in cases of interest, where there are caustics, it will be), the function $\phi(\mathcal{R})$ is single valued for a given path P . Therefore by specifying P with variables $s = (s_1, s_2 \dots)$ the variational equation (2.7) can be expressed as the *vanishing of the gradient of a single-valued function* $\phi = \phi(s; \mathcal{R})$:

$$\frac{\partial \phi}{\partial s_i} = 0 \quad \forall i. \quad (2.9)$$

where i is an index labelling the separate variables $s_1, s_2 \dots$. By these gradient conditions, s and hence P are fixed for those rays k passing through \mathcal{R} .

The s can be chosen in many ways. For example, the path P from \mathcal{R}_0 to \mathcal{R} can be divided into infinitely many segments (Fig. 2.3) with three of the s assigned to the starting point of each. Alternatively, the function $\mathcal{R}(\tau)$ could be Fourier-analyzed for each P , and s taken as the Fourier coefficients. These procedures result in infinitely many s .

However, it is possible to resolve the multivaluedness of \mathcal{S} by choosing paths with a *finite* number of s ; then, of course, the segments of P must themselves be rays. In the simplest procedure of all, P is chosen with a single segment, namely the ray from \mathcal{R}_0 to \mathcal{R} (Fig. 2.4), and only two variables $s = (s_1, s_2)$ are required, corresponding to coordinates of \mathcal{R}_0 on



Fig. 2.3. Path P with many segments, from \mathcal{R}_0 on Σ to \mathcal{R} .

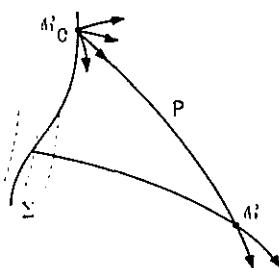


Fig. 2.4. Path P through \mathcal{R} in the family of rays emanating from \mathcal{R}_0 on Σ , together with an extreme ray from Σ to \mathcal{R} . Wavefronts near Σ are shown as dashed lines.

Σ . This will result in a single-valued path integral $\phi(s_1, s_2; \mathcal{R})$, provided only one ray from \mathcal{R}_0 passes through \mathcal{R} – in other words, provided there are no caustics between \mathcal{R}_0 and \mathcal{R} in the family of rays emanating from \mathcal{R}_0 , a condition certainly satisfied for the important case of homogeneous media where the rays are straight, and also in refracting media if \mathcal{R} is not too far from Σ .

In a sufficiently extended refracting medium this procedure may fail by producing a multivalued ϕ . To deal with such cases an alternative method must be employed to construct a single-valued ϕ ; one such method, involving Lagrangian manifolds in the phase space \mathcal{R} , k , was devised by MASLOV [1965] (see also KRAVTSOV [1968]), and will be briefly discussed in § 3.2.

The conclusion from this discussion is that there are a number of ways of representing the branches \mathcal{S}_μ of a multivalued action function $\mathcal{S}(\mathcal{R})$ as the extreme values of a single-valued function $\phi(s; \mathcal{R})$. It helps to think of ϕ for fixed \mathcal{R} as the height of a landscape above a plane with coordinates s . Then the coordinates $s^\mu(\mathcal{R})$ of the extrema (hilltops, minima, saddles) satisfying (2.9) correspond to the rays, and the heights

$$\mathcal{S}_\mu = \phi(s^\mu(\mathcal{R}); \mathcal{R}) \quad (2.10)$$

give the actions.

In the language of catastrophe theory, $\phi(s; \mathcal{R})$ is a *generating function*, and (2.9) defines a *gradient map* from \mathcal{R} to s . The s are called *state variables*. The components of \mathcal{R} are quantities on which the rays depend, and appear as arguments in the action \mathcal{S} (eq. (2.10)); there may be other quantities on which the rays depend, like parameters describing the form of Σ or the action on it (e.g. angles specifying the orientation of a refracting surface relative to an incident beam, or the time in a medium

that is changing). All such quantities are termed *control parameters* and will henceforth be denoted by $C = (C_1, C_2 \dots)$, and the generating function written as $\phi(s; C)$.

Now let C vary, for example by changing the observation point \mathcal{R} . The extrema $s^\mu(C)$ (rays) will change. When two or more extrema coalesce, the map (2.9) is singular, and ϕ is stationary to higher than first order. But we have already seen that coalescence of rays occurs on caustics; therefore *caustics correspond to singularities of gradient maps*.

The condition for C to lie on a caustic is obtained in terms of the generating function as follows: higher order stationarity implies the existence of a set of displacements ds_i , away from the extrema s^μ , for which (2.9) continues to hold, i.e. for which

$$\sum_i ds_i \frac{\partial^2 \phi}{\partial s_i \partial s_j} = 0. \quad (2.11)$$

The condition for these homogeneous equations to have a solution, and hence for C to lie on a caustic, is that the Hessian

$$\mathcal{H}[\phi] \equiv \det \frac{\partial^2 \phi}{\partial s_i \partial s_j} = 0 \quad (2.12)$$

at points $s^\mu(C)$ where (2.9) holds.

2.2. THE CLASSIFICATION OF CATASTROPHES

In catastrophe optics we start from a smooth function ϕ of state variables s and control parameters C , construct the gradient map (2.9), and enquire about the form of its singularities, defined by eq. (2.12), in the control space with coordinates C . We have seen that these singularities are the caustics.

We seek a description of the *structurally stable singularities*. To explain this important concept, consider the singularity S_1 produced by a generating function ϕ_1 . Now let ϕ_1 be perturbed to ϕ_2 (in optics this perturbation might be produced by altering the initial surface Σ , or the propagation medium); the singularity will change to S_2 . Structural stability means that S_1 and S_2 are related by a diffeomorphism of C , that is a smooth reversible change of control parameters. Alternatively, imagine S_1 in a control space made of rubber; then it is structurally stable if it can be

transformed into S_2 by smooth deformation of the rubber (without tearing or self-overlap or infinite strain).

Structural stability is typical, provided the dimensionality of control space does not exceed seven (in which case the definition of structural stability must be modified somewhat – see below). This follows from the mathematics (ZEEMAN [1977]) where it is shown that unstable singularities (e.g. point foci) come from a subset of measure zero in the space of all generating functions. In other words, structural stability is a *generic* property of caustics.

Structurally stable singularities are divided into equivalence classes; all members of a class can be deformed into one another by diffeomorphism. *It is these equivalence classes of singularity that constitute the catastrophes.* Each is described by a standard polynomial, or *normal form*, henceforth denoted by Φ , for the generating function ϕ . In its simplest version, the main classification theorem of catastrophe theory is a list of these standard polynomials. Each polynomial gives a description, up to diffeomorphism, of all singularities in its equivalence class.

The catastrophes are classified by their *codimension* K . K is the dimensionality of control space (i.e. the number of parameters C) minus the dimensionality of the singularity itself. Alternatively, K is the dimensionality of the subspace that must be explored to encounter the singularity. For example, a caustic point in one space dimension, a caustic curve in two dimensions, and a caustic surface in three dimensions, all have codimension 1.

All catastrophes with $K \leq 4$ (the “elementary catastrophes”) are listed in Table 1, with the original names due to THOM [1972] and also the symbols of ARNOL'D [1975]. The subscripts on Arnol'd's symbols denote the multiplicity of the catastrophe, i.e. the number of extrema of Φ (rays)

TABLE 1
Standard polynomials Φ for the elementary catastrophes with codimension $K \leq 4$

Name	Symbol	K	$\Phi(s; C)$
fold	A_2	1	$s^3/3 + Cs$
cusp	A_3	2	$s^4/4 + C_2 s^2/2 + C_1 s$
swallowtail	A_4	3	$s^5/5 + C_3 s^3/3 + C_2 s^2/2 + C_1 s$
elliptic umbilic	D_4	3	$s_1^3 - 3s_1 s_2^2 - C_3(s_1^2 + s_2^2) - C_2 s_2 - C_1 s_1$
hyperbolic umbilic	D_4^*	3	$s_1^3 + s_2^3 - C_3 s_1 s_2 - C_2 s_2 - C_1 s_1$
butterfly	A_5	4	$s^6/6 + C_4 s^4/4 + C_3 s^3/3 + C_2 s^2/2 + C_1 s_1$
parabolic umbilic	D_5	4	$s_1^4 + s_1 s_2^2 + C_4 s_2^2 + C_3 s_1^2 + C_2 s_2 + C_1 s_1$

which coalesce at the most singular point. For a proof of the classification theorem, see ZEEMAN [1977]; for a detailed account of the ideas underlying the proof, see POSTON and STEWART [1978].

Each polynomial consists of two parts: a *germ*, involving only the state variables s , and *unfolding terms* that also involve the controls C (which appear linearly). The germ describes the most singular extremum of the Φ landscape, produced by setting $C = 0$. The unfolding terms describe how the degenerate extremum generifies into a collection of $K + 1$ ordinary extrema as the controls are altered.

Fig. 2.5 shows the topology in control space of the catastrophes in Table 1. The way the caustics unfold away from their singularities is evident: the fold point (Fig. 2.5a) is an isolated singularity; the cusp point (Fig. 2.5b) unfolds into two fold curves; the swallowtail point (Fig. 2.5c) unfolds into two cusped edges (*ribs*) connected by three fold surfaces with a line of self-intersection; the elliptic umbilic point (Fig. 2.5d) unfolds symmetrically into three ribs joined by three fold surfaces, etc. The phenomenon of higher catastrophes unfolding into configurations of lower ones is general (ARNOL'D [1975]).

These catastrophes give the elemental “atomic forms” that stable caustics can adopt. It is important to appreciate that the classification is complete: for $K \leq 4$ all forms other than those in Table 1 are unstable. In this sense the theorem is surprising, for it is far from obvious that in a plane, for example, a cusp point is stable whereas a finite-angled corner is not. In what follows we shall give abundant illustration of the usefulness of the theorem in optics. For the most part we shall confine our attention to Thom's original list as in Table 1 and Fig. 2.5, but we shall occasionally venture into ARNOL'D's [1975] enormous extension of the list for $K > 4$. The novel feature of this extension is that for $K > 7$ it contains catastrophes which are stable not under diffeomorphism but only under homeomorphism, which is a weaker form of equivalence; the optical significance of this difference is being studied in this laboratory (see §§ 5 and 6).

The polynomials Φ in Table 1 are written in terms of the fewest state variables s necessary to describe the singularity. This is the *corank* of the catastrophe. The *cuspoids* A_{2-5} have corank 1, and the *umbilics* D_4^\pm , D_5 have corank 2. It is always possible to add extra state variables s , in terms that are at most quadratic, without affecting the singularity; the extrema of ϕ are non-degenerate in these extra s -directions, and it is easy to verify directly that the new terms leave the Hessian equation (2.12) unaltered.

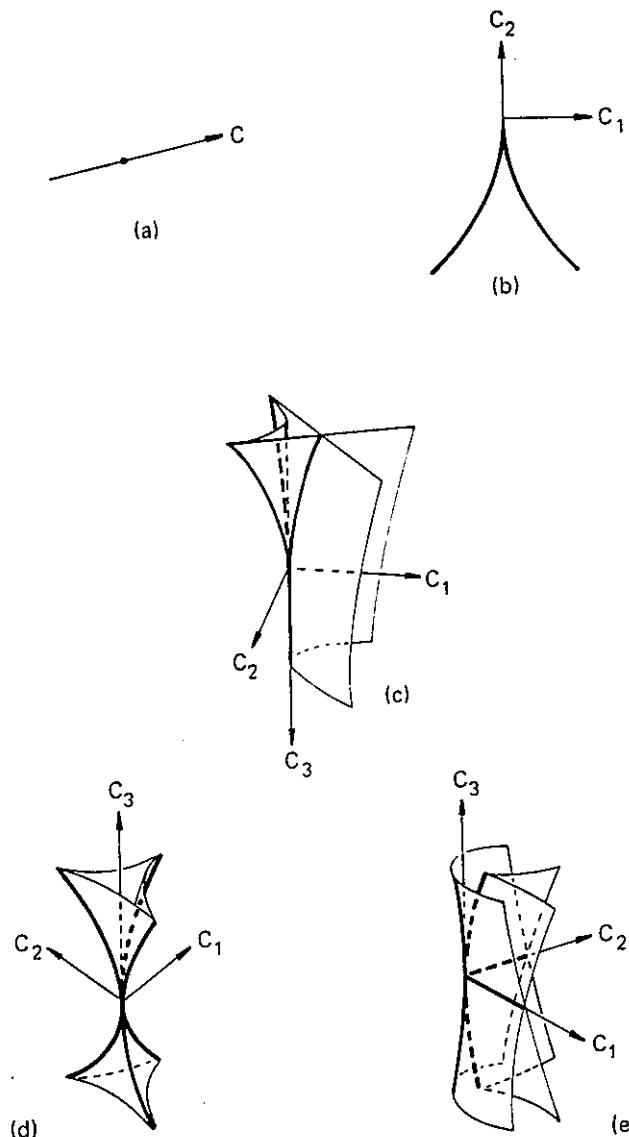


Fig. 2.5. Topology (in control space) of the elementary catastrophes: (a) fold; (b) cusp; (c) swallowtail; (d) elliptic umbilic; (e) hyperbolic umbilic; (f) butterfly; (g) parabolic umbilic.

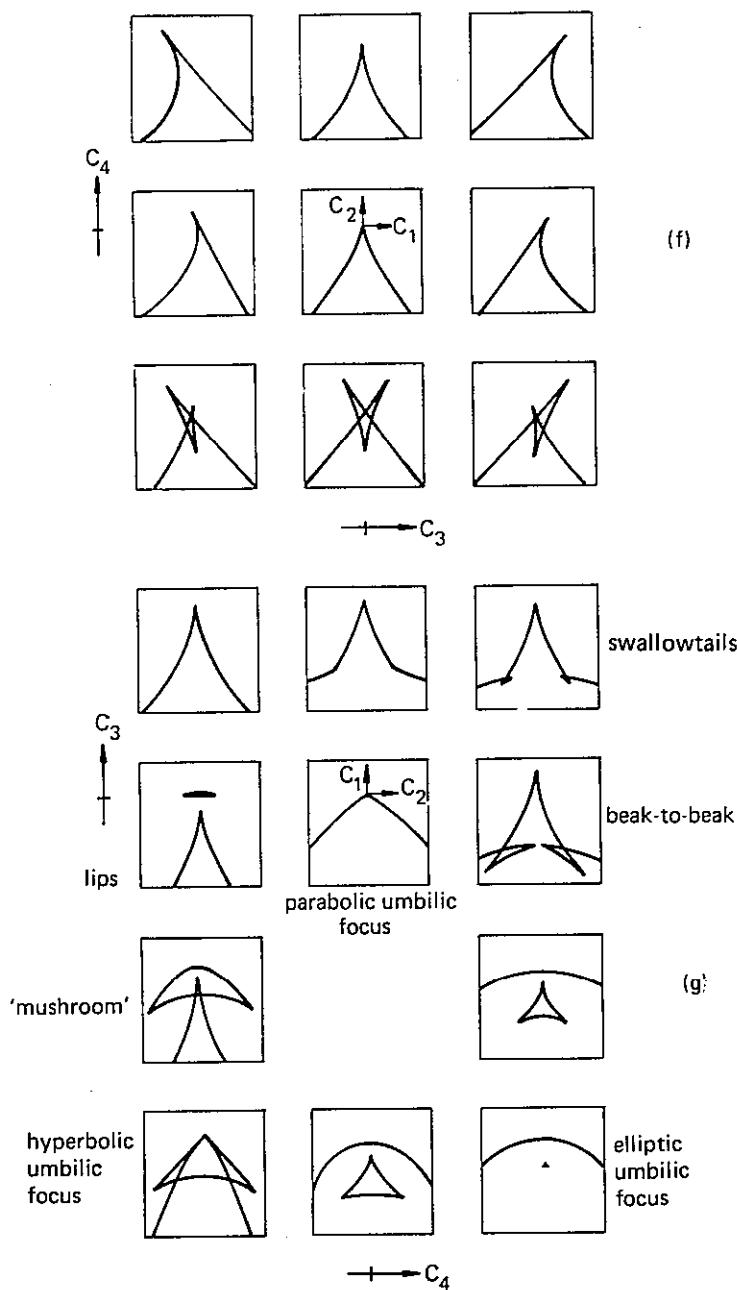
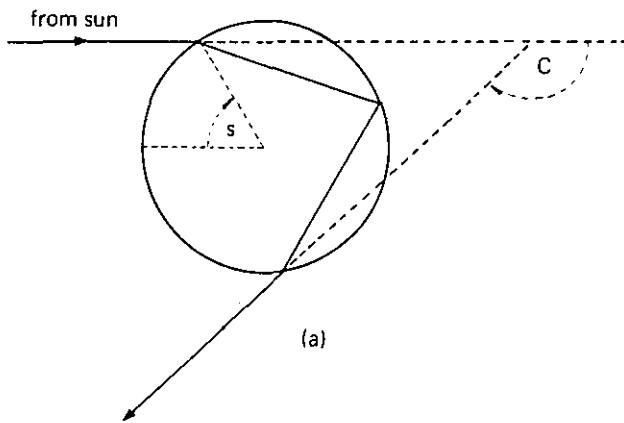


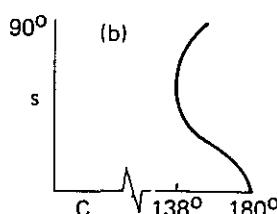
Fig. 2.5 (contd.)

2.3. EXAMPLES OF THE ELEMENTARY CATASTROPHES IN OPTICS

In the *rainbow* (DESCARTES [1637], HUMPHREYS [1940], TRICKER [1970], KHARE and NUSSENZVEIG [1974], NUSSENZVEIG [1969a,b, 1977]), parallel sunrays hit a raindrop (Fig. 2.6a), suffer two refractions and one internal reflection, and are viewed from afar. Because of the rotational symmetry there is only one control parameter C , which can be taken as the angle of deflection of the rays, and there need be only one state variable s , which can be taken as the latitude of the original point of incidence. Elementary optical laws show the graph of $s(C)$ (Fig. 2.6b) to be multivalued: there are two rays if $C > C_0 = 138^\circ$ (for orange light) and none if $C < C_0$. However, the graph of $C(s)$ is single valued and has a minimum at $s = 59^\circ$. The rays form a directional caustic at C_0 and this is a *fold* catastrophe, which is indeed the only stable singularity with codimension 1. In space the caustic surface is asymptotic to a cone with semiangle



(a)



(b)

Fig. 2.6. (a) State variable s and control parameter C for a ray contributing to the rainbow;
 (b) graph of s against C , showing multivaluedness.

$180^\circ - C_0 = 42^\circ$; very close to the drop, the caustic forms a more complicated surface (McDONALD [1963]).

In the *sparkling of sunlight on the sea* (Fig. 2.7) (LONGUET-HIGGINS [1960a]), the eye sees a number of reflected images of the sun, corresponding to specular points, where the wavy water surface is so oriented as to reflect rays into the eye. Now the state variables may be taken as two coordinates $s = (s_1, s_2)$ on the water surface, and the control parameters C are the time t of observation, and the positions \mathcal{R}_E and \mathcal{R}_S of the eye and sun. The generating function $\phi(s_1, s_2; t, \mathcal{R}_E, \mathcal{R}_S)$ is simply proportional to the distance from the sun via s on the sea to the eye along straight paths, and it is then obvious by elementary reasoning that the extrema occur at the specular points s . Over short times, \mathcal{R}_E and \mathcal{R}_S may be considered fixed, so that the only effective control is t , which parameterises the changing form of the water surface. Therefore only fold catastrophes occur stably, and correspond to annihilations and births of pairs of images at particular instants, called "twinkles". The vanishing of the determinant (2.12) at a twinkle means that the water surface has momentarily not merely the right slope to direct light into the eye but also the right curvature to focus it there; in space there is a moving caustic surface which passes through the eye in a twinkle.

The inner surface of an *illuminated tea-cup* produces a caustic surface

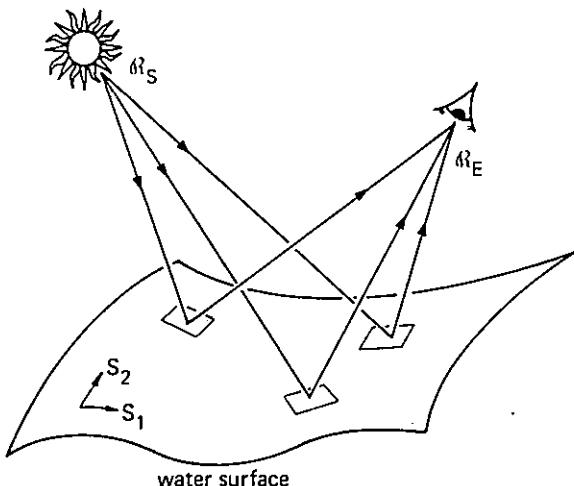


Fig. 2.7. Three sunrays from specular points s_1, s_2 on a water surface. The total length of each ray is proportional to one of the values of the action function S for the family of rays from the sun to the eye.

in space, whose intersection with the tea surface is visible by diffuse reflection as a caustic curve containing a *cusp* point (Fig. 2.5b). This is stable because there are two controls, corresponding to coordinates on the tea, and the stable catastrophes in two dimensions are just fold curves and cusp points.

For a more systematic illustration of the elementary catastrophes we use photographs of caustics produced by reflection from plastic film held in a rigid framework and formed into a curved surface by suction. In the plane of a photograph the only stable caustics are fold curves and cusp points, as in Fig. 2.8. These are sections of fold surfaces and cusped edges (ribs) in three-dimensional space. If a rib is tangent to one of a sequence of plane sections, this gives rise to characteristic *lips* (Fig. 2.9a) or *break-to-break* events (Fig. 2.9b), depending on the sense in which the rib curves relative to the caustic sheets. These events are not catastrophes of codimension three, because spatial points where they occur are not unique but depend on the orientation of the sequence of plane sections.

However, catastrophes with $K=3$ do occur stably at isolated points in the space filled by light reflected by the plastic film, and characteristic sections of unfoldings, close to these points, are shown on Figs. 2.10(a-c)

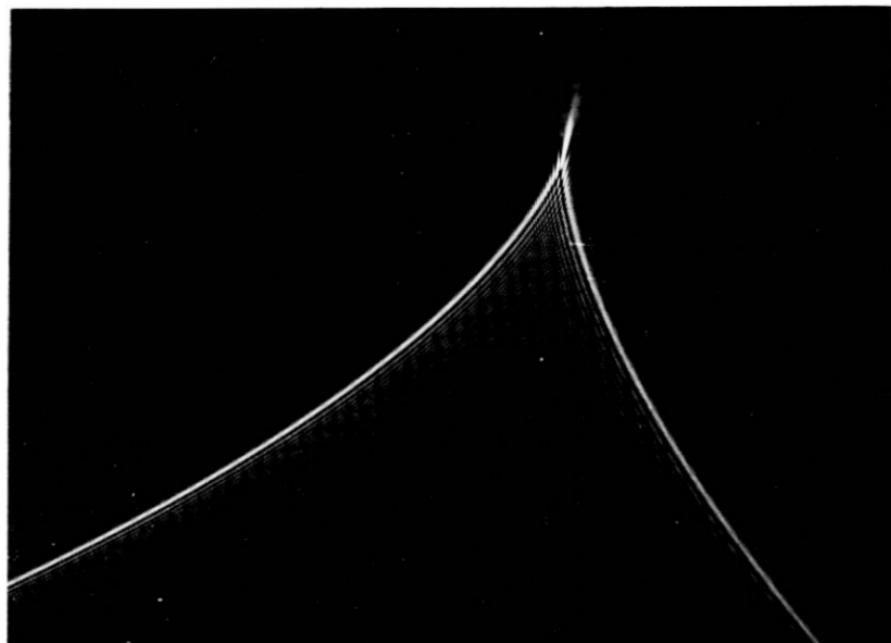


Fig. 2.8. Two fold caustics meeting at a cusp point (courtesy of P. N. Kesterton and D. B. White).

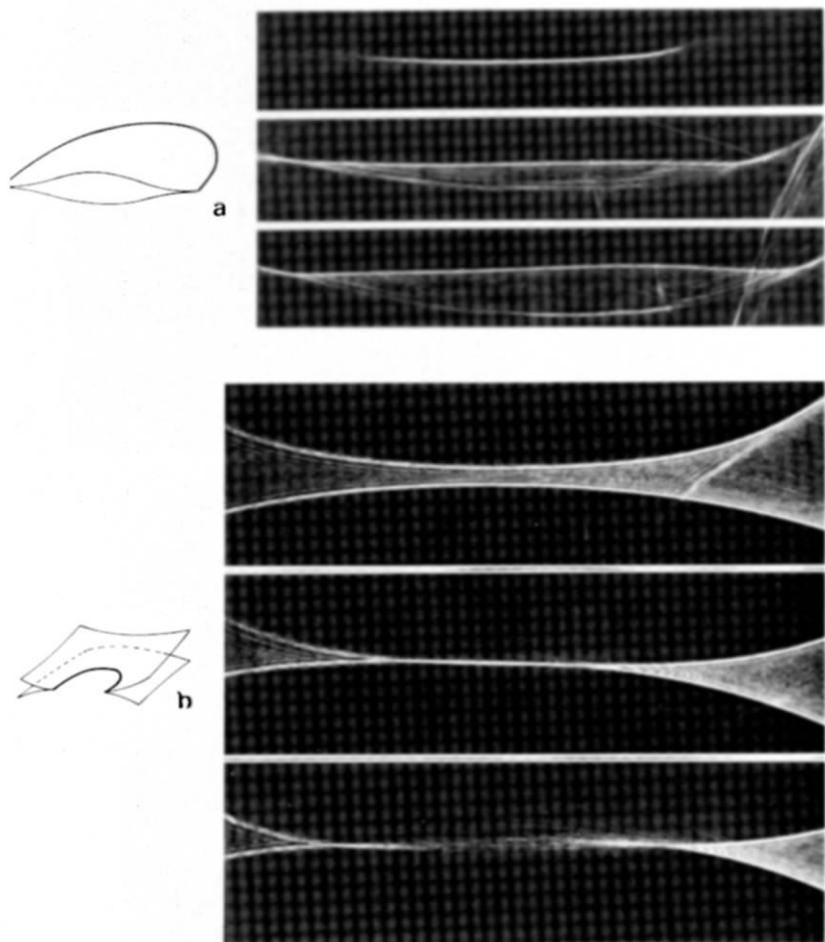


Fig. 2.9. Non-catastrophic events produced when a rib (cusped edge) is tangent to a plane of observation: (a) lips; (b) beak-to-beak (courtesy of P. N. Kesterton and D. B. White).

(cf Figs. 2.5 c–e). Altering the shape of the plastic film gives extra control parameters enabling higher catastrophes to be explored. Fig. 2.10d shows an unfolding of the butterfly, and Figs. 2.10 e and f two unfoldings of the parabolic umbilic, produced in this way (cf Figs. 2.5 f and g).

Beautiful catastrophes in focused electron beams, produced by altering the controls of an electron microscope, can be seen in papers by SCHEFFELS, HAHN and LENZ [1953], LIESEGANG [1953], BARTZ, WEISSENBERG and WISKOTT [1956] and SCHLEICH, HOFFMEISTER, KOOPS and LENZ [1968] (see also GLASER [1956]).

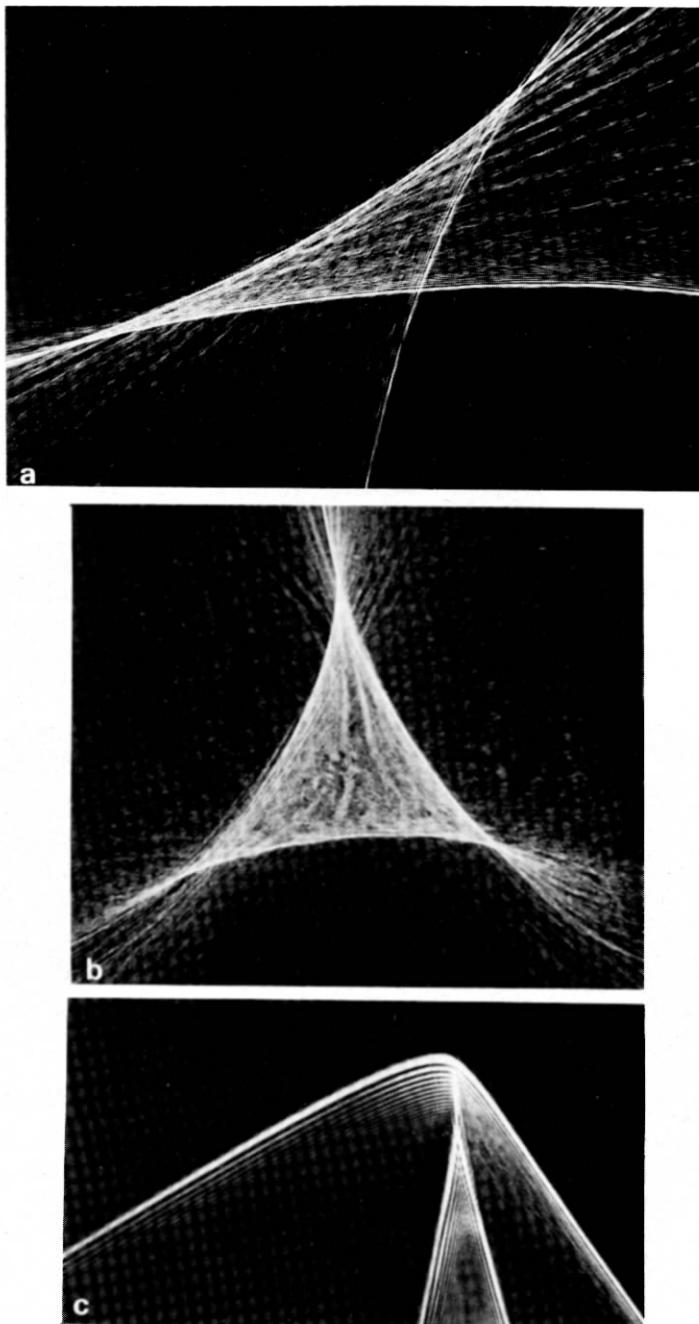


Fig. 2.10. Characteristic sections of caustics close to singular points of (a) swallowtail; (b) elliptic umbilic; (c) hyperbolic umbilic; (d) butterfly; (e) parabolic umbilic (elliptic umbilic piercing fold); (f) parabolic umbilic ("mushroom"). ((a), (b), (d) and (e) courtesy of P. N. Kesterton and D. B. White; (c) and (f) courtesy of S. Q. Hooker and J. C. Rugg.)

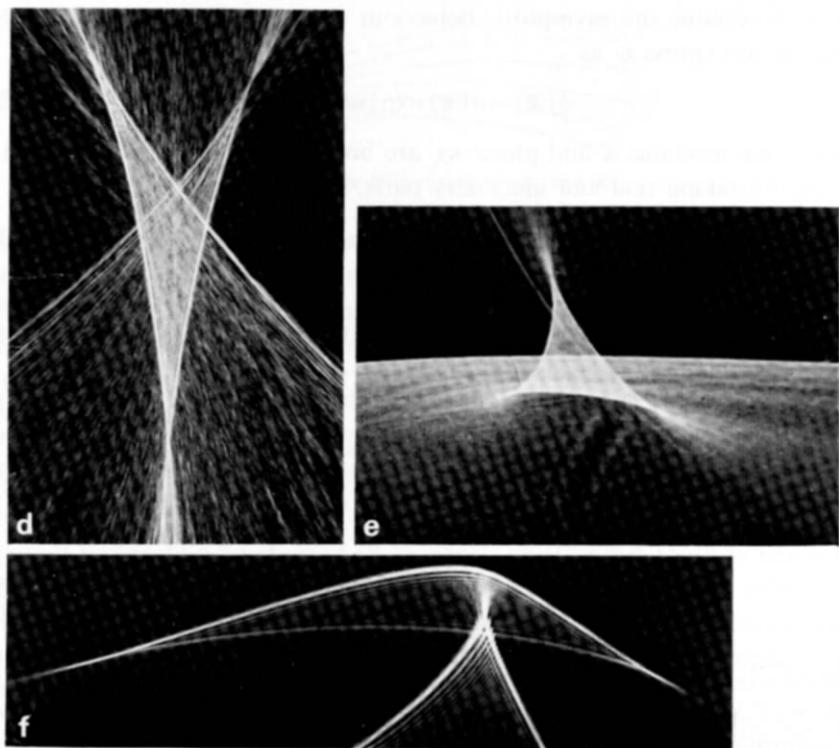


Fig. 2.10 (contd.)

§ 3. Diffraction Catastrophes

3.1. ASYMPTOTIC SOLUTION OF THE WAVE EQUATION

We shall discuss diffraction in terms of the scalar Helmholtz equation:

$$\nabla^2 \psi(\mathcal{R}) + \kappa^2 n^2(\mathcal{R}) \psi(\mathcal{R}) = 0, \quad (3.1)$$

where $\psi(\mathcal{R})$ is a complex scalar wave function. Catastrophe optics is concerned with the approach to the short-wave limit $\kappa \rightarrow \infty$. If the propagation medium is homogeneous ($n = \text{constant}$) the use of scalar theory has been rigorously justified (GREEN and WOLF [1953], WOLF [1959, 1960]). For inhomogeneous media the equations for each component of the electromagnetic field vectors (BORN and WOLF [1975] chap. 1) are not of the form (3.1) but contain extra terms; in the geometrical optics limit these terms describe polarisation effects (BORN and WOLF [1975] chap. 3), which we shall ignore.

To determine the asymptotic behaviour as $\kappa \rightarrow \infty$, the standard procedure is to express ψ as

$$\psi(\mathcal{R}) = a(\mathcal{R}) \exp [i\kappa\chi(\mathcal{R})], \quad (3.2)$$

where the modulus a and phase $\kappa\chi$ are both real. Substitution into (3.1) gives, on taking real and imaginary parts,

$$|\nabla\chi|^2 = n^2 + \nabla^2 a/\kappa^2 a \quad (3.3)$$

and

$$\nabla \cdot (a^2 \nabla\chi) = 0. \quad (3.4)$$

The lowest-order approximation is to neglect the term $\nabla^2 a/\kappa^2 a$ in (3.3). Then χ satisfies the Hamilton-Jacobi equation (2.4), showing that the phase of the wave is κ times the action \mathcal{S} of the underlying ray family. To solve equation (3.4) for the wave intensity a^2 , Gauss's theorem is applied to a thin tube of rays passing through \mathcal{R} (Fig. 3.1), and gives

$$a^2(\mathcal{R}) = a^2(\mathcal{R}_0) \frac{d\mathcal{A}(\mathcal{R}_0)n(\mathcal{R}_0)}{d\mathcal{A}(\mathcal{R})n(\mathcal{R})}, \quad (3.5)$$

where \mathcal{R}_0 is a point on the ray through \mathcal{R} , and $d\mathcal{A}(\mathcal{R}_0)$ and $d\mathcal{A}(\mathcal{R})$ are areas of wavefronts cutting the ray tube at \mathcal{R}_0 and \mathcal{R} . The simplest asymptotic formula for ψ is therefore

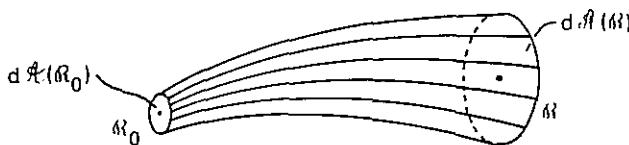
$$\psi(\mathcal{R}) \approx a(\mathcal{R}_0) \left[\frac{n(\mathcal{R}_0) d\mathcal{A}(\mathcal{R}_0)}{n(\mathcal{R}) d\mathcal{A}(\mathcal{R})} \right]^{\frac{1}{2}} \exp [i\kappa\mathcal{S}(\mathcal{R})]. \quad (3.6)$$

For propagation from an initial surface Σ on which ψ is given, \mathcal{R}_0 should be chosen to lie on Σ ; \mathcal{S} is determined as explained in § 2.1, and $d\mathcal{A}(\mathcal{R}_0)/d\mathcal{A}(\mathcal{R})$ can be written as the Jacobian determinant between coordinates on Σ and coordinates on the wavefront through \mathcal{R} (KLINE and KAY [1965]).

If the ray family has caustics between Σ and \mathcal{R} , \mathcal{S} is multivalued. In this case ψ can be represented as a sum of interfering terms of the form (3.6), one for each ray through \mathcal{R} :

$$\psi(\mathcal{R}) \approx \sum_{\mu} a_{\mu}(\mathcal{R}) \exp [i(\kappa\mathcal{S}_{\mu}(\mathcal{R}) - m_{\mu}(\mathcal{R})\pi/2)]. \quad (3.7)$$

In this formula a_{μ} is the amplitude along the μ th ray, constructed according to (3.5), and $m_{\mu}(\mathcal{R})$ is the number of times the μ th ray has touched a caustic en route from Σ to \mathcal{R} . The origin of the phase

Fig. 3.1. Tube of rays through \mathcal{R}_0 and \mathcal{R} .

factor $\exp[-im_\mu\pi/2]$ will emerge in § 3.2; a simple way of understanding it on the basis of (3.6) is to consider $d\mathcal{A}(\mathcal{R})$ as having a simple zero when \mathcal{R} lies on a typical (fold catastrophe) caustic, so that ψ gains a factor $\exp[-i\pi/2]$ each time \mathcal{R} passes through a caustic. The systematic exploitation of the “interfering rays” formula (3.7) is the geometrical theory of diffraction of KELLER [1958].

As an asymptotic approximation to ψ , (3.7) has a serious deficiency: *it fails completely on caustics.* On a caustic, the ray tube area $d\mathcal{A}$ shrinks to zero, so (3.6) predicts that ψ has a singularity of infinite intensity. This cannot represent a solution of (3.1), which must be a smooth function of \mathcal{R} for all finite κ , however large. The error in deriving (3.7) was to neglect the term $\nabla^2 a/\kappa^2$ in (3.3), which in this approximation diverges on caustics no matter how large κ is. This divergence is indeed a catastrophe, and justifies using the term in the context of short-wave optics.

The fact that (3.7) fails at the most important places, where the light is brightest, suggests that the remedy lies not in supplementing (3.7) by correction terms in falling powers of κ (see e.g. AHLUWALIA, LEWIS and BOERSMA [1968]), but in replacing it by an expression whose large- κ analytic properties near caustics are completely different. Indeed we shall see that the diffraction near a caustic depends on its catastrophe type.

3.2. INTEGRAL REPRESENTATIONS FOR SHORT WAVES

To obtain asymptotic representations that do not diverge on caustics, the principle of superposition is invoked, to express ψ as an integral over a collection of contributions of the form (3.6). Each member of the collection is labelled by the values of variables $s = (s_1, s_2 \dots)$, and has action function, denoted by $\phi(s; \mathcal{R})$, chosen to satisfy two conditions. First, ϕ is a single-valued function of \mathcal{R} for fixed s . Second, the values $S_\mu(\mathcal{R})$ of the actions at \mathcal{R} for the ray family under study are the extreme values of $\phi(s; \mathcal{R})$ as s is varied. Therefore ϕ is precisely the generating function of catastrophe theory introduced in § 2.1, and the rays are given

by the gradient map (2.9). The integral representation for ψ is then

$$\psi(\mathcal{R}) = e^{-iN\pi/4} \left(\frac{\kappa}{2\pi} \right)^{N/2} \int \cdots \int d^N s b(s; \mathcal{R}) \exp [i\kappa\phi(s; \mathcal{R})], \quad (3.8)$$

where N is the number of state variables s , and the prefactors have been chosen for later convenience.

Before discussing explicit realisations of this procedure, and the choice of weighting function b , we establish the connection between (3.8) and the simple ray approximation (3.7). This relies on the principle of stationary phase, according to which the main contributions to (3.8) for given \mathcal{R} come from the stationary points, that is from those values $s^*(\mathcal{R})$ of the integration variables for which the gradient map (2.9) holds. The principle can be applied because we are studying the limit $\kappa \rightarrow \infty$ for which the integrand is a rapidly oscillating function of s , so that other than near stationary points, destructive interference occurs. If \mathcal{R} is not near a caustic the stationary points are well separated and the method of stationary phase can be used in its simplest form (see e.g. appendix III of BORN and WOLF [1975]). This gives a series of separate contributions with phases $\kappa\mathcal{S}_\mu$, precisely as in (3.7). Near a caustic, two or more of the stationary points lie close together and their contributions cannot be separated. The nature of this failure of simple stationary phase, and the resulting description of diffraction near caustics in terms of catastrophes will be discussed in § 3.3.

First, however, we point out that because there are many ways of choosing the function ϕ , as described in § 2.1, there are, correspondingly, many integral representations (3.8) for ψ ; all are equivalent as $\kappa \rightarrow \infty$. We discuss in detail the diffraction integral resulting from the choice $\phi(s_1, s_2; \mathcal{R})$ described in § 2.1; ϕ is the optical path along the ray from \mathcal{R}_0 on the initial surface Σ to \mathcal{R} . For convenience we now take Σ to be a wavefront. Then (3.8) has the explicit form

$$\begin{aligned} \psi(\mathcal{R}) = & \frac{-i\kappa}{2\pi} \int ds_1 \int ds_2 a(\mathcal{R}_0(s)) \left[\frac{n(\mathcal{R}_0(s))}{n(\mathcal{R})} \det \frac{\partial^2 \phi(s; \mathcal{R})}{\partial s_i \partial X_j} \right]^{\frac{1}{2}} \\ & \times \exp [i\kappa\phi(s; \mathcal{R})], \end{aligned} \quad (3.9)$$

where the variables X_j in the 2×2 matrix of mixed second derivatives of ϕ denote coordinates in the wavefront through \mathcal{R} of the family of rays emanating from \mathcal{R}_0 .

To see that (3.9) is correct, consider the case where \mathcal{R} is not near a

caustic and apply the method of stationary phase. The contributions s^μ correspond to the points \mathcal{R}_0 from which rays through \mathcal{R} originate on Σ . Standard formulae (e.g. appendix III of BORN and WOLF [1975]) give

$$\psi(\mathcal{R}) \approx \sum_\mu a(\mathcal{R}_0(s^\mu)) \left[\frac{n(\mathcal{R}_0(s^\mu)) \det \partial^2 \phi / \partial s_i \partial X_j}{n(\mathcal{R}) \det \partial^2 \phi / \partial s_i \partial s_j} \right]^{\frac{1}{2}} \times \exp [i(\kappa \mathcal{S}_\mu(\mathcal{R}) - m_\mu(\mathcal{R})\pi/2)], \quad (3.10)$$

where m_μ is the number of negative eigenvalues of the matrix $\partial^2 \phi / \partial s_i \partial s_j$. The ratio of determinants is

$$\frac{\det \partial^2 \phi / \partial s_i \partial X_j}{\det \partial^2 \phi / \partial s_i \partial s_j} = \det \frac{ds_i}{dX_j} = \frac{d\mathcal{A}(\mathcal{R}_0)}{d\mathcal{A}(\mathcal{R})}, \quad (3.11)$$

which is precisely the Jacobian between ray tube area elements in (3.5). Prior to encountering a caustic the path length ϕ has a minimum, so both eigenvalues of $\partial^2 \phi / \partial s_i \partial s_j$ are positive, and $m = 0$ for the (unique) path through \mathcal{R} . On encountering a caustic, $d\mathcal{A}(\mathcal{R})$ passes through zero and ϕ has a saddle in s_1, s_2 , so that one eigenvalue has changed sign and $m = 1$. Therefore the integer $m_\mu(\mathcal{R})$ in (3.10) is the same as that in (3.7). This completes the justification of the integral representation (3.9) by identifying its off-caustic approximation (3.10) with the simple asymptotic formulae (3.6) and (3.7).

The diffraction integral (3.9) is of Kirchhoff type, giving $\psi(\mathcal{R})$ as a superposition of wavelets from points \mathcal{R}_0 on Σ . Each contribution has the simple form (3.6) appropriate to the family of rays issuing from \mathcal{R}_0 . To see this it is necessary to understand the determinant in (3.9). Let \hat{k}_0 (cf (2.2)) be the unit direction vector at \mathcal{R}_0 of the ray from \mathcal{R}_0 to \mathcal{R} . From (2.3), (2.4) and (2.6),

$$\nabla_{\mathcal{R}_0} \phi = -n(\mathcal{R}_0) \hat{k}_0. \quad (3.12)$$

Then the derivatives $\partial \phi / \partial s_i$ in (3.9) give the projections of \hat{k}_0 onto Σ . If θ is the obliquity angle between \hat{k}_0 and the normal to Σ at \mathcal{R}_0 , the determinant in (3.9) is the Jacobian

$$\det \frac{\partial^2 \phi}{\partial s_i \partial X_j} = n(\mathcal{R}_0) \cos \theta \frac{d\hat{k}_0}{d\mathcal{A}(\mathcal{R})}, \quad (3.13)$$

showing that the amplitude is indeed of the form (3.5). Moreover, (3.13) is finite and positive under the conditions discussed in § 2.1, namely that the rays from \mathcal{R}_0 have no caustics en route to \mathcal{R} . Therefore (3.9) gives a smooth representation of ψ , even when the family of rays from Σ has

caustics, near which the stationary-phase approximation (3.10) breaks down.

In a uniform medium the rays are straight, ϕ is simply the distance from \mathcal{R}_0 to \mathcal{R} , and the Jacobian $d\hat{k}_0/d\mathcal{A}(\mathcal{R})$ is the inverse square of this distance. Therefore (3.9) becomes the familiar Kirchhoff integral (chapter 8 of BORN and WOLF [1975]), apart from a slight modification to the obliquity factor, which is of no consequence here since as $\kappa \rightarrow \infty$ all contributions to ψ originate near rays on Σ , for which $\theta = 0$.

We chose to describe (3.9) in detail because it is the simplest and most useful short-wave asymptotic integral representation. However, two others are frequently discussed. The first is the path integral of FEYNMAN [1948] (for simplified accounts see FEYNMAN and HIBBS [1965] and BERRY and MOUNT [1972]). This is based on paths from Σ to \mathcal{R} with infinitely many segments as discussed in § 2.1 (cf Fig. 2.3). The optical path $\phi(s; \mathcal{R})$ has infinitely many variables, so that (3.8) is an infinite-dimensional "functional integral", presenting great difficulties of precise definition and evaluation, especially near caustics (SCHULMAN [1975], LEVIT and SMILANSKY [1977], ALBEVERIO and HÖEGH-KROHN [1977], DANGELMAYR and VERT [1979]). The second alternative to (3.9) is due to MASLOV [1965] (see also KRAVTSOV [1968]), and is specifically designed to apply close to caustics. The wave (3.8) is represented as a Fourier transform where the integration variables s are components of wave vectors k perpendicular to the caustic being studied. Maslov's method is very ingenious, but suffers from the disadvantage of often failing when there are no caustics at all, as in the important case of a uniform medium where Σ is a plane and ψ is simply a plane wave. To our knowledge neither Feynman's nor Maslov's method has ever been employed in wave theory to derive a result not previously obtained otherwise.

3.3. THE HIERARCHY OF CATASTROPHE DIFFRACTION INTEGRALS

Having obtained the integral representations (3.8) for ψ , the remaining problem is to evaluate them for large κ in the important regions, near caustics, where the simple method of stationary phase (leading to (3.7)) breaks down. The reason for the failure lies in approximating the generating function ϕ near each stationary point s^* by a constant plus quadratic terms, followed by exact evaluation of the resulting complex Gaussian integrals – in other words, the ray contributions to ψ are considered separately. But when stationary points approach closer than a distance of

order κ^{-1} the contributions are not separated by a region where destructive interference occurs, and when the stationary points coalesce the action extremum is degenerate - i.e. ϕ is stationary to higher than first order, and the quadratic terms in $s - s^*$ vanish, as well as the linear terms.

The diagnosis suggests the remedy: replace ϕ in (3.8) by a simpler function with the same stationary-point structure, and evaluate the resulting integral (which is more complicated than a Gaussian) exactly. If the caustic is structurally stable it must be equivalent to one of the catastrophes, in the sense explained in § 2.2, so ϕ can be replaced by one of the normal forms Φ which are listed in Table 1 for codimension ≤ 4 . Catastrophe theory guarantees that in doing this we are not simply analyzing one of infinitely many possible diffraction integrals, but a generic integral which will occur in many different contexts; this is what makes the effort of computation (and experiment) worthwhile.

Near the caustic, all stationary points lie close together, and the factor $b(s; \mathcal{R})$ in (3.9) can be evaluated at any one of them and taken outside the integral sign. The factor κ in the exponent can be removed by a scaling to be discussed in § 3.4. The result of this procedure is a series of standard diffraction integrals Ψ , one for each catastrophe, of the form

$$\Psi(C) = \frac{1}{(2\pi)^{N/2}} \int \cdots \int d^N s \exp [i\Phi(s; C)]. \quad (3.14)$$

Following TRINKAUS and DREPPER [1977], we call these integrals *diffraction catastrophes*.

As in § 2.2 we have denoted the catastrophe polynomials by $\Phi(s; C)$, where s represents the state variables and C the control parameters (which include the coordinates of the field point \mathcal{R}). The number of state variables, N , is the corank of the catastrophe. It is the minimum number of state variables necessary for Φ to reproduce the stationary-point structure of ϕ . Extra state variables (such as the infinitely many occurring in Feynman's path integrals) can be transformed by the splitting lemma of catastrophe theory (POSTON and STEWART [1978]) so as to appear as isolated quadratic terms in Φ , which can trivially be integrated out of (3.14). Recall from § 2.1 that in optics N is at most 2.

Comparison of (2.9) and (3.14) shows that the crucial difference between the ray and wave versions of catastrophe optics lies in the way the state variables are eliminated. In ray theory, the s are removed by differentiation via the gradient map (2.9). In wave theory, the s are removed by integration via the diffraction functions (3.14).

The diffraction catastrophes $\Psi(C)$ give *transitional approximations*,

valid as $\kappa \rightarrow \infty$ and close to the caustic. Far from the caustic, the deformation of the generating function ϕ away from the standard polynomial Φ , and the variation of the factor b in (3.8), results in a loss of accuracy; in particular, the stationary-phase approximations of (3.8) and (3.14) do not coincide. To overcome this deficiency, *uniform* approximations, valid near to or far from the caustic, can be obtained by deforming the standard integrals. Rules for carrying out these deformations are given by BERRY [1976]; more mathematically elaborate treatments (including correction terms in descending powers of κ) are given by DUIS-TERMAAT [1974] and GUILLEMIN and STERNBERG [1977]. Uniform approximations to integrals were first derived by CHESTER, FRIEDMAN and URSELL [1957], and their results applied to wave theory by KRAVTSOV [1964a,b] and BERRY [1966]. In addition, LUDWIG [1966] gave explicit formulae for the whole hierarchy of what we now call cuspid diffraction catastrophes. It should be pointed out that the technique of uniform approximation is more general than catastrophe optics, since it can be applied to structurally unstable caustics, such as the glory (NUSSENZVEIG [1969a,b], BERRY [1969a,b, 1976]), whose codimension, as defined in § 2.2, is infinite (see also the end of § 3.4).

The diffraction catastrophes (3.14) constitute standard functions that stably (i.e. up to deformation) represent wave patterns near caustics. Apart from that corresponding to the fold, these functions $\Psi(C)$ cannot be expressed in closed form by means of the special functions of analysis (as tabulated in, say, ABRAMOWITZ and STEGUN [1964]). The first few members of the hierarchy have been studied in detail, and we now discuss these.

(i) *The fold diffraction catastrophe*

From eq. (3.14) and Table 1, this has the formula

$$\Psi(C) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} ds \exp [i(s^3/3 + Cs)] = \sqrt{(2\pi)} \text{Ai}(C) \quad (3.15)$$

where Ai denotes the Airy function (ABRAMOWITZ and STEGUN [1964]), a real function of one variable C . A graph of Ai^2 is shown in Fig. 3.2a. For $C < 0$ there are two rays (stationary points of the integrand) whose interference gives rise to oscillations in $\Psi(C)$. For $C > 0$ there is one contributing complex ray (that is, a complex solution of the mapping (2.9)), and $\Psi(C)$ falls exponentially to zero without oscillation. Fig. 3.2b

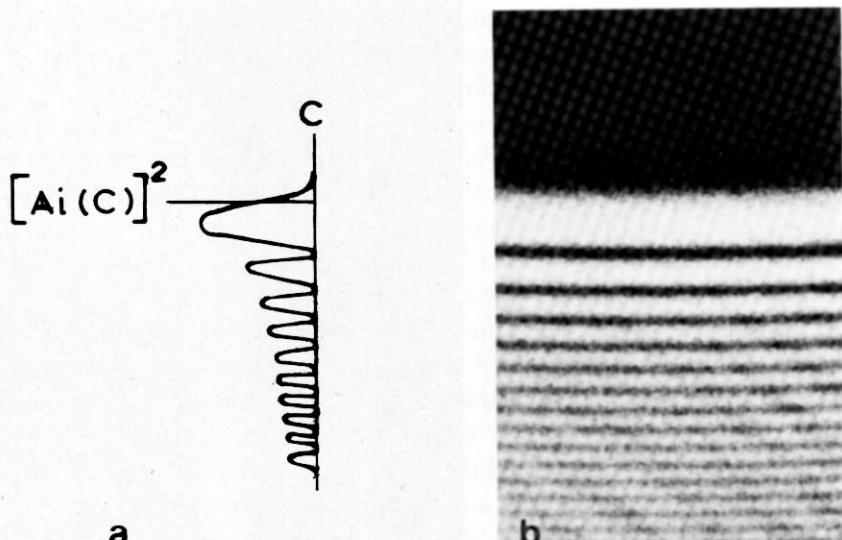


Fig. 3.2. (a) Graph of the fold diffraction catastrophe intensity (square of Airy function); (b) observation of diffraction intensity at a fold caustic.

shows observed diffraction near a fold caustic. AIRY [1838] introduced his function to study diffraction along the asymptote of a caustic, and in particular to establish the law of distribution of supernumerary rainbows (see NUSSENZVEIG [1977] for a clear modern account). LARMOR [1891] made a more detailed study of diffraction at a caustic surface, and obtained power laws for the variation of fringe spacing with the curvature of the caustic and with wavelength (the fringe index to be discussed in § 3.4); he was aware of, but did not study, diffraction at a cusp – see (ii) below.

The Airy function (3.15) is of course very different from the function of the same name describing diffraction in the focal plane of a circular patch of spherical wavefront (BORN and WOLF [1975] p. 396).

(ii) *The cusp diffraction catastrophe*

This has the formula

$$\Psi(C_1, C_2) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} ds \exp [i(s^4/4 + C_2 s^2/2 + C_1 s)], \quad (3.16)$$

and is a complex function of two variables. A map of $|\Psi|$ in the C_1, C_2

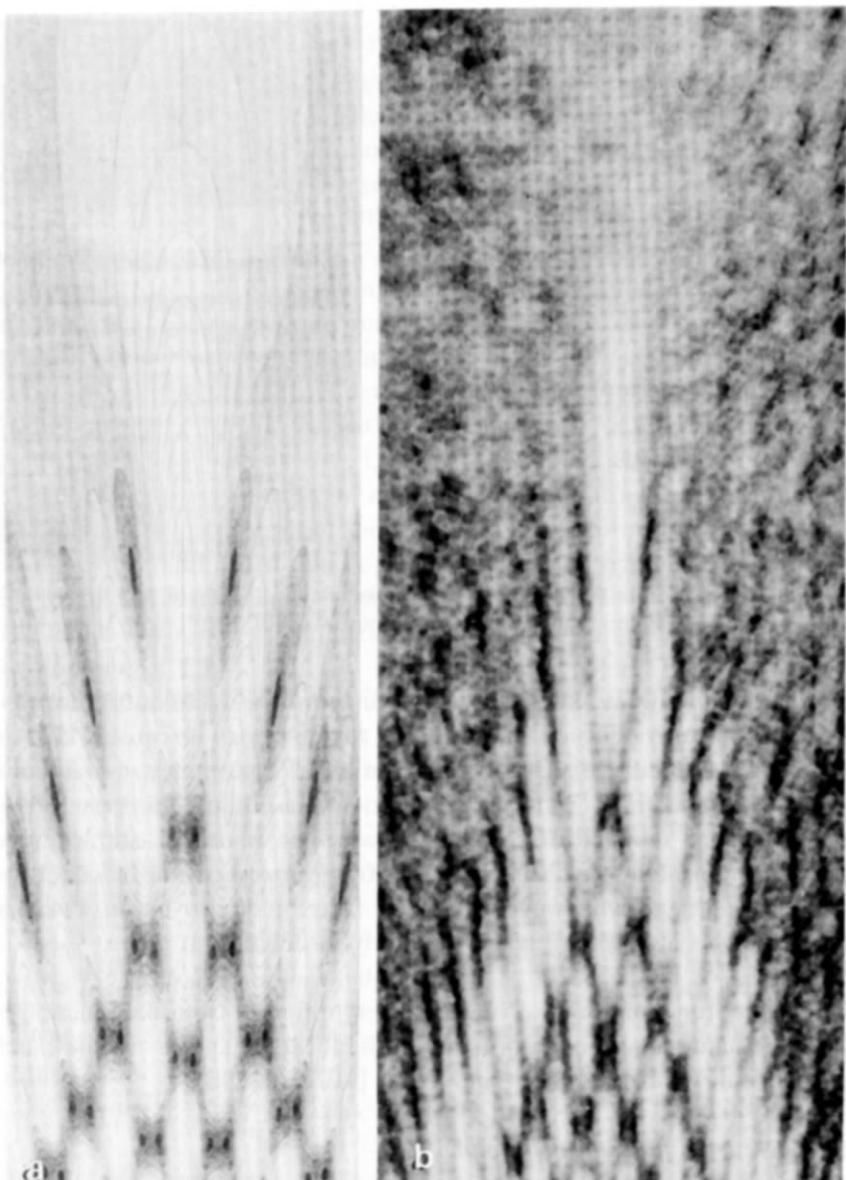


Fig. 3.3. (a) Map of the modulus of the cusp diffraction catastrophe integral (Pearcey function) (courtesy of F. J. Wright); (b) observation of diffraction intensity at a cusped caustic (courtesy of J. F. Nye).

plane is shown in Fig. 3.3a, and should be compared with observed diffraction near the cusp of a caustic, shown in Fig. 3.3b. $\Psi(C_1, C_2)$ was first studied by PEARCEY [1946], who produced contour plots of $|\Psi|$ and the phase of Ψ in what was then a tour de force of computing. More recently, diffraction at a cusped caustic has been considered in radiowave propagation by BUDDEN [1976] and MASLIN [1976a,b], in the context of quantum scattering by BERRY [1975], CONNOR [1976] and TRINKAUS and DREPPER [1977], and in underwater acoustics by HOLFORD [1972].

The most delicate features of this rich wavefield are the *wavefront dislocations* (see Appendix 1 and NYE and BERRY [1974]), which are the points where $|\Psi|$ vanishes and the phase of Ψ has a singularity. They are formed by destructive interference between the contributing rays, and their positions are given very accurately by the method of stationary phase applied to (3.16) (BERRY, NYE and WRIGHT [1979] appendix C). Within the cusp there are three real rays, whose interference produces pairs of dislocation points arranged on a distorted hexagonal lattice. Outside the cusp there is one real ray, and, in certain regions, one complex ray decaying exponentially away from the caustic (WRIGHT [1980]); close to the caustic the complex ray is sufficiently intense for its interference with the real ray to produce a single row of dislocations. The production of dislocations in this way (which also occurs for diffraction catastrophes of higher codimension) is the only circumstance known to us where complex rays affect the topology of a wavefield. Dislocations persist under perturbation, showing that the structural stability of diffraction catastrophes is a very strong property, extending to the finest scales.

Pearcey's function (3.16) should be distinguished from two superficially similar patterns, namely diffraction near the rotationally symmetric cusped cone produced by primary spherical aberration (see Appendix 2), which is structurally unstable in the sense explained in § 2.2 and so not a diffraction catastrophe, and diffraction associated with coma (see (iii) below, and Appendix 2).

(iii) *The elliptic and hyperbolic umbilic diffraction catastrophes*

These are defined by the double integrals

$$\begin{aligned} \Psi(C_1, C_2, C_3) = & \frac{1}{2\pi} \int_{-\infty}^{\infty} ds_1 \int_{-\infty}^{\infty} ds_2 \\ & \times \exp [i(s_1^3 - 3s_1s_2^2 - C_3(s_1^2 + s_2^2) - C_2s_2 - C_1s_1)] \quad (3.17) \end{aligned}$$

for the elliptic umbilic and

$$\Psi(C_1, C_2, C_3) = \frac{1}{2\pi} \int_{-\infty}^{\infty} ds_1 \int_{-\infty}^{\infty} ds_2 \exp [i(s_1^3 + s_2^3 - C_3 s_1 s_2 - C_2 s_2 - C_1 s_1)] \quad (3.18)$$

for the hyperbolic umbilic. They are complex functions of three variables,

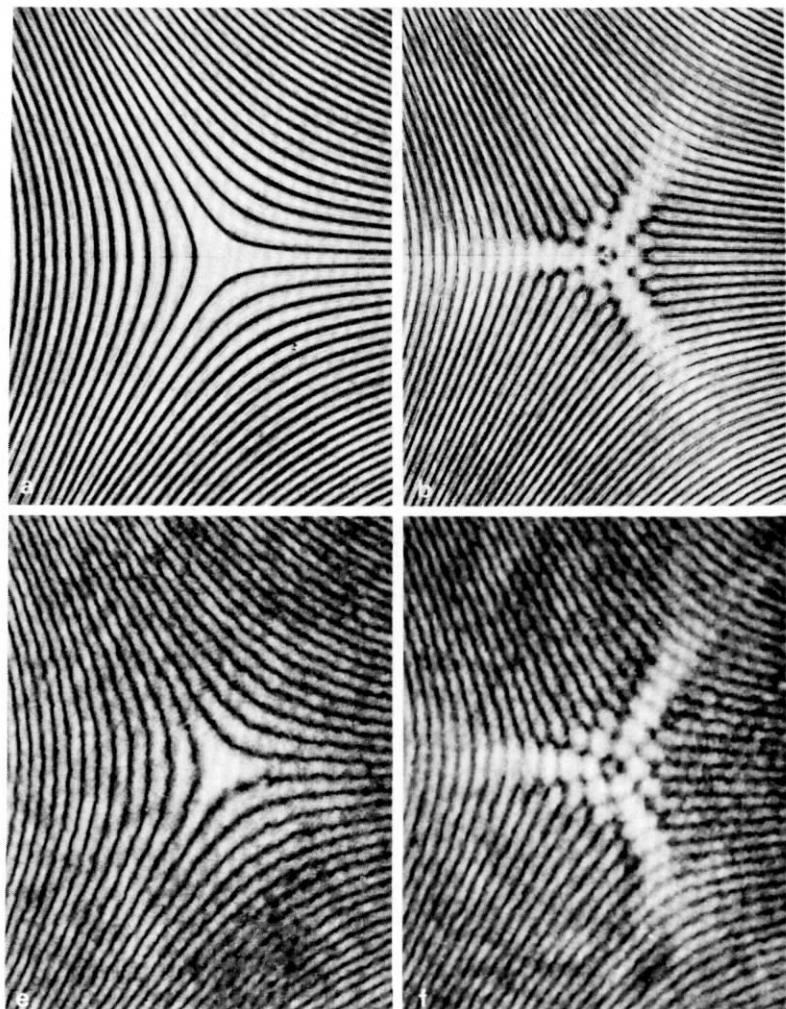


Fig. 3.4. Elliptic umbilic diffraction catastrophe. (a-d): Maps of $|\Psi|$ in the C_1, C_2 plane, for increasing C_3 (courtesy of F. J. Wright); (e-h): corresponding observed diffraction intensity (courtesy of J. F. Nye). (a) and (e) show the focal (singular) section $C_3 = 0$.

realisable as diffraction patterns in space. Both have a mirror plane $C_3 = 0$. A brief analysis of the singular ($C_3 = 0$) sections was given by TRINKAUS and DREPPER [1977], and CONNOR [1976] commented on the integrals in the context of molecular scattering. A complete theoretical and experimental study of the elliptic umbilic diffraction catastrophe has been made by BERRY, NYE and WRIGHT [1979] and a similar one is in preparation for the hyperbolic umbilic.

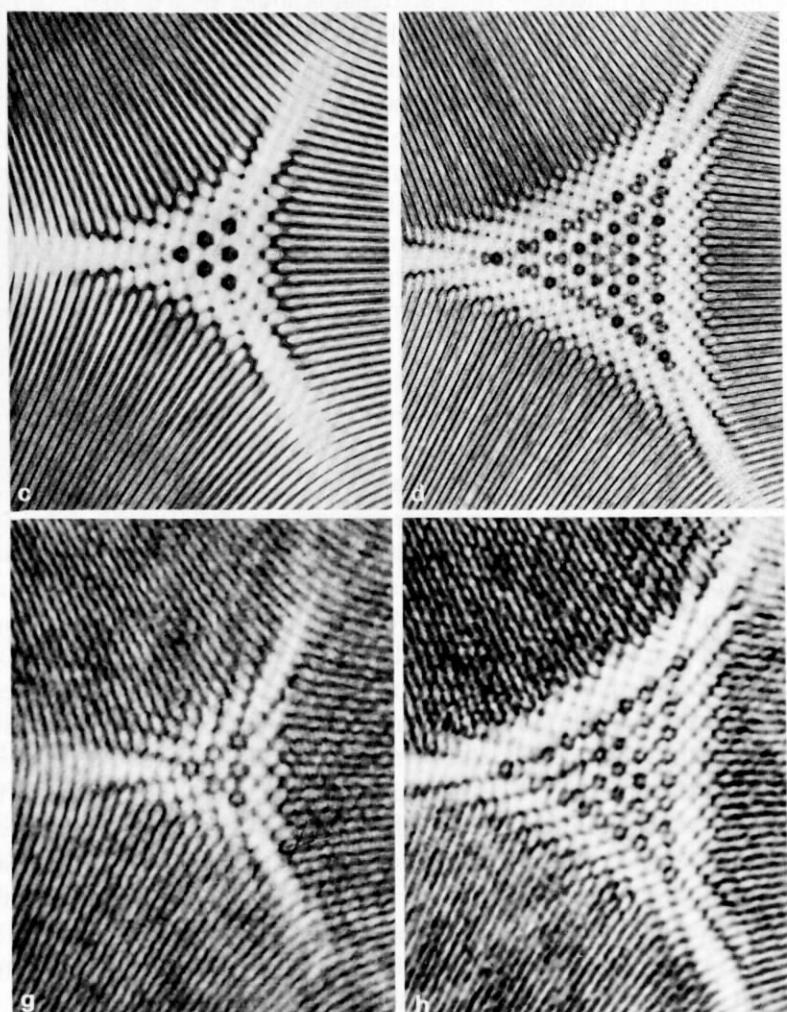


Fig. 3.4. (contd.)

For the elliptic umbilic (eq. (3.17)), maps of $|\Psi|$ in the C_1, C_2 plane, for four values of C_3 , are shown in Figs. 3.4 (a-d). These should be compared with the observed diffraction in Figs. 3.4 (e-h), obtained by focusing a microscope on various planes in the space above a water droplet lens with a triangular perimeter, through which a parallel beam of light has been refracted (see § 5). Note the remarkable agreement between theory and experiment, down to the finest details.

In the plane of the focus of the elliptic umbilic, $C_3 = 0$ (Figs. 3.4a and e), the diffraction pattern has a bright central region, roughly triangular, surrounded by fringes receding to asymptotes $2\pi/3$ apart. Surprisingly, these do not correspond to the ribs, which lie between the asymptotes. As C_3 is increased, the diffraction pattern unfolds in the space bounded by the caustic surface. Formed by the interference of four waves, the pattern consists of intensity maxima and minima resembling hexagonal prisms, stacked in space like the atoms of a distorted crystal with space group $R\bar{3}m$ (Figs. 3.4b, c, f and g). The fringes parallel to the arcs of the caustic arise from the fold diffraction catastrophe: the superposition of the three sets of fringes is responsible for the array of hexagons inside the caustic. Outside the caustic, the interference of two rays produces the fringes which dominated the focal section, but they are faint compared with the fringes inside the caustic, and as C_3 is further increased the cusps dominate the wavefield (Figs. 3.4d and h).

The dislocation lines can be considered as a skeleton on which the whole pattern is built. Within the caustic, not too near the ribs, they form hexagonally puckered rings. As C_3 is increased and the ribs approached, the rings tilt towards one another and eventually join to form "hairpins", whose arms are tightly wound helices which pierce the planes $C_3 = \text{constant}$ in the pairs of dislocation points associated with the cusp diffraction pattern. In the focal plane $C_3 = 0$, Ψ is real and there is an array of dislocation lines formed by the interference between the two real rays; these are the black lines in Fig. 3.4a: the associated regions of low amplitude are greatly extended in the C_3 direction. Outside the caustic, far from the focus, a single complex ray interferes with the two real rays. This results in an infinite sequence of helices lying close to the caustic surfaces, roughly parallel to the ribs, which do not cross the focal plane but join in opposite handed pairs with their equivalents originating from the next rib, like a pair of curly antelope horns.

For the hyperbolic umbilic (eq. (3.18)) we show maps of $|\Psi|$ and comparable observations, for the singular section $C_3 = 0$, showing the

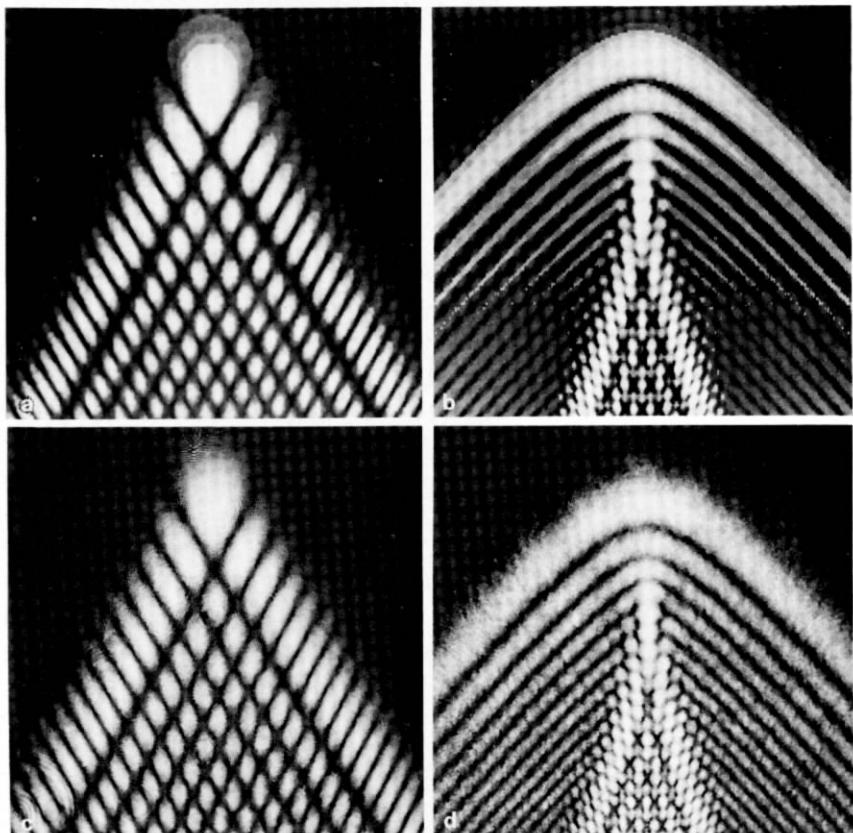


Fig. 3.5. Hyperbolic umbilic diffraction catastrophe. (a) and (b): maps of $|\Psi|$ in the C_1, C_2 plane for (a) the focal section $C_3=0$ and (b) a typical section near the focus (courtesy of F. J. Wright); (c) and (d): corresponding observed diffraction intensity (courtesy of J. F. Nye).

characteristic “parallelograms”, and a typical section $C_3>0$ (Figs. 3.5 a-d). The singular section is familiar in classical aberration theory as coma (BORN and WOLF [1975] pp. 475–479; see also Appendix 2). Note that this is non-generic in that a perturbation will yield a non-singular section: structural stability is a property of the full three-dimensional pattern.

(iv) *The swallowtail diffraction catastrophe*

This is defined by

$$\Psi(C_1, C_2, C_3) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} ds \exp [i(s^5/5 + C_3 s^3/3 + C_2 s^2 + C_1 s)]. \quad (3.19)$$

Many details are known (WRIGHT [1977]), and a complete study is in preparation. Maps of $|\Psi|$ for the singular section $C_3=0$ and typical sections $C_3<0$ and $C_3>0$ are shown in Figs. 3.6 (a-c), together with observations for comparison (Figs. 3.6 (d-f)). Note that, unlike the other codimension 3 diffraction catastrophes, $C_3=0$ is not a mirror plane.

(v) *Diffraction catastrophes whose codimension exceeds three*

It is clear that each diffraction catastrophe will have its characteristic architecture, whose complexity will increase very rapidly with codimension K . It seems unlikely that full analyses of diffraction catastrophes with $K>3$ will be worth undertaking; the parabolic umbilic ($K=4$), for example, is a *family* of three-dimensional diffraction patterns. To identify catastrophes observed experimentally, however, it is valuable to characterise at least the possible two-dimensional singular sections for diffraction catastrophes with $K>3$. Studies of caustics in light refracted by water droplet "lenses" (§ 5) have yielded the characteristic "anvil" and hexagons of the parabolic umbilic (Fig. 3.7), and the straight fringes of E_6 (Fig. 3.8) for which $K=5$ (NYE [1979]).

3.4. DIFFRACTION SCALING LAWS

As $\kappa\rightarrow\infty$, the intensity at a caustic diverges and the scale of the associated diffraction fringes shrinks to zero. It is our purpose in this section to describe the laws governing these phenomena. The results are a series of exponents for each catastrophe. These relate the standard diffraction catastrophes Ψ (eq. (3.14)) to the physical wave functions ψ (eq. (3.8)). Now Ψ involves standard control parameters C , while ψ involves physical controls, independent of κ , that we call C' (and which could represent, for example, coordinates of position \mathcal{R} in space, the time of observation, or orientations of refracting objects). The relation between $\Psi(C)$ and $\psi(C')$ is the following scaling law, written explicitly in terms of the separate control parameters C'_j ($1 \leq j \leq K$):

$$\psi(C') = \kappa^\beta \Psi(\kappa^\sigma C'_j). \quad (3.20)$$

β is a measure of the divergence of $|\psi|$, as $\kappa\rightarrow\infty$, at the most singular control point $C=0$. It was introduced by ARNOLD [1975], who called it

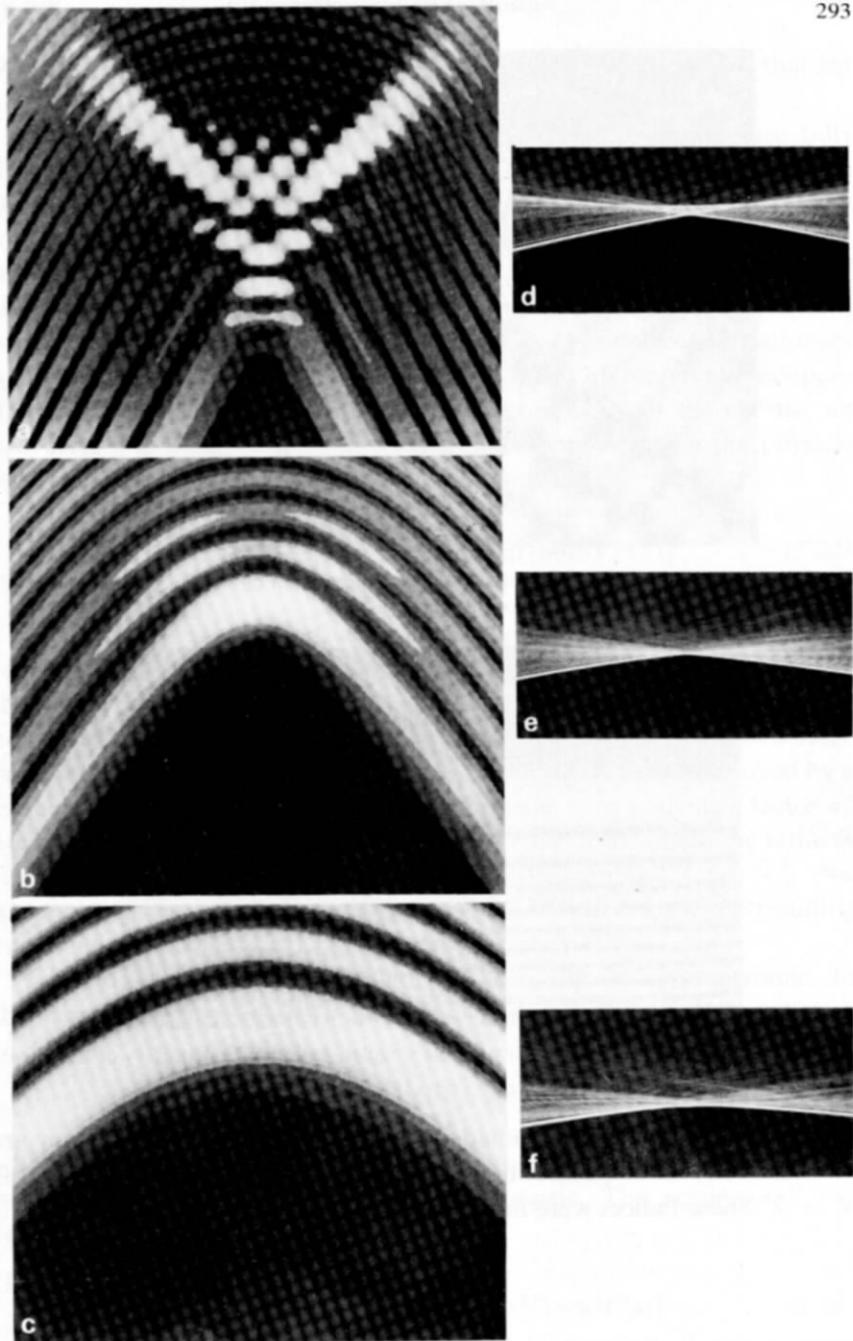


Fig. 3.6. Swallowtail diffraction catastrophe (a-c): maps of $|\Psi|$ in the C_1, C_2 plane for (a) a typical section $C_3 < 0$, (b) the singular section $C_3 = 0$ and (c) a typical section $C_3 > 0$ (courtesy of F. J. Wright); (d-f) observed diffraction intensity (courtesy of P. N. Kesterton and D. B. White).

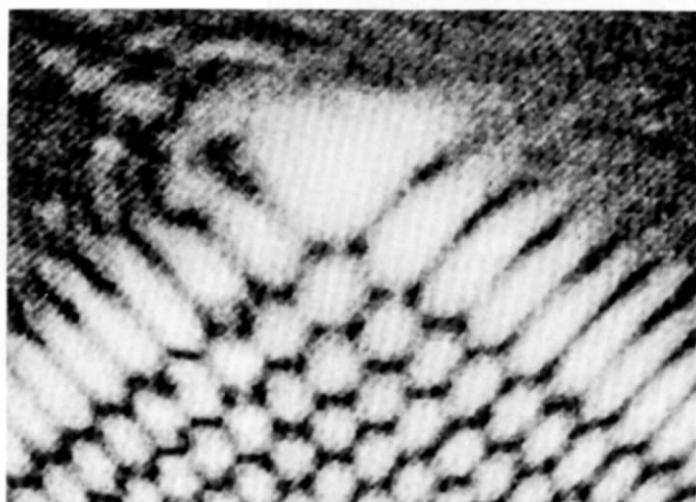


Fig. 3.7. Observed diffraction intensity for the singular section of D_5 (the parabolic umbilic) (courtesy of J. F. Nye).

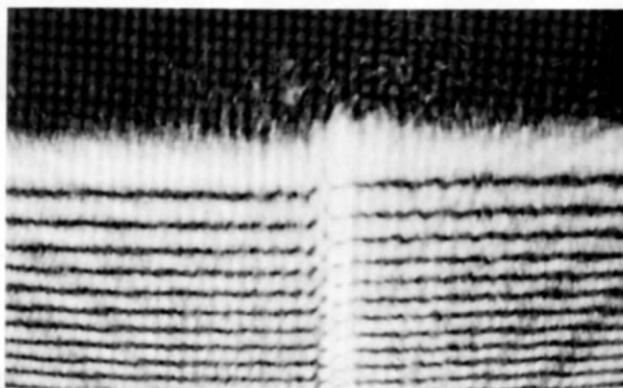


Fig. 3.8. Observed diffraction intensity for the singular section of E_6 (courtesy of J. F. Nye).

the *singularity index*, and studied in detail by VARCHENKO [1976]. σ_i are measures of the fringe spacings in the different control directions C_i , as $\kappa \rightarrow \infty$. These indices were introduced by BERRY [1977]. The sum

$$\gamma \equiv \sum_{j=1}^K \sigma_j \quad (3.21)$$

governs the K -dimensional hypervolume of the main diffraction maximum (in control space); this hypervolume scales as $\kappa^{-\gamma}$. We call γ the

fringe index; it has been studied by ROMERO [1978] who shows that for each catastrophe it is invariant under diffeomorphism.

The derivation of (3.20) and computation of β , σ_i and γ are fully described by BERRY [1977]. Here we give an outline and work through the simplest example in detail. Our starting point is the diffraction integral (3.8) for the wave $\psi(C')$. This involves κ in two ways: as a factor in the exponent, multiplying the generating function ϕ , which has dimensions of length, and as a prefactor $\kappa^{N/2}$, ensuring that far from the caustic, where the integral can be approximated by N -dimensional stationary phase, each ray contribution to the sum (3.7) has an amplitude independent of κ , as the geometrical-optics limit requires. Near the caustic, we can deform ϕ into the normal form Φ and then, denoting the physical state variables in (3.8) by s' , the integral becomes

$$\psi(C') = \left(\frac{\kappa}{2\pi}\right)^{N/2} \int \cdots \int d^N s' \exp[i\kappa\Phi(s'; C')]. \quad (3.22)$$

To carry out the scaling, and hence show that (3.20) is indeed the relation between the κ -dependent (3.22) and the κ -independent (3.14), the crucial observation is that Φ consists of a germ depending on s but not C , together with unfolding terms linear in C . Scaling is a two-stage process. Firstly, the κ -dependence of the germ (in (3.22)) is removed by a suitable linear κ -dependent transformation from s' to s , giving a factor κ^β outside the integral. Secondly, the κ -dependence of the unfolding terms is removed by suitable linear κ -dependent transformations $C'_i = \kappa^{-\sigma_i} C_i$. This procedure leads directly to the relation (3.20). Table 2 shows the resulting exponents β , σ_i and γ for the catastrophes with $K \leq 4$.

We illustrate this procedure with the fold diffraction catastrophe. In this case the integral (3.22) is

$$\psi(C') = \left(\frac{\kappa}{2\pi}\right)^{\frac{1}{2}} \int_{-\infty}^{\infty} ds' \exp[i\kappa(s'^3/3 + C's')], \quad (3.23)$$

where s' is proportional to distance along an initial wavefront and C' is proportional to distance from the fold caustic. The scaling $\kappa s'^3 = s^3$ removes κ from the germ, and gives

$$\psi(C') = \frac{\kappa^{\frac{1}{6}}}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} ds \exp[i(s^3/3 + \kappa^{\frac{1}{3}}C's)], \quad (3.24)$$

which is precisely of the form (3.20) with $\beta = \frac{1}{6}$ and $\sigma_1 = \gamma = \frac{2}{3}$. Near a fold

TABLE 2
Exponents governing the scaling of wave amplitude and fringe spacings as $\kappa \rightarrow \infty$

catastrophe	β	σ_j	γ
fold	1/6	$\sigma_1 = 2/3$	2/3
cusp	1/4	$\sigma_1 = 3/4, \sigma_2 = 1/2$	5/4
swallowtail	3/10	$\sigma_1 = 4/5, \sigma_2 = 3/5, \sigma_3 = 2/5$	9/5
elliptic umbilic	1/3	$\sigma_1 = 2/3, \sigma_2 = 2/3, \sigma_3 = 1/3$	5/3
hyperbolic umbilic	1/3	$\sigma_1 = 2/3, \sigma_2 = 2/3, \sigma_3 = 1/3$	5/3
butterfly	1/3	$\sigma_1 = 5/6, \sigma_2 = 2/3, \sigma_3 = 1/2, \sigma_4 = 1/3$	7/3
parabolic umbilic	3/8	$\sigma_1 = 5/8, \sigma_2 = 3/4, \sigma_3 = 1/2, \sigma_4 = 1/4$	17/8

caustic, therefore, the light intensity $|\psi|^2$ rises to $\mathcal{O}(\kappa^{\frac{1}{3}})$ (AIRY [1838]) and the fringe spacing is $\mathcal{O}(\kappa^{-\frac{1}{3}})$ (LARMOR [1891]).

The singularity index β increases with codimension K (and faster for umbilics than cuspoids). This is reasonable, since the larger K is, the greater is the number of rays coalescing to touch each point of the caustic. For waves in three-dimensional space, the greatest possible value of β is unity and this is attained at the (structurally unstable) perfect focus of a patch of spherical wavefront.

The fringe index γ also increases with K , showing that the intense region near the singularity, where all rays interfere constructively, shrinks very rapidly as $\kappa \rightarrow \infty$ for the higher catastrophes. This shrinking is anisotropic in control space, because for each catastrophe the exponents σ_j are not all the same. Thus for the cusp the fringe spacing along the symmetry axis is $\mathcal{O}(\kappa^{-\frac{1}{3}})$ while that across the cusp is $\mathcal{O}(\kappa^{-\frac{1}{2}})$, explaining why the diffraction patterns near cusps formed by objects large in comparison with the wavelength of light are greatly elongated in the cusp direction. Similar arguments explain the stretching of the elliptic umbilic diffraction pattern along its symmetry axis as $\kappa \rightarrow \infty$ (in the experiment reported by BERRY, NYE and WRIGHT [1979], the hexagonal prismatic maxima near the axis were elongated about 50 times as compared with those in the standard diffraction function (3.17)).

The computation of the exponents for the higher catastrophes on the list of ARNOLD [1975] involves several difficulties, discussed in detail by BERRY [1977] in connection with an application to random short waves which is to be described in § 7.

Outside catastrophe optics, wavelength scaling laws need not take the form (3.20). There are richer possibilities, which we illustrate with a brief discussion of the glory (BRYANT and JARMIIE [1974], TRICKER [1970],

NUSSENZVEIG [1969a,b], KHARE and NUSSENZVEIG [1977]). This is intense backscattering of light from a spherical water droplet. Rays emerging backwards would form a (structurally unstable) focal line extending to infinity, and would give rise to an amplitude $|\psi| \propto \kappa^{\frac{1}{2}}$. Such rays only exist, however, for spheres with refractive indices between $\sqrt{2}$ and 2, a condition which excludes water. For water, the most deflected geometrical-optics rays, which emerge at 14° to the backward direction with total amplitude $\kappa^{-\frac{1}{2}}$, creep by diffraction around the surface of the drop, suffering wavelength dependent attenuation as they do so. It is these creeping rays which form the focal line, giving rise to a final backscattered wave amplitude

$$|\psi| \propto \kappa^{\frac{1}{2}} \exp[-D(\kappa a)^{\frac{1}{2}}], \quad (3.25)$$

where a is the radius of the drop and D a constant (NUSSENZVEIG [1969a,b]). Therefore the glory is a focusing effect which disappears in the geometrical-optics limit!

§ 4. The Geometry of Wavefronts and their Normals

4.1. PARAXIAL OPTICS OF A UNIFORM MEDIUM

In the most general focusing problems, we wish to consider the propagation of rays from some specified initial surface Σ . As in § 2.1 we consider monochromatic waves in isotropic media so that the rays are normal to the geometrical-optics wavefronts. Moreover we shall now consider in detail a uniform medium, where the rays are straight. This describes a whole class of problems where an initial wavefront W emerges from a non-planar refracting boundary such as a water/air or glass/air interface; for convenience we identify W with Σ . W is specified by its deviation $f(r)$ from the plane $r = (x, y)$, $Z = 0$, and propagates in the Z -direction (Fig. 4.1).

We now make the important stipulation that $f(r)$ has small slopes, so that *paraxial optics* is applicable, and of course we require that $f(r)$ is smooth on wavelength scales (i.e. its curvatures are small) so that geometrical optics and related short-wave approximations are valid. These restrictions imply that the rays propagate at small angles $\theta = \sin \theta = \tan \theta$ to some optical axis, in this case the Z -direction. Note that there is no restriction on the distance of a ray from the Z -axis, merely the angle

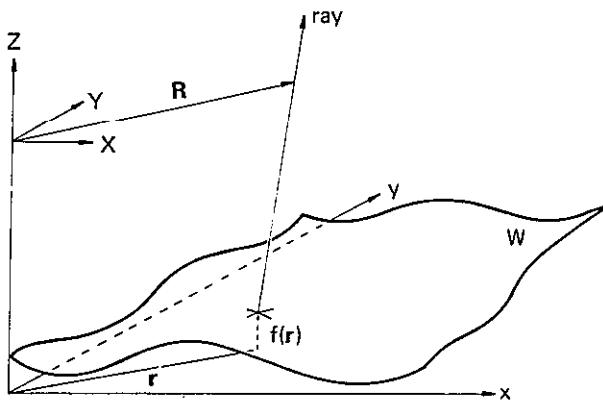


Fig. 4.1. Coordinates of wavefront and observation space.

between them – indeed in § 6 we shall consider quite large extents of wavefront using this model. Paraxial optics linearizes Snell's law, with the useful consequence that W may be regarded as an inversion of the refracting boundary, scaled by the ratio of the refractive indices, and we can infer the geometry of the one directly from that of the other. Henceforth we shall consider only paraxial catastrophe optics, which is, we emphasize, an entirely self-consistent theory. There is a complete formal analogy between paraxial ray and wave optics and the non-relativistic classical and quantum mechanics of a particle in two dimensions (MARCUSE [1972], HANNAY [1977]).

In the three-dimensional observation space \mathcal{R} we shall generically encounter catastrophes with $K \leq 3$. But in many situations we shall need to consider not just one possible W but whole families of initial wavefronts which can be explored by changing control parameters, such as rotating a water droplet lens so that gravity affects its shape (§ 5) or considering the time evolution of a rippling water surface (§ 6). In this way we can explore the unfoldings of catastrophes with $K > 3$ and even encounter their singularities. It is important to note that a singular point is interesting not just for its own sake, but more so for the geometry it imposes on its surroundings. In this way higher singularities *organize* lower ones; in later sections we shall talk of caustics in three space dimensions being organized by catastrophes of much higher codimension. Thus the singular points of the family of initial wavefronts assume special importance and we must analyze them in detail, to locate and determine the local configuration of the most important parts of the caustics; the global configuration is constrained by these local forms.

We denote the family of initial wavefronts by $f = f(\mathbf{r}; A)$, where A represents control parameters influencing the form of the wavefront, and we denote spatial observation points by $\mathcal{R} \equiv (\mathbf{R}, Z)$ where $\mathbf{R} \equiv (X, Y)$ is a two-dimensional vector. Then the optical path length from \mathbf{r} on W to (\mathbf{R}, Z) is ϕ , where

$$\phi(\mathbf{r}; \mathbf{R}, Z, A) = \sqrt{[(Z - f(\mathbf{r}; A))^2 + (\mathbf{R} - \mathbf{r})^2]}. \quad (4.1)$$

For small angles (paraxiality) we can expand the square root, and the additional assumption of small curvatures means that focusing occurs for $Z \gg f$. If we write the control parameters \mathbf{R} , Z and A together as C , then

$$\phi = \phi(\mathbf{r}; C) = Z - f(\mathbf{r}; A) + \frac{1}{2Z} (\mathbf{R} - \mathbf{r})^2. \quad (4.2)$$

This is precisely the generating function $\phi(s; C)$ of § 2.1, the state variables s being coordinates x, y in the \mathbf{r} plane.

For rays, we require ϕ to be stationary with respect to \mathbf{r} , as in (2.9), thus forming the gradient map

$$\mathbf{R} = \mathbf{r} - Z \nabla_{\mathbf{r}} f(\mathbf{r}; A). \quad (4.3)$$

The rays focus at caustics when the Hessian (2.12) vanishes, giving

$$\mathcal{H}[\phi] = (f_{xx} - 1/Z)(f_{yy} - 1/Z) - f_{xy}^2 = 0 \quad (4.4)$$

where subscripts denote differentiation. This is quadratic in Z , giving the two conjugate levels at which rays from \mathbf{r} focus. These values of Z are the two principal radii of curvature of f .

4.2. THE CURVATURE LANDSCAPE

In paraxial optics, rays from all points on W with a given principal curvature \mathfrak{C} focus in the same plane \mathbf{R} , at $Z = 1/\mathfrak{C}$. Thus it is advantageous to consider not the geometry of the height function f of W , but the geometry of the curvature function. As with a height function, it is useful to visualize this as a “landscape”. But it is a peculiar landscape, consisting of two surfaces which never intersect; they may join, but only at special points (see § 4.3), and there are no cliffs (infinite slopes). The curvature landscape is generally not obviously related to the height landscape. NYE [1978] considers a curvature difference landscape which is single valued; this has certain advantages in the particular case of water droplets he

considers (see § 5.2) but, as we shall shortly show, the information thereby lost is in general some of the most important to be had.

We can conveniently obtain the principal curvatures $\mathfrak{C}_1, \mathfrak{C}_2$ from (4.4) as

$$1/Z_{1,2} = \mathfrak{C}_{1,2} = \frac{1}{2}(f_{xx} + f_{yy}) \pm \sqrt{\left[\frac{1}{4}(f_{xx} - f_{yy})^2 + f_{xy}^2\right]}. \quad (4.5)$$

Note that the total curvature is just $\nabla^2 f$: we shall use this fact when discussing water droplet optics in § 5.

All points with the same principal curvature lie on a contour of the curvature landscape. These contours of curvature \mathfrak{C} in $r = (x, y)$ map via (4.3) to the caustics in the planes $Z = \text{constant} = 1/\mathfrak{C}$ (Fig. 4.2). A line element dR of the caustic is perpendicular to the direction i of the principal curvature which generates it. In the notation of Fig. 4.2,

$$dR = (dr \cdot j)(1 - \mathfrak{C}_i/\mathfrak{C}_j)j. \quad (4.6)$$

Clearly something interesting happens when $dr \cdot j = 0$; dr and i are parallel at such a point (the two principal directions of curvature being everywhere orthogonal) and as we pass through such a tangency dR changes sign and the caustic reverses direction and crosses its tangent, thus forming a cusp. We shall return to this in § 4.4.

The families of caustics in \mathcal{R} are completely determined by the geometry of f . We can consider the caustics to be organized by their most singular points, whose local unfoldings constrain the possible topologies of the three-dimensional caustic structures. It is very informative to

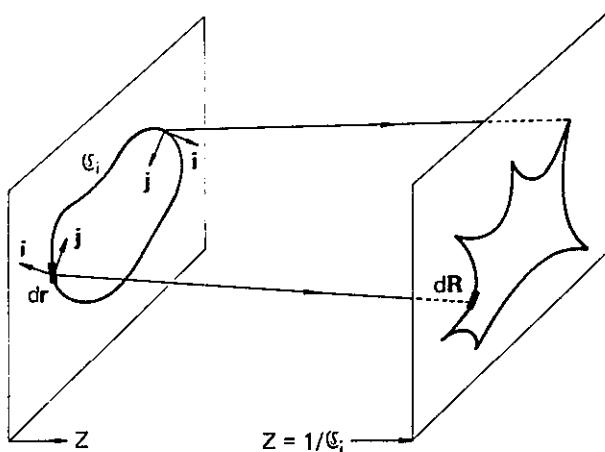


Fig. 4.2. Caustics are images of contours of principal curvature.

analyze the corresponding singular points of the curvature landscape: catastrophe theory will tell us the local behaviour of the caustics, while other differential-geometric properties will tell us about their global connectivity.

The simplest stationary points – maxima, minima and saddles – of the curvature landscape do not correspond to singular points in \mathcal{R} , but to places where a rib reaches an extreme value of Z , at which point it is tangential to a plane \mathbf{R} , and retreats again smoothly. Maxima and minima correspond to ribs convex upwards and downwards respectively: in a sequence of two-dimensional sections, these are *lips* (Fig. 2.9a). Saddles correspond to *beak-to-beak* events (Fig. 2.9b).

4.3. UMBILICS

The most important places on the curvature landscape are the locally conical points where the two surfaces meet, called (for obvious reasons) umbilic points. At an umbilic point, $\mathfrak{C}_1 = \mathfrak{C}_2$, the curvature contours are singular, and the caustic has a singular point “focus”. Note that umbilic points are the only possible connections between the two surfaces of the curvature landscape. In classical differential geometry, umbilic points were studied by DARBOUX [1896], GULLSTRAND [1905], FORSYTH [1912] and, more recently, by PORTEOUS [1971].

The condition for a point to be umbilic is, from (4.5),

$$f_{xx} - f_{yy} = f_{xy} = 0. \quad (4.7)$$

Umbilic points are classifiable in three different ways, discussed in detail by BERRY and HANNAY [1977], in terms of the coefficients of the cubic terms of a local expansion of f . These are third derivatives evaluated at the umbilic point, and we employ the notation

$$\alpha \equiv f_{xxx}, \quad \beta \equiv f_{xxy}, \quad \gamma \equiv f_{xyy}, \quad \delta \equiv f_{yyy}. \quad (4.8)$$

The unfolding of the point focus of the caustic is an elliptic (E) or hyperbolic (H) umbilic catastrophe as the contours of curvature near the umbilic point are ellipses or hyperbolae, yielding the *catastrophe* classification of an umbilic point:

$$\text{if } C(\alpha, \beta, \gamma, \delta) \equiv 4(\alpha\gamma - \beta^2)(\beta\delta - \gamma^2) - (\alpha\delta - \beta\gamma)^2 \begin{cases} > 0 & \text{then } E \\ < 0 & \text{then } H. \end{cases} \quad (4.9)$$

These two cases are smoothly connected by the parabolic umbilic, for which the contours are parabolae; this case is non-generic in three dimensions because the parabolic umbilic has $K=4$. The umbilic is parabolic if $C(\alpha, \beta, \gamma, \delta)=0$, unless the other classification functions (see below) vanish, in which case it is an umbilic with $K>4$. The transition $E \leftrightarrow H$ corresponds to a smooth change in the angle of the axis of the double cone connection between the two curvature surfaces, with respect to the propagation direction (Fig. 4.3).

The principal axes of curvature generate an orthogonal net of lines of curvature (not to be confused with contours of curvature) whose configuration around a generic umbilic point is one of the three shown in Fig. 4.4. Lemon (L) is distinguished from star (S) and monstar (M) by there being one straight line through the singularity rather than three. Corresponding to this is the *pattern* classification:

$$\text{if } P(\alpha, \beta, \gamma, \delta) \equiv 4[3\gamma(\alpha - 2\gamma) - (\delta - 2\beta)^2][3\beta(\delta - 2\beta) - (\alpha - 2\gamma)^2] \\ - [(\delta - 2\beta)(\alpha - 2\gamma) - 9\beta\gamma]^2 \begin{cases} > 0 & \text{then } M \text{ or } S \\ < 0 & \text{then } L. \end{cases} \quad (4.10)$$

During a circuit of the umbilic the principal axes of curvature rotate one half-turn; co-rotation (counter-rotation) is denoted by $+\frac{1}{2}$ ($-\frac{1}{2}$). This is the basis of the *index* classification:

$$\text{if } J(\alpha, \beta, \gamma, \delta) \equiv \alpha\gamma - \gamma^2 + \beta\delta - \beta^2 \begin{cases} > 0 & \text{then } +\frac{1}{2} (L \text{ or } M) \\ \leq 0 & \text{then } -\frac{1}{2} (S). \end{cases} \quad (4.11)$$

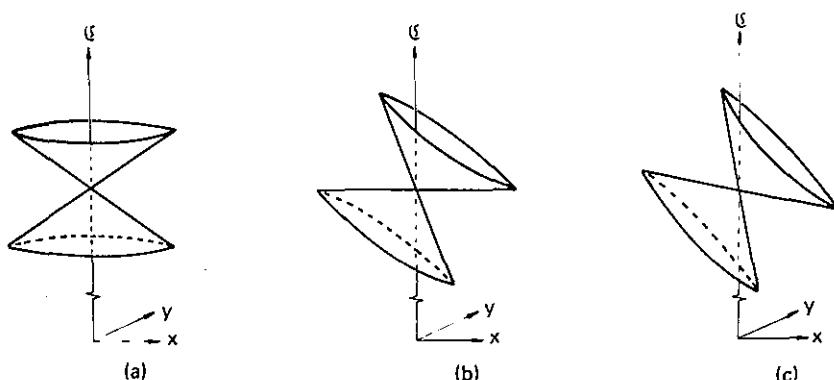


Fig. 4.3. Orientation of the conical (umbilical) connection between the two surfaces of the curvature "landscape": curvatures in the neighbourhood of (a) elliptic, (b) parabolic, (c) hyperbolic umbilic points.

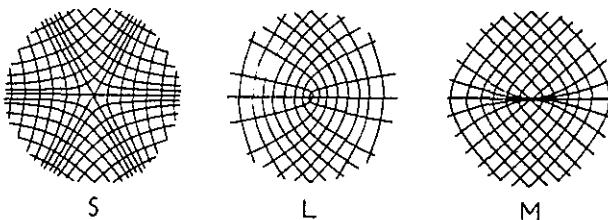


Fig. 4.4. Lines of curvature around umbilic points with star (*S*), lemon (*L*) and monstar (*M*) configurations.

The threefold classification of umbilics is summarized in Fig. 4.5; this diagram illustrates the following important implications of the inequalities (4.9) and (4.11): all elliptic umbilics have index $-\frac{1}{2}$, and all umbilics with index $+\frac{1}{2}$ are hyperbolic. The classification of umbilic catastrophes with $K > 4$ depends on terms higher than cubic; we do not discuss these cases further.

Obviously, the catastrophe classification dictates the local form of the caustics. The pattern classification serves to distinguish the rare monstars from the asymptotically indistinguishable lemons (Figs. 4.4 and 4.5), but we know of no circumstance in which this distinction affects the caustics.

The index classification constrains the configuration of caustics remote from the focus. It is especially useful in conjunction with the following conservation laws:

(i) The number of rotations of the principal axes of curvature of f around any closed curve on W is the algebraic sum of the indices of the enclosed umbilic points. If A is varied, so that the number and positions of the umbilics vary, then provided no umbilic crosses the curve the algebraic sum of the indices is invariant.

(ii) The net index of umbilics on any unbounded surface is equal to the Euler-Poincaré characteristic of the surface (SPIVAK [1975]). We shall make use of this in §§ 6 and 7.

As A is varied, the umbilics move along definite trajectories in the parameter space. Analyzing these, and studying the possible umbilic

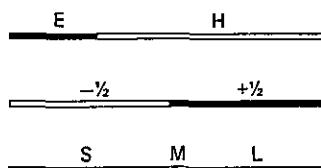


Fig. 4.5. The interrelation of the three classifications of umbilic points.

"reactions" (NYE [1979]) are valuable tools for investigating catastrophes of codimension $K > 3$. Rule (i) above shows at once that two elliptic umbilics cannot be born or annihilate, as they must both have index $-\frac{1}{2}$. The only umbilic reactions with $K \leq 4$ are:

Birth/annihilation: $H_{-\frac{1}{2}} + H_{+\frac{1}{2}} \leftrightarrow \text{nothing}$

Transformation: $E_{-\frac{1}{2}} \leftrightarrow \text{parabolic umbilic} \leftrightarrow H_{-\frac{1}{2}}$. (4.12)

4.4. CUSPOIDS

On a plane screen $Z = \text{constant}$, the only stable caustics are fold lines which may be interrupted by cusp points. We recall that these are the images of contours $\mathfrak{C} = 1/Z$, and that a line element of caustic (Fig. 4.2) is perpendicular to that direction of principal curvature which generates it (eq. (4.6)). The points where lines of curvature touch a contour of curvature generate cusps in the focal plane. If we consider all the contours, the loci of all such points are *riblines* in \mathbf{r} ; their images in (\mathbf{R}, Z) are *ribs* of the caustic.

There may be points where the contours and lines of curvature not only touch but osculate (i.e. have the same curvature): these are swallowtail points. If we consider the family of wavefronts parametrized by A , there may also stably occur points where the contours and lines of curvature touch, osculate and also have the same rate of change of curvature: these are butterfly points. The hierarchy is shown in Fig. 4.6; we could continue similarly for cuspoids with $K > 4$. Of course, a cusp point must lie on a ribline; at a swallowtail point the ribline touches both the contour and

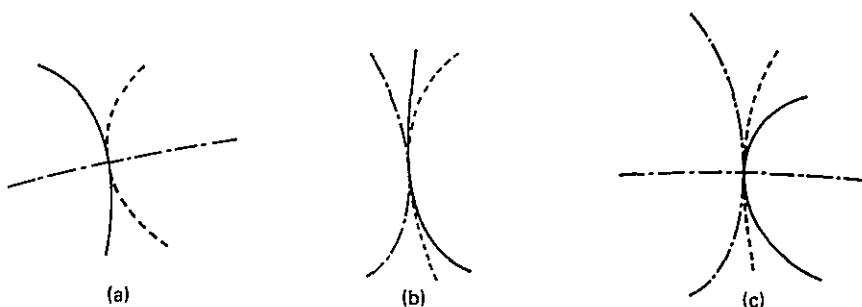


Fig. 4.6. Geometry of a wavefront in the neighbourhood of a cuspid singularity; (a) cusp; (b) swallowtail; (c) butterfly. Contours of \mathfrak{C} (full lines), lines of curvature (dashed lines), riblines (chain lines).

line of curvature, whilst at a butterfly point there are two riblines, one of which touches both the contour and line of curvature.

As with umbilics, analysis of the trajectories of the cuspoids (ribs and riblines and their analogues in higher dimensions) is extremely helpful in elucidating the geometry of higher catastrophes.

§ 5. Liquid Droplet Lenses

5.1. THE IRREGULAR DROPLET AS A LENS

The distortion of an image viewed through a raindrop is familiar to all spectacle wearers, for an irregular water droplet on an inhomogeneously dirty surface is an extremely aberrant lens. LARMOR [1891] wrote "on looking at a bright lamp, sufficiently distant to be treated as a luminous point, through a plate of glass covered by fine raindrops, the caustic surfaces after refraction through the drops are produced within the eye itself, and their sections by the plane of the retina appear as bright curves projected into the field of vision". MINNAERT [1954] also drew attention to the optics of such droplets, but neither author saw any reason to study the forms of the seemingly infinite variety of caustics produced in this way: it is in such circumstances as these that we can use catastrophe optics and realise the power of the notions of structural stability and genericity.

Although much of the following discussion is couched in terms of geometrical optics, note that in real experiments we do not work with light of zero wavelength, and diffraction is significant: we invoke the structural stability of the diffraction catastrophes (§ 3) to relate observations to the underlying geometrical caustics.

A droplet on a flat transparent substrate illuminated normally by a plane wave is well described by the paraxial optics of § 4. The form of the drop surface is constrained by the surface tension of the liquid, and the differential equation to be satisfied depends on both the size of the drop and the orientation of the substrate, the crucial condition being whether or not the internal pressure may be considered to be uniform throughout the drop. On a horizontal substrate, a water drop of more than 3 or 4 mm across is large enough for gravity to affect its shape and cause the internal pressure to be non-uniform. In § 5.2 we consider the optics of small droplets, for which the effect of gravity is negligible. For a water drop of even 1 mm or so in size on a vertical surface the effect of gravity is

significant. Tilting the substrate, and hence varying the effect of gravity on the drop, provides an additional control parameter affecting the caustics. This will be considered in § 5.3.

We discuss caustics of refraction, produced by droplets of water, as this is most convenient for experiments. The discussion carries over in principle for caustics in reflected light, and for any liquid, with the proviso that the different surface tension will of course affect the size of drop for which gravity becomes significant. We remark in passing that the caustics formed by reflection from plastic film, used to illustrate the geometry of the elementary catastrophes in § 2.3, are on a much larger scale and diffraction is much less significant. The surface tension condition for uniform internal pressure (eq. (5.2)) is obeyed approximately by the plastic film; conditions analogous to non-uniform internal pressure can be explored in a controlled way with the plastic film as part of the wall of a vessel full of liquid: the pressure gradient depends on the density of the liquid. These techniques have not proved as convenient for research as the use of water droplets.

5.2. DROPLETS WITH UNIFORM INTERNAL PRESSURE

Observations with the naked eye such as those described by LARMOR [1891] are of the directional caustics at infinity – the *far field*. BERRY [1976] considered this case in detail for small horizontal drops (infinity in this instance is anything greater than a few centimetres). The far field caustic is the image of the contour of zero curvature on the initial wavefront W . If W is specified by its deviation f from a reference plane (see § 4), this is the line where the Gaussian curvature

$$\mathfrak{C}_1 \mathfrak{C}_2 = f_{xx} f_{yy} - f_{xy}^2 \quad (5.1)$$

vanishes. Berry pointed out that as the far field is a two-dimensional space of directions, the only structurally stable caustics are fold lines interrupted by cusp points. He showed that if a surface-tension dominated droplet has a far field caustic (and it need not – for example, if it is shaped like the cap of an ellipsoid), the fold caustics must be concave outwards and there must be at least three cusps. This can be verified experimentally by shining a broadened laser beam vertically through a small irregular droplet on a horizontal flat glass plate and observing the far field caustic on the ceiling of a darkened room (Fig. 5.1). Such a

caustic, perhaps a metre across, with its delicate diffraction structure in the red light of a helium-neon laser, is a most beautiful demonstration of catastrophe optics.

The directional caustic at infinity is not the special asymptotic form it might appear: real and virtual far fields can be identified (with a 180° rotation) and the three-dimensional caustic structure is thus continuous from the real near field through the far field to the virtual near field. By using suitable lenses, it is possible in principle to cause any two-dimensional section to be at infinity, but we reserve the term far field for the caustic at infinity when there are no lenses in the space beyond W .

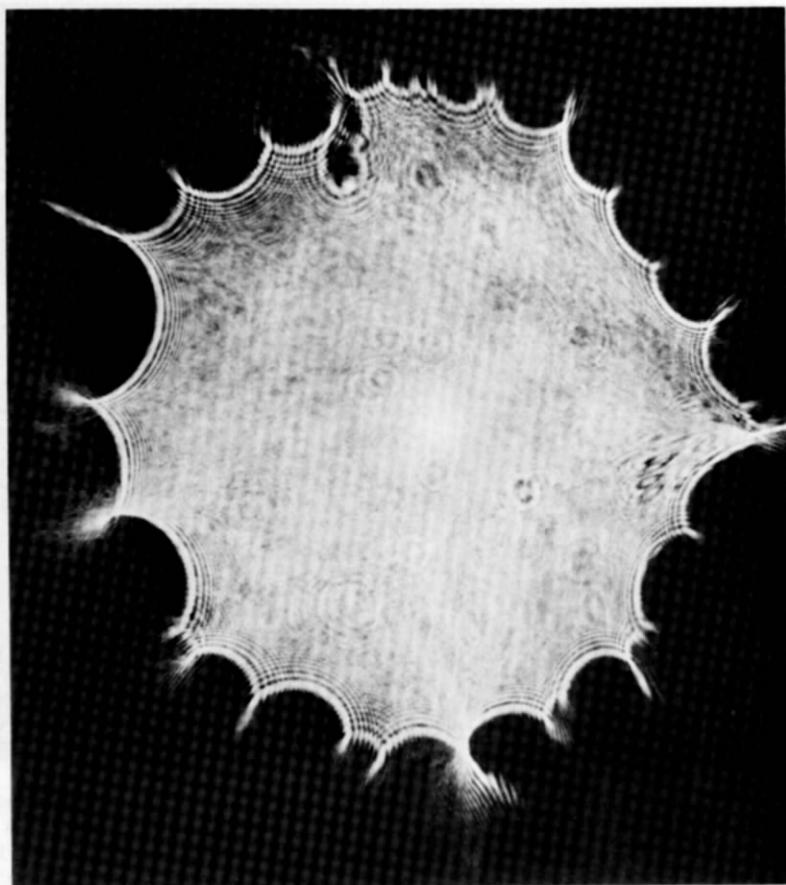


Fig. 5.1. The far field caustic from a small irregular water drop on a horizontal flat glass plate.

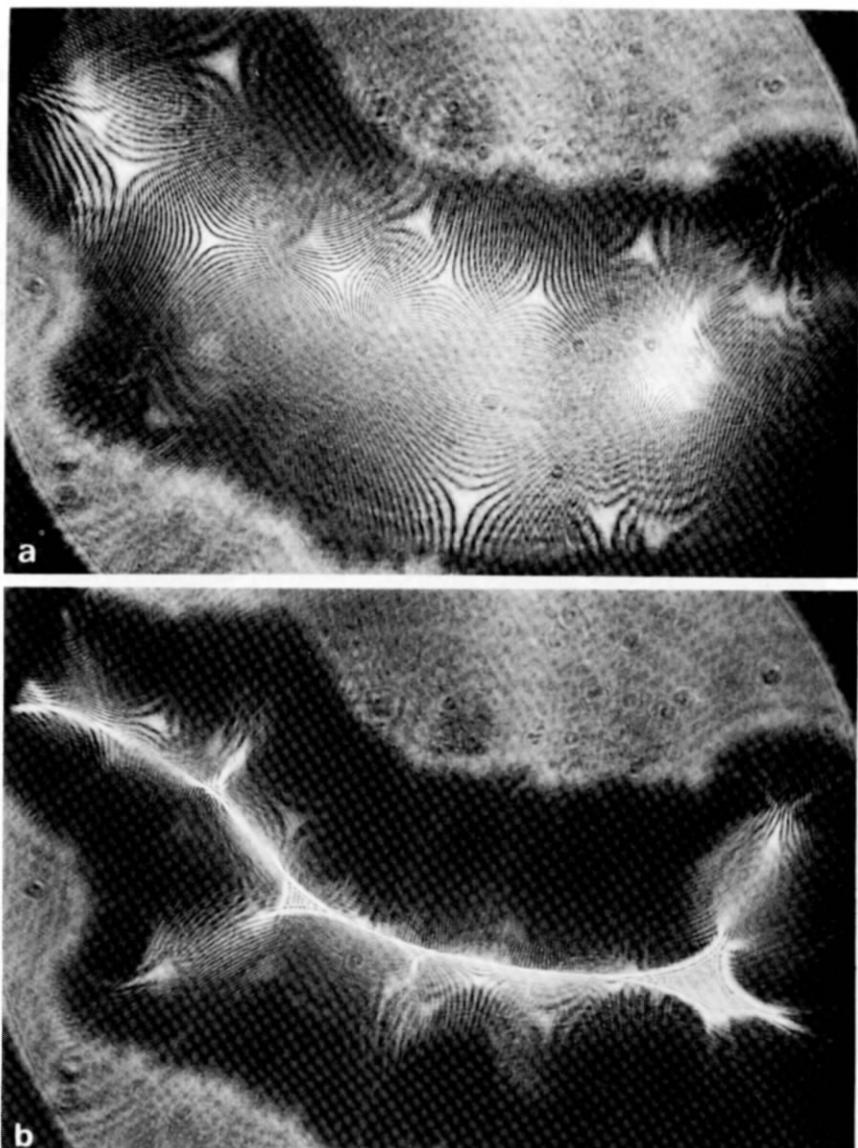
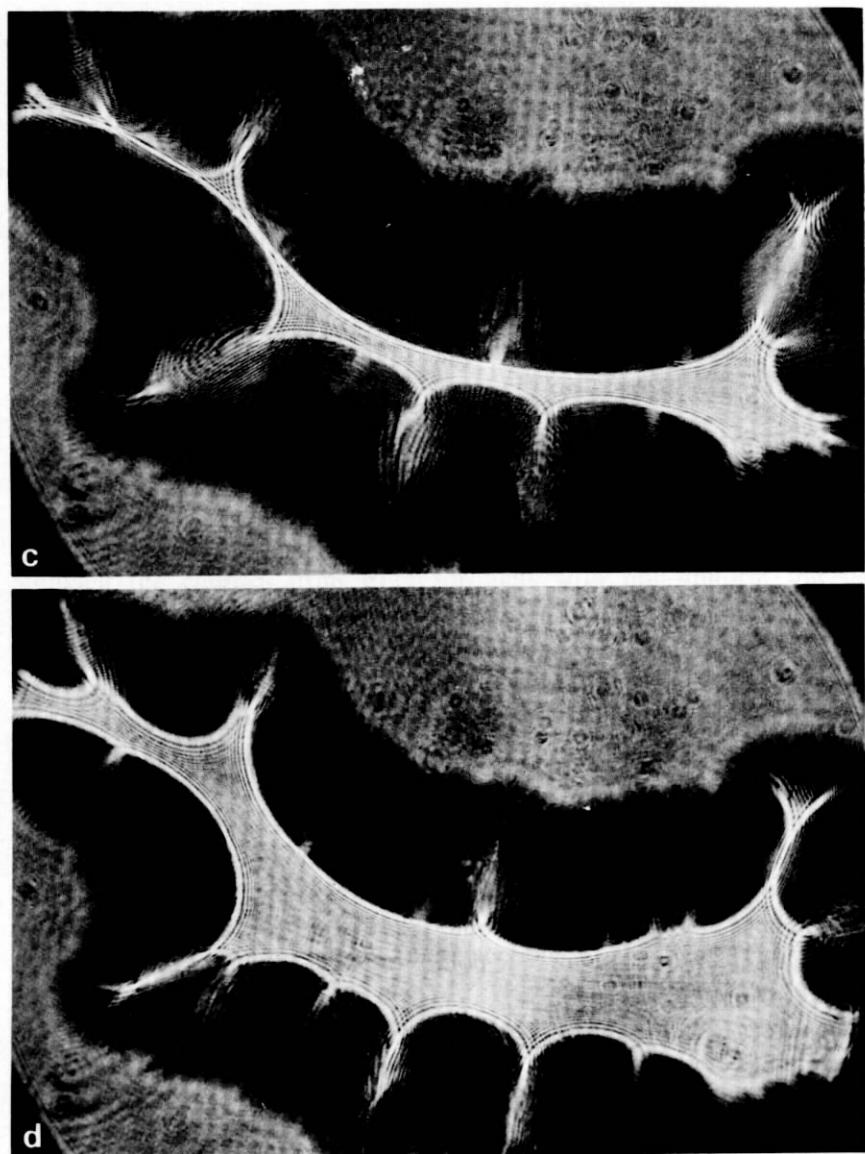


Fig. 5.2. A sequence of two-dimensional sections of the near field of a small irregular water drop on a horizontal microscope slide. In (a) the microscope is focused onto the plane of the elliptic umbilic foci; in (b) to (d) the plane of focus moves towards the drop (courtesy of J. F. Nye).

Fig. 5.2 (*contd.*)

Studying the spatial caustics beyond W is clearly more general, and we expect to find generically catastrophes of codimension $K \leq 3$. It turns out, however, that the restriction to uniform internal pressures rules out the hyperbolic umbilic, and the only possible singularities are the fold, cusp, swallowtail and elliptic umbilic (NYE [1978]).

If the excess internal pressure p is uniform throughout a drop of refractive index n , the surface tension γ constrains f to be a solution of

$$\nabla^2 f = (n - 1)p/\gamma = \mathfrak{K} \quad (5.2)$$

where \mathfrak{K} is a constant. In paraxial optics, the sum of the principal curvatures (4.5) is equal to the Laplacian, so

$$\nabla^2 f = \mathfrak{C}_1 + \mathfrak{C}_2 = \mathfrak{K}. \quad (5.3)$$

NYE [1978] defines the curvature difference as

$$\mathfrak{D} = |\mathfrak{C}_1 - \mathfrak{C}_2|. \quad (5.4)$$

As a consequence of the relationship (5.3), the contour maps of each of \mathfrak{C}_1 , \mathfrak{C}_2 , \mathfrak{D} and the Gaussian curvature $\mathfrak{C}_1\mathfrak{C}_2$ are identical, only the labels on the individual contours being different. For convenience, Nye considered the caustic structure in terms of the single valued \mathfrak{D} -landscape, and showed that it has locally conical minima all at zero level, the only other stationary points being saddles; there are no maxima.

At an umbilic point on W , $\mathfrak{D} = 0$, $\mathfrak{C}_1 = \mathfrak{C}_2$ and the rays focus at a height $Z = 2/\mathfrak{K}$. It is important to note that all the umbilic points focus at this same level, and are all elliptic. Without the restriction on drop shape, we do not normally see all the umbilic foci at once on a two-dimensional screen; of course, all the umbilic points are simultaneously present on the wavefront.

Contours $0 < \mathfrak{D} < \mathfrak{C}_1 + \mathfrak{C}_2$ image at two conjugate levels (NYE [1978]) above and below the umbilic focusing level. For $\mathfrak{D} = \mathfrak{C}_1 + \mathfrak{C}_2$ one caustic is at infinity and the other at $1/\mathfrak{K}$, and for $\mathfrak{D} > \mathfrak{C}_1 + \mathfrak{C}_2$ one caustic is virtual and its conjugate is real, lying between $1/\mathfrak{K}$ and W .

Because of their size and shape, the “focal length” of the droplets (the height of the umbilic focusing level) is small. NYE [1978] studied the near field experimentally by focusing a microscope on successive planes above irregular horizontal droplets illuminated normally by a broadened parallel laser beam. There is indeed a unique plane of focus, containing many elliptic umbilic diffraction “stars” (Fig. 5.2a). As the plane of focus of the

microscope is altered, they unfold and interact by beak-to-beak events and swallowtails to give many-cusped figures close to the drop and in the far field (Figs. 5.2 b-d). To quote Nye, "In three dimensions, one has to picture a brilliant gothic palace: pointed arches of light, on a microscopic scale, with the tilted triangular columns of different sizes springing from a polished floor, which reflects their tracery of interactions . . .". NYE [1978] studied in detail the possible interactions between the different branches of the caustics emanating from different elliptic umbilic foci. Note that there are no lips events: these correspond to maxima of the \mathfrak{D} landscape, which are precluded by the surface tension condition (5.2). To reduce the complexity and show up systematic features, Nye studied the caustics from drops with controlled perimeters by filling with water appropriately shaped holes cut in opaque tape stuck to a glass microscope slide. This technique was used with a triangular hole to study the elliptic umbilic diffraction catastrophe (BERRY, NYE and WRIGHT [1979]; see § 3) and, as we shall shortly see, is proving to be an invaluable research tool for studying the caustics from water droplets with non-uniform internal pressure.

5.3. DROPLETS WITH NON-UNIFORM INTERNAL PRESSURE

In the caustics from an irregular droplet clinging to an inclined glass surface, all the catastrophes of codimension up to three are seen, including the hyperbolic umbilic (NYE [1979]). Patterns like Fig. 5.3 can be seen with the naked eye, close to a drop on a vertical surface, by viewing a distant point light source through a raindrop on a window or on one's spectacles. In two-dimensional plane sections of the near field of such caustics observed with a microscope (as described above), the most conspicuous new generic features are folds which are convex outwards and hyperbolic umbilic foci (Figs. 5.3 and 3.5c).

If the glass substrate is rotated in its own plane, the caustic does not rotate rigidly but continuously rearranges itself such that any fold which is convex outwards stays in the same position in the field of view. Clearly this is an effect of gravity: the drop perimeter remains fixed, but the thickest part of the drop will always be pendulous downwards. In drops not affected by gravity, folds must be concave outwards (BERRY [1976]).

As the plane of focus is changed, there is a complicated sequence of interactions between the various parts of the caustic. A transition like that



Fig. 5.3. The far field caustic from an irregular water drop on a vertical flat glass plate (courtesy of J. F. Nye).

between Figs. 5.4a and 5.4b is typical: it can be accomplished by changing the plane of focus, by tilting the glass substrate out of the vertical, or even dynamically "as can be verified when walking in the rain, wearing spectacles and viewing a distant light: at each pace the elliptic umbilic stars pierce the fold in striking unison" (NYE [1979]). The physically accessible "controls" are not necessarily all independent; their relation-

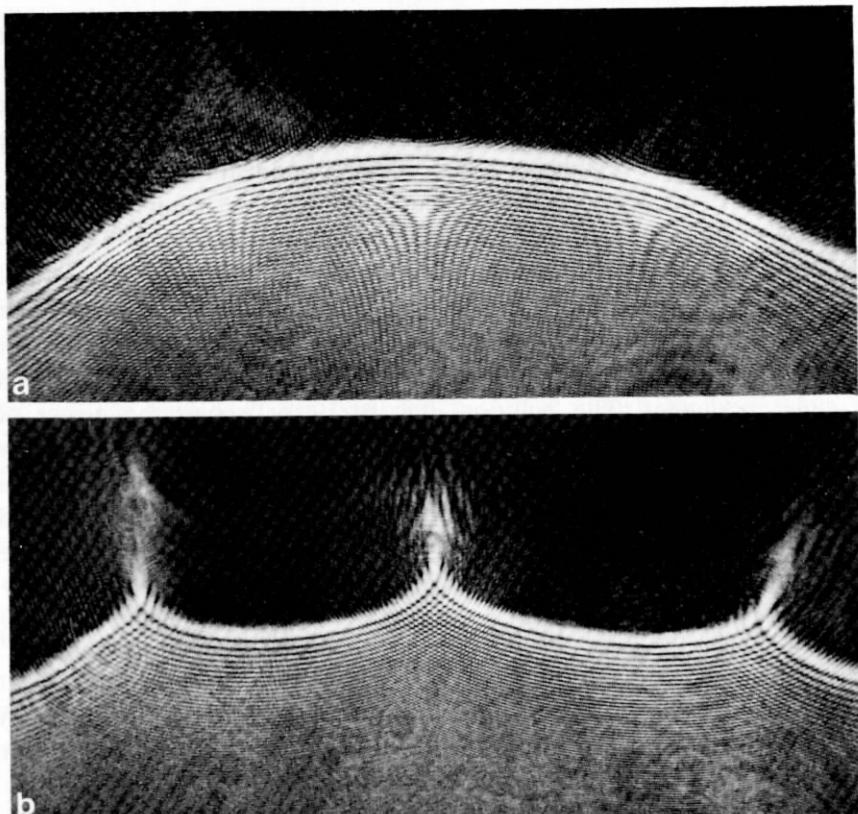


Fig. 5.4. (a, b) Near field caustics formed by the region of an irregular water drop whose shape is affected by gravity (courtesy of J. F. Nye).

ship with the control parameters of catastrophe theory need not be simple, but can usually be inferred from experimental evidence.

In spite of the natural irregularities of any drop, the caustics that it produces have a remarkable degree of organization, and NYE [1979] showed that this organization can be understood by regarding higher-dimensional catastrophes as organizing centres. By performing controlled experiments using the "hole in tape" method outlined above, rotating the substrate about its axes and letting gravity alter the drop shape without changing its perimeter, he obtained degenerate umbilic points on the surface corresponding to D_5 (the parabolic umbilic, $K = 4$) and E_6 ($K = 5$). These experiments have confirmed the geometry of these catastrophes (GODWIN [1971], CALLAHAN [1977]) and elucidated the characteristics of

the corresponding diffraction structure (§ 3; Figs. 3.7 and 3.8). In a randomly irregular drop, the degeneracy is removed and the patterns can be understood in terms of unfoldings of D_5 and E_6 . In addition to these catastrophes, there are persistent global features which point to there being a still higher organizing centre at work, conjectured by Nye [1979] to be X_9 ($K = 8$). Current work is confirming this (see below).

As we saw most explicitly for the umbilic classifications in § 4, catastrophe theory is about the local forms of the unfoldings of singularities: typically one works in terms of expansions about umbilic (or other singular) points. Nye makes no attempt to fit solutions to boundary conditions at the edge of the drop, except for the case of a drop with a circular perimeter. This solution, due to Berry, shows the power of catastrophe optics and the value of the concept of structural stability. For a thin droplet of density ρ , refractive index n and surface tension γ , on a substrate at an angle θ to the horizontal, with the x axis as in Fig. 5.5 and the y axis normal to the page, the equation governing the form of the wavefront W is

$$\nabla^2 f = \frac{n-1}{\gamma} (p + g\rho x \sin \theta), \quad (5.5)$$

where g is the acceleration of gravity. In dimensionless variables f' , p' ,

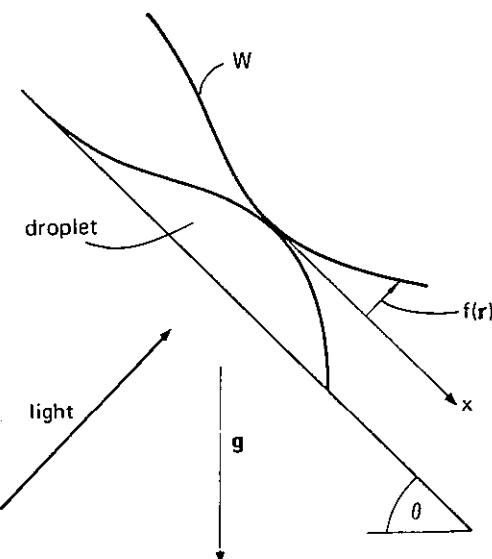


Fig. 5.5. Coordinates for a drop with non-uniform internal pressure.

(5.5) reduces to

$$\nabla^2 f' = p' + x \quad (5.6)$$

which, for a drop with a circular perimeter of radius b , has the exact solution

$$f' = \frac{1}{4}p'(x^2 + y^2) + \frac{1}{8}x(x^2 + y^2 - b^2). \quad (5.7)$$

The origin is an umbilic point, but f has non-zero slope there, so the umbilic focus is off-axis. The catastrophe classification (4.9) shows that the umbilic is hyperbolic. Without gravity, the drop would be a spherical cap with a structurally unstable point focus. Gravity removes the indeterminacy and produces a structurally stable hyperbolic umbilic focus: the geometrical caustic in the focal plane can be shown to be two lines, $X = \pm\sqrt{3}Y$, $X \leq 0$, which is a V shape with arms at 60° : see NYE [1979] for details. Fig. 3.5c was produced in this way. Structural stability guarantees that a drop in an imperfect circular or oval hole also produces a hyperbolic umbilic focus, with the arms of the V at an angle close to 60° . A good example is shown by TANNER [1978] in a study of the optics and motion of oil droplets.

The above case makes clear not only the identification of the off-axis lens aberration known as primary coma with the hyperbolic umbilic (Appendix 2) but the reason for the difficulty of avoiding it: the aberration is structurally stable!

NYE [1979] refers briefly to the “strikingly different” caustic patterns, only resolvable under high magnification, from small water droplets of such thickness that gravity affects their shape even when they are horizontal. A systematic study now in progress is showing that the central regions of such caustics can be understood in terms of the catastrophe X_9 , ($K = 8$). In the next section we shall see that X_9 also arises in the context of the caustic networks formed by refraction and reflection by the surface of moving water or rippled (bathroom) glass.

§ 6. Caustic Networks

6.1. NETWORKS AND JUNCTIONS

Mascrosopic caustic networks are a familiar sight as ever-changing patterns of lines of sunlight focused onto the bottom of a swimming pool or the sea bed after refraction by the wavy water surface (MINNAERT

[1954], STEVENS [1976]). They also occur in reflection, for example on the soffit of a bridge or the overhanging side of a boat; it is sometimes possible to see both refracted and reflected caustics from the same stretch of water. Refraction of starlight by refractive-index inhomogeneities caused by atmospheric turbulence produces caustic networks, which can be seen crossing the defocused image disc of a strongly twinkling star viewed under high magnification through an astronomical telescope. More amenable to detailed study are the stationary networks of caustics formed when a plane wave is reflected or refracted by rippled glass (BERRY [1976, 1978a], UPSTILL [1979]). Caustic networks can also occur in an electron microscope, when electrons are focused by an electrode in the form of a wire gauze (BARTZ, WEISSENBERG and WISKOTT [1956]).

These two-dimensional networks are of course sections of complicated caustic surfaces in three dimensions, which are likely to be organized by catastrophes of higher codimension, as there are many possible parameters which change the details of the caustics. It is natural to analyze these families of caustics in terms of the geometrical-optics configuration of their two-dimensional sections, using the paraxial optics of § 4 and keeping in mind that in most laboratory scale experiments (and also the observation of starlight caustics) we see the characteristic diffraction structures enumerated in § 3.

Apart from the rules related to the indices of umbilic points (§ 4), no general theorems about caustic networks are known. It is possible, however, to construct periodic and quasi-periodic models for wavy surfaces, whose calculated caustic networks agree well with many observations, both in terms of local structure and global configurations (UPSTILL [1979]). More generally, it has proved instructive to consider a network as being constructed from *junctions*, whose details are to be understood in terms of sections of catastrophes, and which are to be connected by fold lines in the simplest possible way. We shall illustrate these considerations by analyzing two common morphologies of caustic network (§§ 6.2 and 6.3) for which catastrophe optics reveals previously unsuspected fine structure.

6.2. SINUOUS NETWORKS

The pattern of overlapping wavy lines shown in Fig. 6.1 is a type of caustic network often seen in sunlight focused by rippling water. The lines

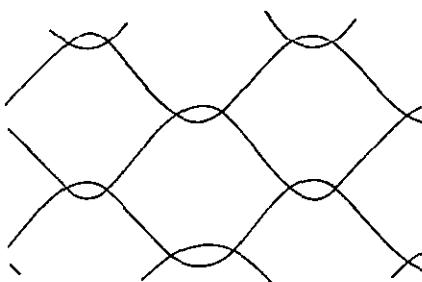


Fig. 6.1. A common morphology of caustic network formed by a wavy water surface, unresolved.

cannot all be single folds which merely cross one another, for if we attempt to label the regions between the lines according to the number of rays contributing, which changes by ± 2 across a fold, we find ambiguities, equivalent regions having different labels. This suggests that Fig. 6.1 is an imperfect resolution of a more complicated caustic network whose fine structure is blurred by the rapid motion of the water and by the half-degree angular diameter of the sun.

The first step in resolving Fig. 6.1 is to realise that below each wave in a single train of long-crested water waves is a caustic surface (Fig. 6.2) whose two sheets meet along a cusped edge and intersect the sea bed in a line pair. This makes it likely that some (but not all) of the lines in Fig. 6.1 are line pairs. The sinuous network repeats in two directions, suggesting a refracting surface consisting of two trains of water waves. This is a special case of a model surface proposed by UPSTILL [1979], consisting of a superposition of n sinusoids with amplitudes A , wave vectors k and phases χ , so that

$$f = \sum_{i=1}^n A_i \cos(k_i \cdot r + \chi_i). \quad (6.1)$$

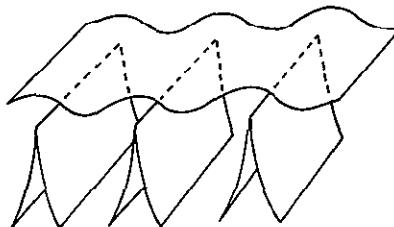


Fig. 6.2. Caustic surfaces below a single train of water waves.

For $n > 2$, f is periodic only if the \mathbf{k}_i are lattice vectors; but for $n = 2$, f is periodic for any linearly independent $\mathbf{k}_1, \mathbf{k}_2$. UPSTILL [1979] studied in detail the caustics for the case $n = 2$, verifying his analysis by experiments using both a ripple tank and sinusoidally grooved glass slides. All the possible two-dimensional caustic sections are encountered if we let both sinusoids have unit wavelength and one sinusoid have unit amplitude, varying the amplitude A of the other from 0 to 1, the angle θ between them from 0 to $\pi/2$, and the distance Z of the focal plane from 0 to $+\infty$ (this surface is invertible, so the virtual field $Z < 0$ is the same as the real field $Z > 0$).

Consider first the effect of increasing A , with θ and Z fixed. Clearly if $A = 0$ there is just one sinusoid and the caustics are just parallel line pairs as in Fig. 6.2. For $0 < A \ll 1$ the line pairs start to kink (Fig. 6.3a) but, as argued above, as A increases the fold lines do not simply overlap. Instead, there are swallowtail catastrophes which unfold asymmetrically (Fig. 6.3b). Finally, for $A \approx 1$, the swallowtails link up with one another by beak-to-beak events (Fig. 6.3c) to form the network shown unresolved in Fig. 6.1; far from being composed of sinuous lines, the network is thus revealed as an array of overlapping giant lip shapes.

Now consider Fig. 6.3c as part of a focal sequence. Figs. 6.4a–c show one unit cell of such a sequence for increasing Z , obtained experimentally by refracting a broadened parallel laser beam through two sinusoidally grooved slides at about 60° to one another, with $A \leq 1$. Below the plane of Fig. 6.4a is a lips event, whilst between Figs. 6.4a and 6.4b two swallowtails unfold, four cusps from neighbouring unit cells enter the field of view, and there are four beak-to-beak events. The configuration of Fig. 6.4b persists for a considerable range of Z . Between Figs. 6.4b and 6.4c there is another lips event at right angles to the first, and beyond Fig. 6.4c the four cusps approach the fold lines which are concave towards them to give four hyperbolic umbilic catastrophes in the far field. One unit of the periodic pattern is then a large diamond shape, each line being three coincident folds. As the far field is approached, however, neighbouring units overlap and the complete pattern becomes a confusion of caustic lines.

Finally, consider the effect of changing θ . For $\theta = 0$ the surface is a single sinusoid. If $0 < \theta \ll \pi/2$ caustic patterns in planes $Z = \text{constant}$ are highly elongated; under poor resolution the typical section of Fig. 6.4b looks like two triple junctions – common features of caustic networks (cf § 6.3). As θ increases, the caustics stretch in the Z direction: whilst the

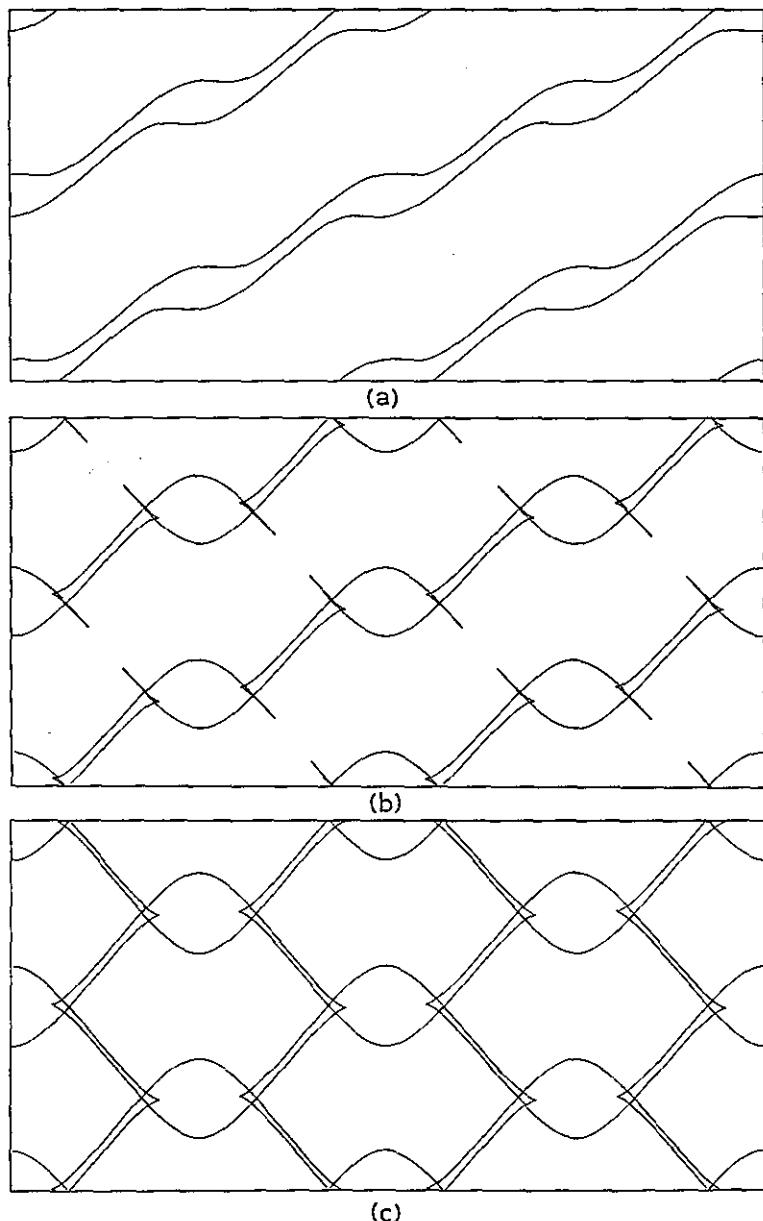


Fig. 6.3. Computed caustic sections from a surface consisting of a superposition of two sinusoids, for fixed θ and Z . (a) $0 < A \ll 1$, (b) $A < 1$, (c) $A \approx 1$.

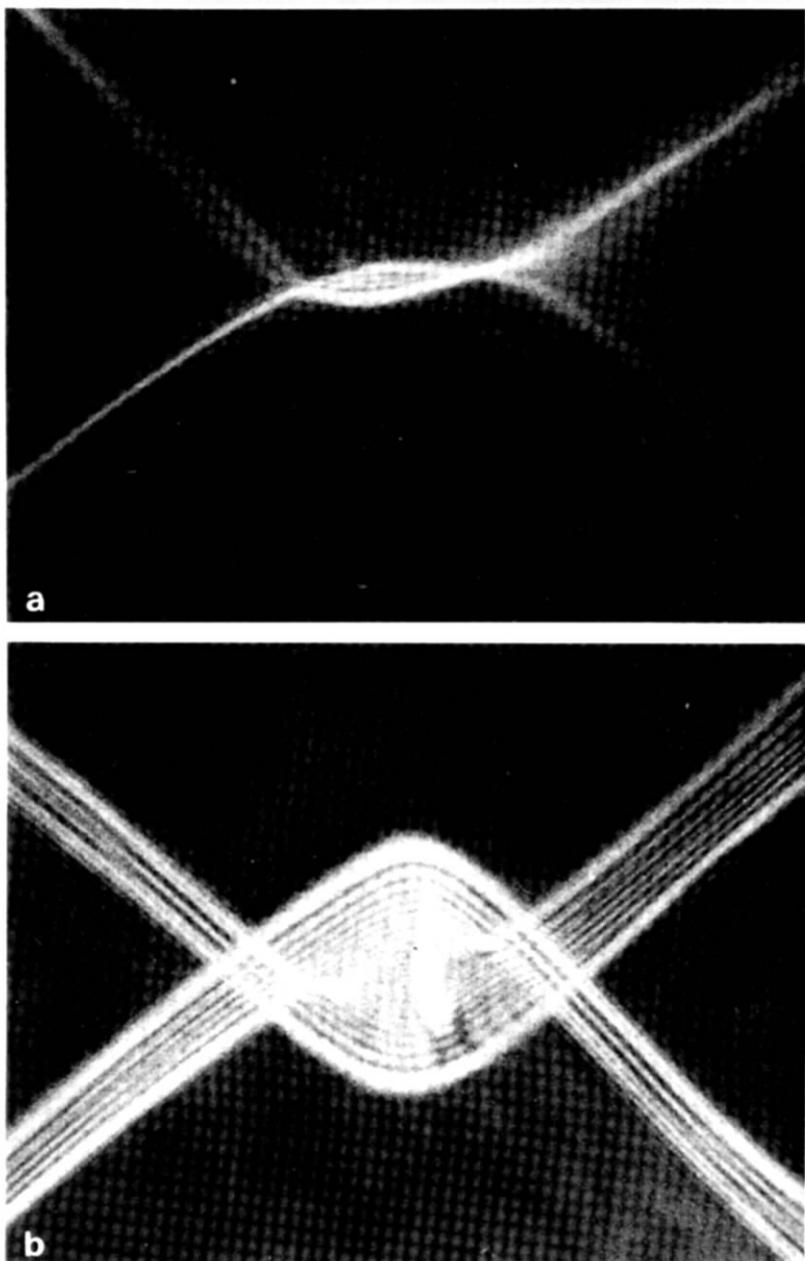


Fig. 6.4. Observation of one unit cell of caustic of refraction from two superposed sinusoidally grooved glass slides. Typical sequence of sections for fixed θ and A ; Z increasing from (a) to (c).

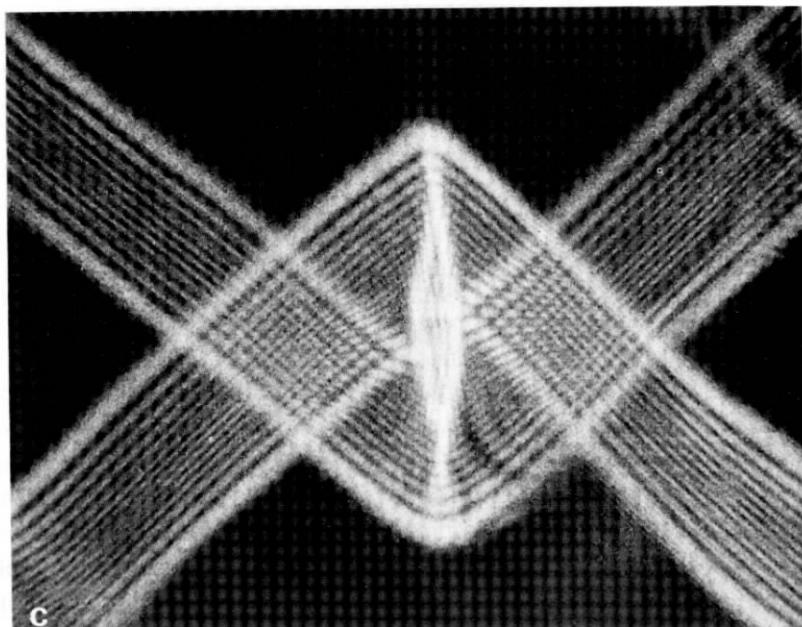


Fig. 6.4 (contd.)

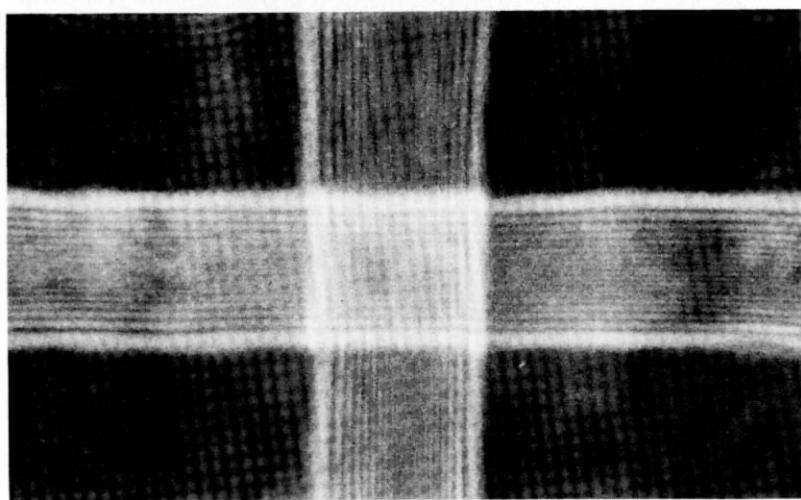


Fig. 6.5. Typical section, as Fig. 6.4 but with sinusoids at right angles.

first lips event (preceding Fig. 6.4a) and the hyperbolic umbilic foci (after Fig. 6.4c) remain fixed, the events between them occur at progressively lower Z . The limit $\theta = \pi/2$ is a degenerate case, for which all events occur simultaneously and there are hyperbolic umbilic singularities for a range of focusing levels, corresponding to lines of umbilics (rather than points) on f . It is as if there are two of the patterns from one sinusoid alone superposed orthogonally, except that within the region of intersection the caustic surface is not one but three coincident fold catastrophes; this is clearly shown experimentally by the greater intensity of that region (Fig. 6.5).

6.3. TRIPLE JUNCTIONS

The polygonal line pattern shown in Fig. 6.6 is another common type of caustic network. BERRY and NYE [1977] drew attention to the prevalence of triple junctions. This might seem unremarkable, since triple junctions are common, stable features of natural systems of lines and surfaces as diverse as mud cracks, animal markings and foams. But no catastrophe has a two-dimensional section consisting of three lines meeting at a point, and this leads to the conclusion that the caustic triple junction must be illusory, so the appearance of the network in Fig. 6.6, like the sinuous network in Fig. 6.1, is a consequence of imperfect resolution.

Such a junction could be produced by the meeting of three water waves, in which case each emerging line will be a line pair (cf Fig. 6.2). BERRY and NYE [1977] studied the fine structure of such junctions by refracting light through a glass "lollipop", on which three smooth grooves had been impressed in the molten state, and explored the three-dimensional caustic surface in a sequence of two-dimensional sections using a microscope. They found the central feature of the junction to be a section of an elliptic umbilic catastrophe, surrounded by a "foliage" of smoothly curving fold lines (Fig. 6.7). As the plane of focus approaches

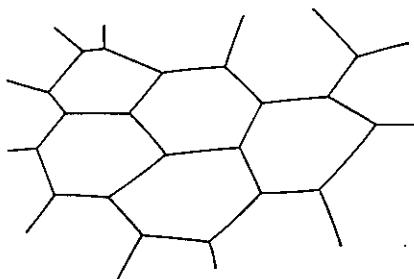


Fig. 6.6. Another common morphology of caustic network, unresolved.

the “lollipop” (Fig. 6.8), the elliptic umbilic shrinks to its singular point and unfolds again, one or more of its cusps pierce the “foliage”, and the unfolded elliptic umbilic and folds interact via beak-to-beak events and swallowtails, leaving three line pairs ending in inward pointing cusps which recede from the field of view.

These triple junctions link up to form a network resembling chain mail. The existence of such networks can be confirmed experimentally by producing caustics of refraction from a type of bathroom window glass (Pilkington’s “Atlantic”), as shown in Fig. 6.9. However, it is impossible for all the umbilic catastrophes produced by an undulating wavefront to be of elliptic type; this follows from the arguments outlined in § 4.3, according to which all elliptic umbilics have index $-\frac{1}{2}$, whereas the surface has net index zero. Therefore there must be hyperbolic umbilic points, of index $+\frac{1}{2}$, on the surface: where do they focus? Recall that the caustics have continuity through the far field (§ 5.2). This implies that, provided there are no hyperbolic umbilics in the real field, the missing ones must lie in the virtual field. This possibility is exemplified by “Atlantic” glass,

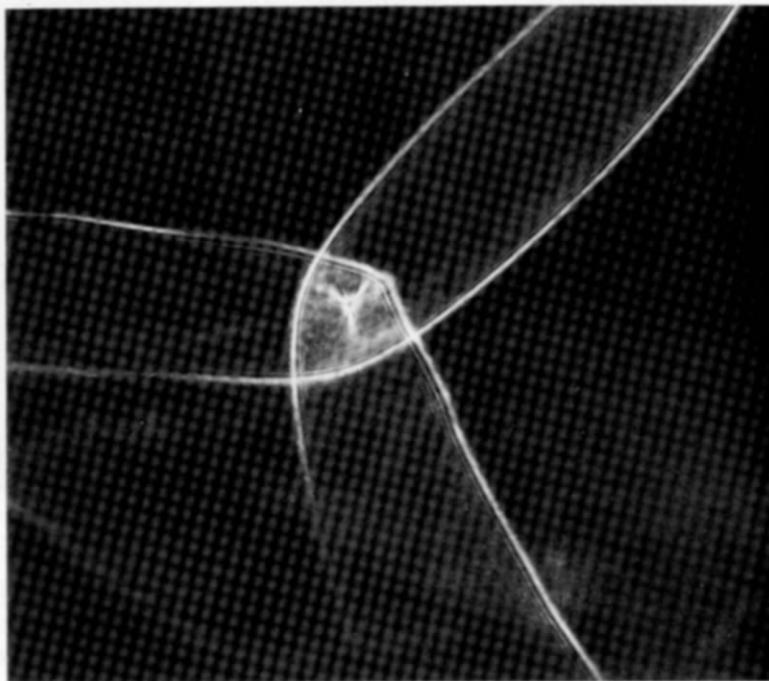


Fig. 6.7. An elliptic umbilic centred caustic triple junction resolved (courtesy of J. F. Nye).

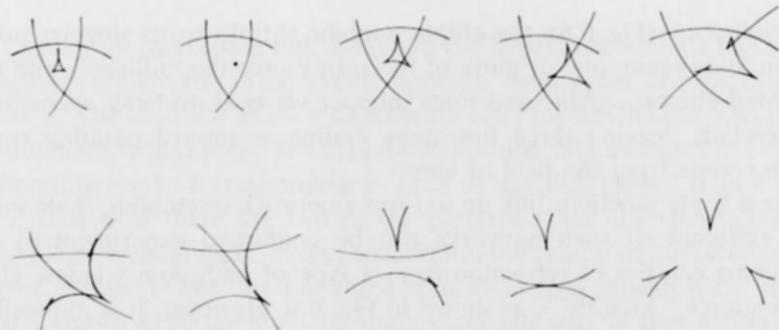


Fig. 6.8. Sequence of sections showing how the caustic junction of Fig. 6.7 is observed to evolve for decreasing Z .

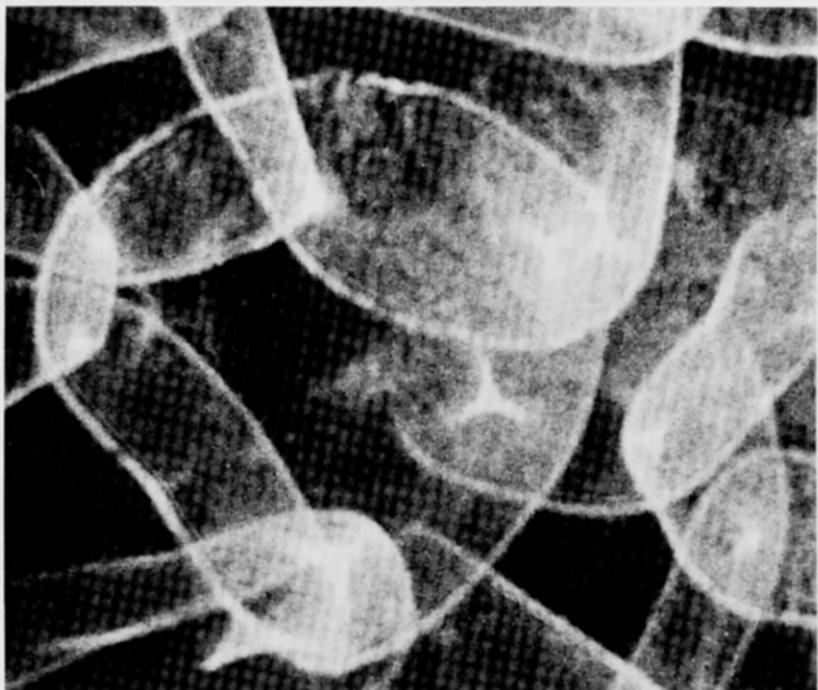


Fig. 6.9. Network of elliptic umbilic centred triple junctions in the real caustic of refraction from Pilkington's "Atlantic" glass.

whose virtual near field caustics consist of a lattice of spiky "spangles" each containing two hyperbolic umbilics (Fig. 6.10). (The existence of additional umbilics of either sort is not precluded, but their net index must be zero.) Note that, in contrast to the case considered in § 6.2, real and virtual fields are different: therefore this surface is non-invertible.

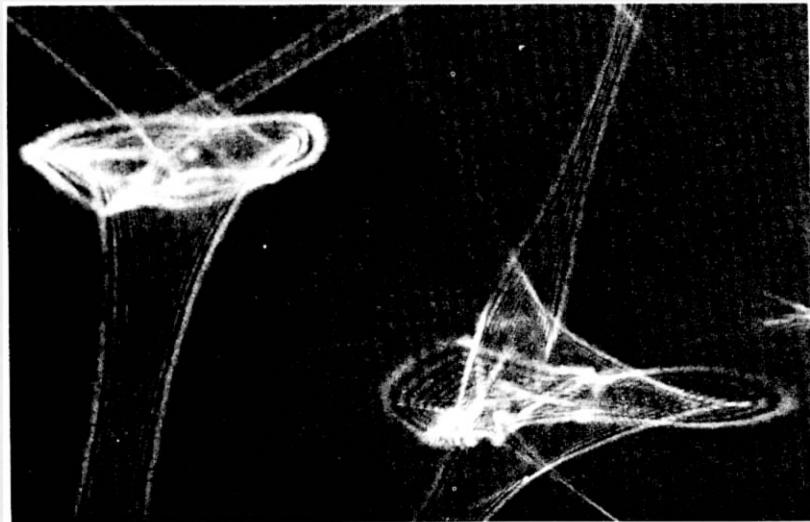


Fig. 6.10. Virtual caustic of refraction from Pilkington's "Atlantic" glass, showing hyperbolic umbilics.

An obvious global model surface corresponding to the meeting of three trains of water waves is eq. (6.1) with $n = 3$. We have confirmed that in the appropriate region of the (A, \mathbf{k}, χ) parameter space this model gives both the "chain mail" network and the lattice of "spangles", correctly reproducing the details of the entire observed real and virtual focal sequence. Moreover, this model describes various other experimentally observable configurations and transitions, including the transition from the networks with elliptic umbilic centred triple junctions to the sinuous networks discussed in § 6.2 (UPSTILL and BERRY, in preparation). Recall that two trains of water waves superposed at a fairly acute angle gave caustics which under poor resolution looked like triple junctions with a fine structure comprising only one cusp (cf Fig. 6.4b). Therefore three waves are not necessary for the occurrence of caustic triple junctions; it is the curvature landscape that matters, not the height landscape (§ 4.2).

6.4. JUNCTIONS ARE ORGANIZED BY 0X_9

The extended caustic networks generated by the surface (6.1) are not part of the unfolding of any single catastrophe (except perhaps one with $K = \infty$). We are finding, however, that all fine structures yet discovered in

the junctions of caustic networks are sections of the universal unfolding of the compact subset 0X_9 of the catastrophe X_9 , (ARNOL'D [1975], ZEEMAN [1977], CALLAHAN [1978]), which has codimension 8. Some plane sections of X_9 possessing discrete rotation symmetry were calculated by LENZ and MECKING [1961] (before the development of catastrophe theory) to simulate caustics in electron optics (see also LENZ [1956]).

It seems that the importance of 0X_9 in the context of caustic networks stems from the fact that it is the simplest catastrophe of corank 2 which is *compact*, i.e. for which the generating function $\Phi(s; C)$ has an absolute minimum, rising to $+\infty$ for any control point C , as $s \rightarrow \infty$ in all directions in state space. The generating function ϕ for an unbounded initial wavefront must also be compact because through any observation point there is an absolutely minimal ray (POSTON [1980]). A non-compact ϕ obtained by expansion about some point can only give the local form of the caustic. By further expanding until ϕ is compact, we obtain the simplest possible models for significant units of caustic networks, such as junctions. Thus we expect such units to be organized (at the least) by 0X_9 , and we are currently engaged in a comprehensive study of this. Of course, it is possible to obtain even higher compact catastrophes describing larger features of networks, such as several linked junctions.

§ 7. Statistical Catastrophe Optics

7.1. GAUSSIAN RANDOM WAVEFRONTS

The conditions under which light propagates are frequently specified not precisely but only in a statistical sense. Examples are starlight twinkling through a randomly turbulent atmosphere, and sunlight glittering on randomly rippling water. When the wave number κ is large enough, catastrophe optics is applicable and gives a classification of caustics and diffraction catastrophes and a description of caustic networks. The additional feature of randomness immediately suggests questions of a statistical character. What is the average number of rays passing through a given point? What is the average density of umbilic foci in space? How do the wave intensity fluctuations depend on κ ? The study of such problems is a promising new field of research, lying in the intersection of wave theory, topology, and probability.

It is of course necessary to choose an ensemble over which averages are

to be taken; the appropriate choice depends on the physical system. In §§ 7.2 and 7.3 we restrict ourselves to the only case where progress has been made in elucidating the statistical geometry of rays and caustics: the paraxial optics of § 4 with the height function f of the initial wavefront W chosen to be a *Gaussian random function* of \mathbf{r} . This is a superposition of infinitely many sinusoids with random phases, given by eq. (6.1) with $n \rightarrow \infty$; each choice of phases χ gives a sample f in the ensemble and hence a sample family of rays and caustics. The mean value $\langle Q \rangle$ of any quantity Q that depends on the choice of phases is defined as the average over all phases

$$\langle Q \rangle = \lim_{n \rightarrow \infty} \frac{1}{(2\pi)^n} \int_0^{2\pi} d\chi_1 \cdots \int_0^{2\pi} d\chi_n Q(\chi). \quad (7.1)$$

By standard noise theory (RICE [1944, 1945], LONGUET-HIGGINS [1957]), $\langle f \rangle = 0$, f and its derivatives are normally distributed, and odd and even derivatives are statistically independent (so that, for example, $\langle ff_x \rangle = 0$). The statistics of f are completely determined by the autocorrelation function $\rho(r)$ or the power spectrum $E(k)$, defined by

$$\rho(r) \equiv \langle f(\mathbf{r}_0) f(\mathbf{r}_0 + \mathbf{r}) \rangle \equiv \int_{-\infty}^{\infty} dk E(k) \exp[ik \cdot \mathbf{r}] \quad (7.2)$$

(these functions do not depend on \mathbf{r}_0).

We begin in § 7.2 with the case where W is corrugated, i.e. where f depends only on one coordinate x so that the ray families are effectively two-dimensional; it is possible to obtain a full statistical description of the rays and caustics. The general case, where f varies in two dimensions so that the ray families are three-dimensional, is considered in § 7.3. Before the development of catastrophe theory, the topology of ray families was not understood. Nevertheless, LONGUET-HIGGINS [1956, 1957, 1958, 1959, 1960a,b] pioneered techniques for evaluating the averages required to describe the statistical geometry, devoting considerable attention to the effects of the time-dependence of W . His methods were recently employed by BERRY and HANNAY [1977] to calculate the statistics of umbilic points. We confine ourselves in §§ 7.2 and 7.3 to statements of results, which were derived from the material in these papers. Unless otherwise indicated, the results are new.

In § 7.4 we consider statistics that are not purely geometric: the wave intensity fluctuations. These scale with κ according to universal laws, independent of any particular statistical ensemble or propagation model.

7.2. STATISTICAL GEOMETRY IN TWO DIMENSIONS

If W is corrugated, there is only one state variable x and from eq. (4.3) the equation determining which ray(s) pass through the control point X, Z (Fig. 7.1) is

$$X = x - Zf_x. \quad (7.3)$$

Stable caustics must be organised by the cuspoid catastrophes, since only these have corank one. For a typical f in the ensemble, the XZ plane contains fold lines, which from eq. (4.4) satisfy

$$Zf_{xx} = 1 \quad (7.4)$$

(as well as eq. (7.3)), and cusp points, which satisfy the additional condition

$$f_{xxx} = 0. \quad (7.5)$$

The autocorrelation ρ (eq. (7.2)) depends only on x and determines the statistical geometry via the quantities

$$F_1 \equiv \left[\left\langle \left(\frac{df}{dx^i} \right)^2 \right\rangle \right]^{\frac{1}{2}} = \left[(-1)^i \frac{d^{2i}}{dx^{2i}} \rho(0) \right]^{\frac{1}{2}} \quad (7.6)$$

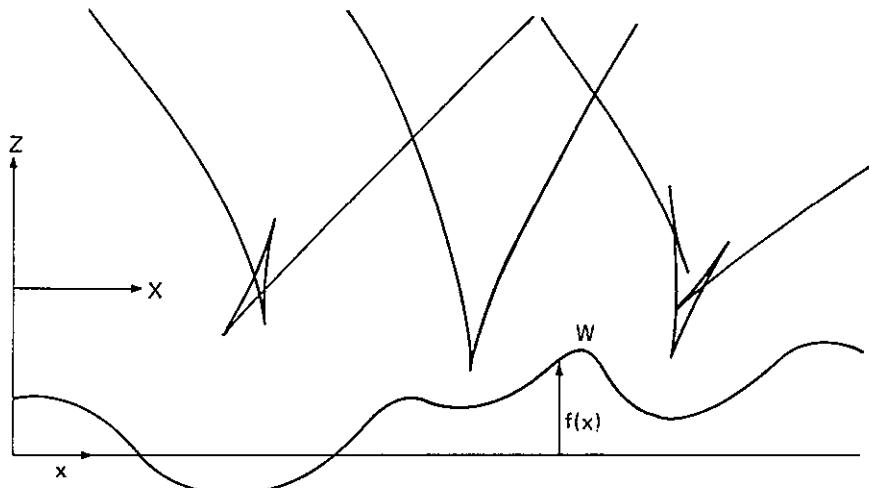


Fig. 7.1. Clusters of cusps in the caustics above a corrugated wavefront.

Consider first the most elementary average, namely the *mean number of rays* passing through a point at height Z . This is independent of the X coordinate of the point, and will be denoted by $N_r(Z)$. The rays must satisfy (7.3), so that

$$\begin{aligned} N_r(Z) &= \left\langle \int_{-\infty}^{\infty} dx \delta(X - x + Zf_x) |1 - Zf_{xx}| \right\rangle \\ &= \langle |1 - Zf_{xx}| \rangle, \end{aligned} \quad (7.7)$$

where the second equality follows from the statistical independence of f_x and f_{xx} . LONGUET-HIGGINS [1960a] evaluated the average over f_{xx} , obtaining

$$N_r(Z) = \operatorname{erf}[(ZF_2\sqrt{2})^{-1}] + \frac{2ZF_2}{\sqrt{(2\pi)}} \exp[-(2Z^2F_2^2)^{-1}], \quad (7.8)$$

where erf denotes the error function. Thus N_r increases from 1 close to W to $\sim Z$ far from W .

Next consider the *mean linear density of fold crossings* at height Z , denoted by $N_f(Z)$. It is convenient to count the folds crossed in a very large distance L and then transform from X to x ; bearing in mind that folds satisfy (7.4), we get

$$\begin{aligned} N_f(Z) &= \lim_{L \rightarrow \infty} \left\langle \int_0^L dx \delta(1 - Zf_{xx}) |Zf_{xxx}| \right\rangle \\ &= \langle \delta(1 - Zf_{xx}) \rangle \langle |Zf_{xxx}| \rangle. \end{aligned} \quad (7.9)$$

This reduces to

$$N_f(Z) = \frac{F_3}{\pi F_2} \exp[-(2Z^2F_2^2)^{-1}]. \quad (7.10)$$

The asymptotic value, $F_3/\pi F_2$, is the density of inflections of f from which the far field caustics originate.

Clearly, the final statistic in this hierarchy is the *areal density of cusp points* at height Z , denoted by $N_c(Z)$. Arguments similar to those used to obtain N_f , together with the cusp conditions (7.4) and (7.5), lead to

$$\begin{aligned} N_c(Z) &= \lim_{L \rightarrow \infty} \left\langle \int_0^L dx \delta(f_{xxx}) |f_{xxxx}| \delta(Z - 1/f_{xx}) \right\rangle \\ &= \langle \delta(f_{xxx}) \rangle \langle \delta(f_{xx} - 1/Z) |f_{xxxx}| \rangle / Z^2. \end{aligned} \quad (7.11)$$

This gives

$$N_c(Z) = \frac{2F_3\sqrt{(\mathcal{N}^2 - 1)} \exp [-(2Z^2 F_2^2)^{-1}]}{Z^2 F_2^2 (2\pi)^{\frac{3}{2}}} \\ \times \left\{ \exp [-(2Z^2 F_2^2 (\mathcal{N}^2 - 1))^{-1}] + \frac{1}{ZF_2} \left(\frac{\pi}{2(\mathcal{N}^2 - 1)} \right)^{\frac{1}{2}} \operatorname{erf} [(ZF_2\sqrt{(\mathcal{N}^2 - 1)^{-1}})] \right\}, \quad (7.12)$$

where \mathcal{N} is the quantity

$$\mathcal{N} \equiv F_4 F_2 / F_3^2, \quad (7.13)$$

whose significance will soon be apparent. N_c vanishes near to and far from W , reaching a maximum for all \mathcal{N} when $Z \sim F_2^{-1}$; not surprisingly, the cusps are most dense near the level where the centres of curvature of W are most dense.

Globally, the fold lines consist of branches beginning and ending in the far field (Fig. 7.1). Each branch contains a *cluster* of an odd number of cusps, corresponding to the zeroes of f_{xx} between successive zeroes of f_{xx} . It is not hard to show that the mean number of cusps in a cluster is

$$\frac{2}{N_f(\infty)} \int_0^\infty dZ N_c(Z) = \mathcal{N}, \quad (7.14)$$

where \mathcal{N} is also given by eq. (7.13).

One gets a very clear picture of the statistical geometry of the caustics by evaluating \mathcal{N} for different functions ρ . Since W is smooth, ρ must be an even function of x , and so

$$\rho(x) = \sum_{n=0}^{\infty} a_n x^{2n}. \quad (7.15)$$

Then (7.13) and (7.6) give

$$\mathcal{N} = \left(\frac{28}{15} |a_2 a_4 / a_3^2| \right)^{\frac{1}{2}}. \quad (7.16)$$

The simplest case is the (nonrandom) sinusoidal wavefront, for which $\rho = F_0^2 \cos \alpha x$ and (7.16) gives $\mathcal{N} = 1$. If the correlations decay in a single-scale Gaussian manner, so that $\rho = F_0^2 \exp [-\alpha x^2]$, then $\mathcal{N} = \sqrt{\frac{7}{5}} = 1.183$; if the decay is Lorentzian, so that $\rho = F_0^2 / (1 + \alpha x^2)$, then $\mathcal{N} = \sqrt{\frac{28}{15}} = 1.366$. Bearing in mind that the number of cusps must be odd, we see that most clusters contain just one cusp. For wavefronts with two or more scales, however, it is not difficult to construct cases where \mathcal{N} is arbitrarily large.

7.3. STATISTICAL GEOMETRY IN THREE DIMENSIONS

If W is curved in two dimensions, there are two state variables $\mathbf{r} = (x, y)$. The rays in control space X, Y, Z satisfy (4.3) and the caustics satisfy (4.4) as well. For a generic f the highest singularities are elliptic and hyperbolic umbilic foci, and these satisfy the additional conditions (4.7).

In the most general case, the correlations on W will be anisotropic. Here, however, we restrict ourselves to the case of isotropic disorder, where ρ and E (eq. (7.2)) are functions only of the lengths r and k of the vectors \mathbf{r} and \mathbf{k} . The statistical geometry is described in terms of the moments

$$M_i \equiv 2\pi \int_0^\infty dk k^{i+1} E(k). \quad (7.17)$$

Note that $M_4 = \langle (\nabla^2 f)^2 \rangle$.

The *mean number of rays* passing through a point at height Z is given by (cf (7.7))

$$\begin{aligned} N_r(Z) &= \left\langle \int \int \int_{\mathbf{w}} d\mathbf{r} \delta(\mathbf{R} - \mathbf{r} + Z\nabla f) \left| \det \frac{d}{d\mathbf{r}} (\mathbf{r} - Z\nabla f) \right| \right\rangle \\ &= \langle 1 - Z\nabla^2 f + Z^2(f_{xx}f_{yy} - f_{xy}^2) \rangle. \end{aligned} \quad (7.18)$$

This average was evaluated by LONGUET-HIGGINS [1960a] as an estimate of the number of reflected images of the sun seen from different heights above the sea. He obtained the result

$$N_r(Z) = 1 + \frac{M_4 Z^2}{2\sqrt{3}} \exp[-4/(3M_4 Z^2)]. \quad (7.19)$$

Thus N_r increases from 1 close to W to $\sim Z^2$ far from W .

In a systematic development, the next statistic would be the average number of times a line of unit length pierces fold surfaces, or the average length of fold cutting unit area, followed by the average number of times unit area is pierced by ribs, and the spatial density of swallowtail points. However, all these cuspid statistics are extraordinarily difficult to calculate. One result which can be derived relatively easily is that as $Z \rightarrow \infty$ the mean length of fold cutting unit area is proportional to Z ; this explains the confusion of caustic lines in the far field of extended wavefronts, noted in § 6, and contrasts with the two-dimensional case for which the corresponding quantity $N_f(Z)$ (eq. (7.10)) saturates as $Z \rightarrow \infty$.

Fortunately, it is not difficult to calculate the *spatial density of umbilic foci* at height Z , denoted by $N_u(Z)$. By analogy with (7.11), use of (4.7) and (4.4) gives

$$\begin{aligned} N_u(Z) &= \lim_{A \rightarrow \infty} \left\langle \iint_A d\mathbf{r} \delta(f_{xx} - f_{yy}) \delta(f_{xy}) \right. \\ &\quad \times \left. \left| \det \frac{\partial(f_{xx} - f_{yy}, f_{xy})}{\partial(x, y)} \right| \right\rangle \\ &= \langle \delta(f_{xx} - f_{yy}) \delta(f_{xy}) \delta(f_{xx} - 1/Z) \rangle \\ &\quad \times \langle |f_{xxx}f_{yyy} - f_{xxy}^2 + f_{xxy} - f_{xxy}^2| \rangle / Z^2. \end{aligned} \quad (7.20)$$

This leads to the result

$$N_u(Z) = \frac{M_6 \exp[-2/(M_4 Z^2)]}{(2\pi M_4)^3 Z^2}. \quad (7.21)$$

N_u vanishes near to and far from W , and reaches a maximum at $Z = \sqrt{(2/M_4)}$, showing that umbilic foci are most dense where the centres of principal curvature of W are most dense.

The *density of umbilic points* on W is simply

$$\int_{-\infty}^{\infty} N_u(Z) dZ = M_6 / (4\pi M_4). \quad (7.22)$$

On the average, equal numbers of these umbilics have indices $+\frac{1}{2}$ and $-\frac{1}{2}$; this follows from the second index rule in § 4.3. However, as suggested by the inequalities in § 4.3, elliptic umbilics are less common than hyperbolic ones. By calculating an average involving the catastrophe classification function (4.9), BERRY and HANNAY [1977] showed that the fraction α_E of umbilics of elliptic type is

$$\alpha_E = \frac{9}{50} + \frac{8}{\pi} \int_{\frac{1}{2}}^1 du \frac{u(1-u^2)}{(1+u^2)^3} \arcsin \left[\left\{ \frac{(1-u)^2(3u+1)}{16u^3} \right\}^{\frac{1}{2}} \right] = 0.268. \quad (7.23)$$

This result does not depend on the nature of the correlations on W , provided these are isotropic.

7.4. RANDOM DIFFRACTION CATASTROPHES IN TWINKLING LIGHT

A random structure S produces random caustics, as in the special case just discussed. Associated with these caustics are random waves ψ . The

statistics of ψ as $\kappa \rightarrow \infty$, in particular the probability distribution of the intensity $|\psi|^2$, are not well understood theoretically. Experimentally, it is convenient to measure the moments \mathcal{I}_m of the fluctuating intensity, defined by

$$\mathcal{I}_m \equiv \langle |\psi|^{2m} \rangle, \quad (7.24)$$

where $\langle \rangle$ denotes averaging over the ensemble of S . If ψ were a Gaussian random function of position and/or time, \mathcal{I}_m would be equal to $\mathcal{I}_1 m!$ for all cases considered here. For large κ , however, the intensity fluctuations are much stronger than those of a Gaussian wave, even when the randomness of S is Gaussian as in the model of § 7.1; these strong fluctuations were measured by JAKEMAN, PIKE and PUSEY [1976] in twinkling starlight.

Conventional techniques based on perturbation theory (USCINSKI [1977]) fail completely for strong fluctuations. Elaborate analysis for paraxial wave propagation from a Gaussian random wavefront (§ 4 and § 7.1), by SHISHOV [1971], BUCKLEY [1971a,b] and JAKEMAN and McWHIRTER [1977] has led to an understanding of \mathcal{I}_2 and shown that $\mathcal{I}_2 \sim \ln \kappa$ as $\kappa \rightarrow \infty$. Therefore \mathcal{I}_2 , and a fortiori the higher moments, diverge in the geometrical-optics limit.

As has long been known, the non-Gaussian strong fluctuations in ψ and associated divergences of \mathcal{I}_m are due to focusing (MERCIER [1962], SALPETER [1967]). When $\kappa = \infty$, $|\psi|$ is infinite on the caustics of any ray family in the ensemble; during a time, space or ensemble average these infinities pass across the radiation detector. It follows from conservation of energy, and can also be shown directly, that these infinities in the intensity $|\psi|^2$ are integrable, so that \mathcal{I}_1 is finite. But higher powers of $|\psi|^2$ have non-integrable divergences, so that $\mathcal{I}_{m \geq 2}$ diverge. Prior to catastrophe theory, the elaboration of this idea was frustrated by lack of understanding of the geometry of caustics.

In practice several factors prevent \mathcal{I}_m being infinite. The source may be spatially and temporally incoherent, as with the finite size and polychromaticity of the sun, and this blurs caustic networks on the sea bed (§ 6). Most fundamental, however, is the finite value of κ , which causes the divergences of \mathcal{I}_m to be softened by diffraction.

It was suggested by BERRY [1977] that a natural measure of this effect is the set of *critical exponents* ν_m , defined by

$$\nu_m \equiv \lim_{\kappa \rightarrow \infty} d(\log \mathcal{I}_m) / d(\log \kappa). \quad (7.25)$$

Up to logarithmic factors such as appears in \mathcal{J}_2 , the existence of the ν_m implies the scaling laws

$$\mathcal{J}_m \sim \kappa^{\nu_m}. \quad (7.26)$$

The ν_m can be determined by catastrophe theory; we give a simplified version of the argument presented by BERRY [1977].

The main result is that as m increases, the exponents depend on catastrophes of ever higher codimension K . To see that this must be so, consider the familiar case where the \mathcal{J}_m are measured by time-averaging the intensity at fixed position \mathcal{R} . Over long times, diffraction catastrophes of arbitrarily large K will pass arbitrarily close to \mathcal{R} . The higher catastrophes are rare but cause large localized fluctuations in $|\psi|^2$. But large rare fluctuations are precisely what dominate high moments \mathcal{J}_m , which therefore depend on catastrophes of large K .

To calculate the ν_m , a scaling argument is employed. This depends crucially on the fact that the ensemble of S is parameterized by a large number of variables C (e.g. the phases χ in § 7.1). ψ depends on C , which therefore constitutes a tremendous augmentation of control space, fully explored during the ensemble averaging for \mathcal{J}_m which (cf eq. (7.1)) is given by

$$\mathcal{J}_m = \int \cdots \int dC P(C) |\psi(C)|^{2m}, \quad (7.27)$$

where $P(C)$ is the density of realizations of S over the ensemble. For large κ the integral will be dominated by those values of C lying on caustics. \mathcal{J}_m will be the sum of contributions \mathcal{J}_{mj} from the different catastrophes in the hierarchy of § 2.2, here labelled j . Each contribution scales differently with κ and can be written

$$\mathcal{J}_{mj} \sim \kappa^{\nu_{mj}}, \quad (7.28)$$

where ν_{mj} is thus the exponent governing the contribution of the j th catastrophe to the m th moment.

ν_{mj} is determined by two quantities describing the architecture of the j th diffraction catastrophe as discussed in § 3.4, namely the singularity index, here written β_j and the fringe index, here written γ_j . The contribution \mathcal{J}_{mj} is estimated from eq. (7.27) as follows:

$$\begin{aligned} \mathcal{J}_{mj} &\sim (\max |\psi|)^{2m} \times \left(\frac{\text{hypervolume of diffraction}}{\text{maximum in control space}} \right) \\ &\sim \kappa^{2m\beta_j} \kappa^{-\gamma_j}, \end{aligned} \quad (7.29)$$

whence (7.28) gives

$$\nu_{mj} = 2m\beta_j - \gamma_j. \quad (7.30)$$

Eqs. (7.28) and (7.30) establish how each catastrophe contributes to each moment. Summing over all catastrophes gives the m th moment as

$$\mathcal{J}_m \sim \sum_i B_{mj} \kappa^{\nu_{mj}} \quad \text{as } \kappa \rightarrow \infty, \quad (7.31)$$

where B_{mj} are constants. The dominant term is clearly the one with the largest ν_{mj} , so that from (7.26) the critical exponents are given by

$$\nu_m = \max_j (2m\beta_j - \gamma_j). \quad (7.32)$$

Eq. (7.32) is the main result. Calculation of the ν_m (a tedious exercise) shows that the maximum over j does actually exist: for fixed m , ν_{mj} first increases and then decreases as catastrophes with increasing K are considered. The resulting ν_m are sets of rational numbers depending only on the dimensionality of the space in which ψ propagates. The reason for this is that although K can be arbitrarily large, the corank (§ 2.2) cannot exceed the dimensionality of the wavefronts. Therefore for waves in the plane only catastrophes of corank one (cuspoids) can contribute, while for waves in space only the catastrophes of corank one or two can contribute; the extra singularities make ν_m larger in the latter case. Apart from this dimension dependence, the ν_m are universal: they do not depend on the details of S , only on the fact that it is smooth and its randomness is described by many parameters C .

Table 3 shows the first few ν_m , and the dominant catastrophes. The value $\nu_2 = 0$ does not mean that \mathcal{J}_2 is not singular as $\kappa \rightarrow \infty$; in fact, as already mentioned $\mathcal{J}_2 \sim \ln \kappa$. These results constitute testable predictions

TABLE 3
Critical exponents ν_m and dominant catastrophes for $2 \leq m \leq 5$

m	ν_m (waves in the plane)	Dominant catastrophe	ν_m (waves in space)	Dominant catastrophe(s)
2	0	A_2	0	A_2
3	1/3	A_2	1/3	A_2 and D_4
4	3/4	A_3	1	D_4
5	5/4	A_3	5/3	D_4 and E_6

about the wavelength dependence of the moments \mathcal{I}_m of random short waves. Usually \mathcal{I}_m are measured as functions of other parameters, such as distance from a turbulent medium, or strength of turbulence, but it should be possible for ν_m to be measured directly on the basis of the definition (7.25).

For $m > 5$, serious difficulties arise in the calculation of ν_m for waves in space because of the appearance of catastrophes with modality (ARNOLD [1975]). Making plausible assumptions, BERRY [1977] was able to calculate critical exponents up to ν_{13} .

§ 8. Concluding Remarks

Catastrophe optics belongs in a more general context, in which wave motion is viewed, unconventionally, in terms of the contrast and interplay between the morphologies of three extreme regimes (BERRY [1979a]). Firstly, if the wavelength λ is small in comparison with scales of variation of diffracting objects or refracting media, the wavefield is dominated by the caustics and associated diffraction patterns which have been the subject of this review. Secondly, when waves propagate in environments which can be modelled by a hierarchy of scales extending to the infinitely small, caustics cannot occur and the limit $\lambda \rightarrow 0$ is not geometrical optics; MANDELBROT [1977] calls such hierarchical structures *fractals*, and BERRY [1979b] calls the corresponding waves *diffractals*. And thirdly, when waves are explored on the scale of λ , the principal features are wavefronts, which are dominated by their singularities, in the form of lines in space; these are the wavefront dislocations described in Appendix 1.

Finally, we emphasize that catastrophe optics, although founded on differential topology, is now firmly part of physics. Its development has been characterized by an intimate combination of observation, experiment, analysis and computation. This has been especially fertile in studies of diffraction catastrophes (§ 3), liquid droplet optics (§ 5) and caustic networks (§ 6). The successes so far achieved amply justify the injection into optics of concepts like structural stability and the unfolding of a singularity, which must surely be unfamiliar and strange to most physicists. Catastrophe optics is bringing within the compass of our understanding whole realms of phenomena with the pleasant property, uncommon in physics nowadays, that they are accessible not only in the laboratory but also in nature, where they can be seen by everybody.

Acknowledgments

It is a pleasure to thank our colleagues, Dr. J. H. Hannay, Professor J. F. Nye, F.R.S., and Dr. F. J. Wright, for many invaluable discussions and for their assistance with the preparation of this review.

Appendix 1: Wavefronts

It is natural to define wavefronts as surfaces of constant phase. There are, however, two ways of defining phase, leading to two sorts of wavefront whose geometric properties – in particular their singularities – are very different. Firstly, there are *diffraction wavefronts*, defined in terms of the phase $\kappa\chi$ of the exact complex scalar wave function ψ (eq. (3.2)) by

$$\kappa\chi(\mathcal{R}) = A \mod 2\pi, \quad (\text{A1.1})$$

where A is constant. Secondly, there are *geometrical wavefronts*, defined in terms of the phases $\kappa\mathcal{S}_\mu$ of contributions to the short-wave approximation to ψ (eq. (3.7)) by

$$\kappa\mathcal{S}_\mu(\mathcal{R}) = B, \quad (\text{A1.2})$$

where B is constant; geometrical wavefronts are contour surfaces of action, i.e. of solutions of the Hamilton–Jacobi equation (2.4). We shall give a brief description of these two sorts of wavefronts, emphasizing their differences.

Diffraction wavefronts, defined by given values of A in eq. (A1.1), are surfaces in space (or lines in the plane), intersecting neither themselves nor wavefronts with different values of A . They have singularities where the function $\chi(\mathcal{R})$ is singular. Since ψ is a smooth function, χ may be singular only where the amplitude a in eq. (3.2) is zero. Because two equations must be satisfied ($\text{Re } \psi = \text{Im } \psi = 0$), singularities of diffraction wavefronts generically take the form of lines in space, called *wavefront dislocations* by analogy with similar structures in crystals (NYE and BERRY [1974]). On a dislocation line, infinitely many diffraction wavefronts join. During a circuit around a dislocation line, the phase $\kappa\chi$ changes by $2n\pi$, where n is an integer defining the dislocation strength. As explained in detail by NYE and BERRY [1974], wavefront dislocations may be of edge or screw type, they may be curved and, when ψ is time-dependent, they may interact in a variety of ways. Near a dislocation line, the vector field

∇_X has a vortex structure, a feature emphasized (in the context of quantum-mechanical waves) by HIRSCHFELDER, CHRISTOPH and PALKE [1974], HIRSCHFELDER, GOEBEL and BRUCH [1974], and HIRSCHFELDER and TANG [1976a,b] (see also DIRAC [1931] and RIESS [1970a,b, 1976]).

As explained above, wavefront dislocations are singularities derived from the complex function ψ . An alternative approach, in which dislocations are studied in terms of the topography of the function $\text{Re } \psi(\mathcal{R})$ whose stable features are analyzed using catastrophe theory, has been explored by WRIGHT [1979].

Considered as wave morphologies (BERRY [1979a]), wavefront dislocations are complementary to caustics, for the following reasons: at a dislocation the wave amplitude is zero, while on a caustic the geometrical-optics amplitude (§ 3.1) is infinite; observation of a dislocation requires discrimination of wave structure on the scale of λ , a circumstance in which caustics are blurred by diffraction (§ 3.3), whereas observation of a caustic takes place on scales large compared with λ , a circumstance in which phase details, and hence dislocations, are hard to discern. This complementarity is exemplified by the diffraction catastrophes described in § 3.3, all having caustics as their grossest features and (except the fold) dislocations as their finest details. However, dislocations are by no means associated only with caustics, because they occur generically in cases outside the scope of catastrophe optics where λ is not small in comparison with other relevant scales (see, for example, BORN and WOLF [1975] p. 576). In general, caustics are the singularities of ray theory, whereas dislocations are the singularities of wave theory.

We now turn to the geometrical wavefronts, defined by given values of B in eq. (A1.2). These are surfaces in space, with multiple branches labelled by the index μ , which corresponds to the different rays passing through each point \mathcal{R} (§ 2.1). Because of this multiplicity, several geometrical wavefronts can pass through \mathcal{R} , as shown in Fig. 2.1a where they intersect in pairs, and in Fig. 2.1b where they intersect in threes. Also illustrated in Fig. 2.1b is the fact that a geometrical wavefront may self-intersect.

The geometrical wavefronts are singular where two or more branches S_μ coincide; this occurs on the caustics of the associated ray family (§ 2.1). An individual wavefront in space generically has line singularities, but the family of wavefronts (for all B) has as its singularity the whole caustic surface. These wavefront singularities are called Legendre singularities, for mathematical reasons explained by ARNOL'D [1975] and

SEWELL [1977, 1978]. They form a hierarchy corresponding to the hierarchy of catastrophes. If the caustic is a cuspoid catastrophe of codimension K , the wavefronts have the form of sections through the singular set of a cuspoid catastrophe of codimension $K+1$. Thus, for example, if the caustic is a fold line in the plane, the wavefronts are cusped lines (Fig. 2.1a), while if the caustic is a cusp in the plane the wavefronts are sections of a swallowtail (Fig. 2.1b and WRIGHT [1979]). A general proof of this result for cuspoids is given by SEWELL [1977]. But if the caustic is an umbilic catastrophe, the wavefronts are not sections of higher catastrophes, and their forms have to be computed ad hoc (SEWELL [1978]). Pictures of sequences of wavefronts for all stable caustics that can occur in space are shown by ARNOLD [1976] and POSTON and STEWART [1978] p. 280.

It is evident that geometrical and diffraction wavefronts, and their singularities, are very different in character. In monochromatic waves, the geometrical wavefronts are of little physical significance, whereas the diffraction wavefronts are related to observable features of the wave function. For shock waves, however, the geometrical wavefronts are important because they correspond to propagating discontinuities of the field; for a brief account and references see POSTON and STEWART [1978]. For the intermediate case of quasi-monochromatic pulses, the gross structure is of a geometrical wavefront but there are dislocations on fine scales (WRIGHT [1977], HUMPHREY, NYE and WRIGHT, in preparation).

Appendix 2: Primary Aberrations

Consider the caustic in the image space of an optical system with rotation symmetry, for a point source of monochromatic light. This caustic can be classified both in terms of catastrophe theory and in terms of the traditional aberrations. Our purpose here is to describe the relationship between these two classifications for the simplest case: primary geometrical (Seidel) aberrations. Most common is the so-called third order theory, in which $\sin \theta$ is approximated by $\theta - \theta^3/3!$ in the ray formulae, but we will use the equivalent procedure of calculating the optical path length to fourth order in the paraxial geometrical optics of § 4. Our analysis is based on unpublished notes by Dr. J. H. Hannay.

The wavefront W emerging from a general optical system with rotation symmetry about the Z axis will be specified, as in § 4, by its deviation $f(\mathbf{r})$

from the plane $Z=0$. f depends parametrically on the source point ρ (Fig. A2.1), the five Seidel aberration coefficients B (spherical aberration), C (astigmatism), D (curvature), E (distortion) and F (coma), and on coefficients G and H describing the shape of the aberration-free wavefront. Rotation symmetry requires that f be a function only of invariants under co-rotation of r and ρ , and this gives, on choosing signs and numerical factors to agree with the usual development of aberration theory (BORN and WOLF [1975] chap. 5),

$$f(r) = \frac{1}{4}Br^4 + C(\rho \cdot r)^2 + \frac{1}{2}D\rho^2r^2 - E\rho^2(\rho \cdot r) - F(\rho \cdot r)r^2 - G\rho \cdot r - Hr^2, \quad (\text{A2.1})$$

where r denotes $|\mathbf{r}|$ and ρ denotes $|\rho|$. Taking coordinates $\rho = (\xi, \eta)$ and $\mathbf{r} = (x, y)$, and choosing $\xi = 0$ without loss of generality, we obtain

$$f(x, y) = \frac{1}{4}B(x^4 + 2x^2y^2 + y^4) + C\eta^2y^2 + \frac{1}{2}D\eta^2(x^2 + y^2) - E\eta^3y - F\eta(x^2y + y^3) - G\eta y - H(x^2 + y^2). \quad (\text{A2.2})$$

The most important features of W are its umbilic points (§ 4.3), which occur where the conditions (4.7) are satisfied. A short calculation shows that there are two umbilics whose relative positions depend on the

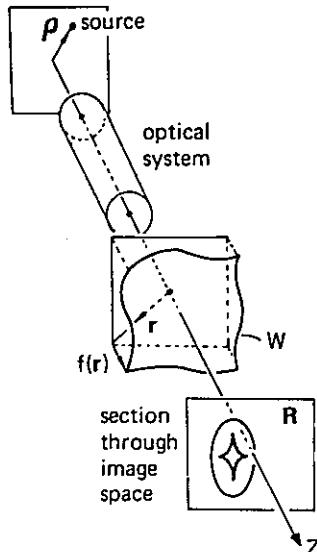


Fig. A2.1. Rotationally symmetric optical system; a source at ρ produces a wavefront W , giving rise to a caustic in the image space (R, Z).

quantity

$$Q = \frac{\eta^2}{B} \left(\frac{F^2}{B} - C \right), \quad (\text{A2.3})$$

and whose coordinates are

$$\begin{aligned} x = 0, \quad y = F\eta/B \pm \sqrt{Q} & \quad \text{for } Q > 0 \\ x = \pm \sqrt{|Q|}, \quad y = F\eta/B & \quad \text{for } Q < 0 \end{aligned} \quad (\text{A2.4})$$

Substitution of third derivatives of f , evaluated at these points, into the catastrophe classification (4.9) shows that the umbilics are always hyperbolic.

The wavefront (A2.2) has mirror symmetry about the lines $x = 0$ and $y = F\eta/B$, which are thus necessarily *riblines* (§ 4.4): contours of curvature must cross mirror lines normally – or be straight along them, which is clearly not possible in this case – and are therefore tangent to that principal axis of curvature which is normal to the mirror line, the other principal axis necessarily being along the mirror line. Hence it is not difficult to establish that the structurally stable form of the caustic is the sequence of sections $Z = \text{constant}$ shown in Fig. A2.2; the direction of increasing Z depends on the sign of B in (A2.1). The only catastrophes of codimension three are the two hyperbolic umbilic foci: there are no elliptic umbilic foci and no swallowtails. Note that in practice not all of the caustic need be seen in the image space; part of it may be virtual, and part of it may be missing due to aperture effects.

Clearly the curvature and distortion coefficients D and E do not affect the form of the caustic but merely its spatial position, and henceforth we ignore them. If the other aberration coefficients B , C and F are all zero,

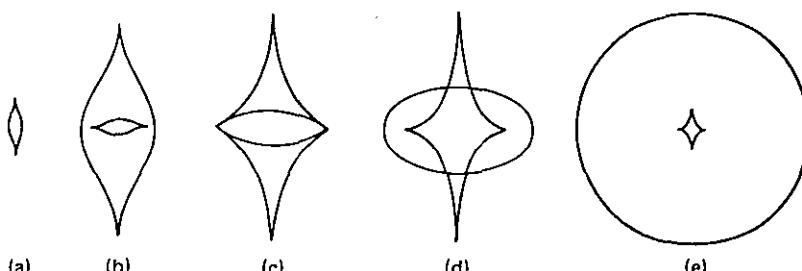


Fig. A2.2. Structurally stable caustic of primary aberration; sequence of sections $Z = \text{constant}$ showing (a) lips event, (b) conjugate lips event, (c) two hyperbolic umbilic foci, (d) and (e) unfolding into conical outer caustic and four-cusped “needle”.

the caustic is a perfect point focus, which is structurally unstable. If the astigmatism coefficient C alone is not zero, the caustic consists of two straight orthogonal focal lines in different planes, which again is structurally unstable. In the generic case where B and F are non-zero, these focal lines become the ribs of Fig. A2.2, and the two umbilic foci come into the field from infinity.

The behaviour of the umbilics is determined principally by Q (eq. (A2.3)). When $Q = 0$, we have “especially good” imaging; both umbilics and both lips events are coincident, the inner four-cusped needle in Fig. A2.2 collapses to a degenerate line caustic and the outer caustic surface is a rotationally symmetric cusped cone (Fig. A2.3). This caustic is not structurally stable, and corresponds to a catastrophe of infinite codimension. It occurs not only for $C = F = 0$, in which case the only contribution to the caustic is from the term with spherical aberration coefficient B , but also for $F^2 = BC$, which is the transition between the two possible orientations of the caustic of Fig. A2.2.

If $B = 0$, one umbilic point is at the origin and the other at infinity in the r plane. The generating function ϕ of eq. (4.2) is simply the hyperbolic umbilic, and may be put into the normal form Φ of Table 1 by the transformation $x \rightarrow x - y$, $y \rightarrow (x + y)/\sqrt{3}$. This is the case of primary coma, whose coefficient F multiplies the germ of the hyperbolic umbilic. Note that C need not be zero – the degree of astigmatism simultaneously present does not change the topology of the caustic in this case. The existence of hyperbolic umbilic foci does not, however, depend on there being a nonvanishing coma coefficient; if F alone is zero, then Q is nonzero and the effect of astigmatism ($C \neq 0$) in addition to spherical aberration ($B \neq 0$) is to break the degeneracy of the caustic in Fig. A2.3 to give that of Fig. A2.2.

The geometrical caustics will of course be clothed with diffraction, as described for the stable caustics of Fig. A2.2 by the diffraction catastrophes of § 3.3 and for the unstable caustic of Fig. A2.3 in chapter 9 of BORN and WOLF [1975].

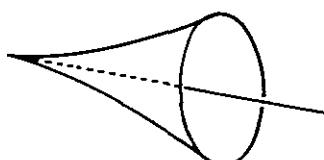


Fig. A2.3. Structurally unstable caustic of primary aberration when the two umbilics are coincident.

References

- ABRAMOWITZ, M. and I. A. STEGUN, eds., 1964, *Handbook of Mathematical Functions* (Dover, New York).
- AHLUWALIA, D. S., R. M. LEWIS and J. BOERSMA, 1968, SIAM J. Appl. Math. **16**, 783–807.
- AIRY, G. B., 1838, Trans. Camb. Phil. Soc. **6**, 379–403.
- ALBEVERIO, S. and R. HÖEGH-KROHN, 1977, Inventiones Math. **40**, 59–106.
- ARNOL'D, V. I., 1975, Russ. Math. Survs. **30**, 1–75.
- ARNOL'D, V. I., 1976, Commun. Pure Appl. Math. **29**, 557–582 [Errata **30**, 1977, 823].
- BARTZ, G., G. WEISSENBERG and D. WISKOTT, 1956, Radex Rdsch., pp. 163–165.
- BERRY, M. V., 1966, Proc. Phys. Soc. **89**, 479–490.
- BERRY, M. V., 1969a, Science Progress (Oxford) **57**, 43–64.
- BERRY, M. V., 1969b, J. Phys. **B2**, 381–392.
- BERRY, M. V., 1975, J. Phys. **A8**, 566–584.
- BERRY, M. V., 1976, Adv. Phys. **25**, 1–26.
- BERRY, M. V., 1977, J. Phys. **A10**, 2061–2081.
- BERRY, M. V., 1978a, La Recherche **9**, 760–768.
- BERRY, M. V., 1978b, Regular and Irregular Motion, in: Topics in non-linear dynamics, ed. S. Jorna, Am. Inst. Phys. Conf. Proc. **46**, 16–120.
- BERRY, M. V., 1979a, Some geometric aspects of wave motion, in: Geometry of the Laplace operator, Proc. Amer. Math. Soc. Conf., Honolulu 1979, to be published.
- BERRY, M. V., 1979b, J. Phys. **A12**, 781–797.
- BERRY, M. V. and J. H. HANNAY, 1977, J. Phys. **A10**, 1809–1821.
- BERRY, M. V. and K. E. MOUNT, 1972, Repts. Prog. Phys. **35**, 315–397.
- BERRY, M. V. and J. F. NYE, 1977, Nature **267**, 34–36.
- BERRY, M. V., J. F. NYE and F. J. WRIGHT, 1979, Phil. Trans. R. Soc. Lond. **A291**, 453–484.
- BORN, M. and E. WOLF, 1975, *Principles of Optics*, 5th ed. (Pergamon Press, Oxford).
- BRYANT, H. C. and N. JARMIE, 1974, Sci. Am. **231**, 60–71.
- BUCKLEY, R., 1971a, Aust. J. Phys. **24**, 351–371.
- BUCKLEY, R., 1971b, Aust. J. Phys. **24**, 373–396.
- BUDDEN, K. G., 1976, Proc. R. Soc. Lond. **A350**, 143–164.
- CALLAHAN, J., 1977, Geometry of E_6 and anorexia, Preprint, Mathematics Institute, University of Warwick, U.K.
- CALLAHAN, J., 1978, Special bifurcations of the double cusp, Preprint, Mathematics Institute, University of Warwick, U.K.
- CHESTER, C., B. FRIEDMAN and F. URSELL, 1957, Proc. Camb. Phil. Soc. **53**, 599–611.
- CONNOR, J. N. L., 1976, Molec. Phys. **31**, 33–55.
- DANGELMAYR, G. and W. VEIT, 1979, Ann. Phys. New York **118**, 108–138.
- DARBOUX, G., 1896, *Leçons sur la théorie générale des surfaces*, Vol. 4 (Gauthier-Villars, Paris) Note VII [Reprinted: 1972 (Chelsea, New York)].
- DESCARTES, R., 1637, *Les Météores*, Discours VIII [also in: 1902, *Oeuvres de Descartes*, Vol. VI, eds. C. Adam and P. Tannery (Léopold Cerf, Paris)].
- DIRAC, P. A. M., 1931, Proc. R. Soc. Lond. **A133**, 60–72.
- DUISTERMAAT, J. J., 1974, Commun. Pure Appl. Math. **27**, 207–281.
- FEYNMAN, R. P., 1948, Rev. Mod. Phys. **20**, 367–387.
- FEYNMAN, R. P. and A. R. Hibbs, 1965, *Quantum mechanics and path integrals* (McGraw-Hill, New York).
- FORSYTH, A. R., 1912, *Differential geometry* (Cambridge University Press, Cambridge) chap. 4.

- GLASER, W., 1956, in: *Handbuch der Physik* **33**, ed. S. Flügge (Springer, Berlin) pp. 273-284.
- GODWIN, A. N., 1971, *Publ. Math. IHES* **40**, 117-138.
- GOLUBITSKY, M., 1978, *SIAM Review* **20**, 352-387.
- GREEN, H. S. and E. WOLF, 1953, *Proc. Phys. Soc.* **A66**, 1129-1137.
- GUILLEMIN, V. and S. STERNBERG, 1977, *Geometric asymptotics*, American Mathematical Society Survey No. **14** (Providence, USA).
- GULLSTRAND, A., 1905, *Acta. Math.* **29**, 59-100.
- HANNAY, J. H., 1977, Paraxial optics and statistical problems of wave propagation, Ph.D. Thesis, Magdalene College, University of Cambridge, U.K.
- HIRSCHFELDER, J. O., A. C. CHRISTOPH and W. E. PALKE, 1974, *J. Chem. Phys.* **61**, 5435-5455.
- HIRSCHFELDER, J. O., C. J. GOEBEL and L. W. BRUCH, 1974, *J. Chem. Phys.* **61**, 5456-5459.
- HIRSCHFELDER, J. O. and K. T. TANG, 1976a, *J. Chem. Phys.* **64**, 760-785.
- HIRSCHFELDER, J. O. and K. T. TANG, 1976b, *J. Chem. Phys.* **65**, 470-486.
- HOLFORD, R. L., 1972, Modifications to ray theory near cusped caustics, Bell Telephone Labs. Report.
- HUMPHREYS, W. J., 1940, *Physics of the air*, 3rd ed. (McGraw-Hill, New York).
- JAKEMAN, E. and J. G. McWHIRTER, 1977, *J. Phys.* **A10**, 1599-1643.
- JAKEMAN, E., E. R. PIKE and P. N. PUSEY, 1976, *Nature* **263**, 215-217.
- KELLER, J. B., 1958, *Proc. Symp. Appl. Math.* **8**, 27-52.
- KHARE, V. and H. M. NUSSENZVEIG, 1974, *Phys. Rev. Lett.* **33**, 976-980.
- KHARE, V. and H. M. NUSSENZVEIG, 1977, *Phys. Rev. Lett.* **38**, 1279-1282.
- KLINE, M. and I. W. KAY, 1965, *Electromagnetic theory and geometrical optics* (Interscience, New York).
- KRAVTSOV, Yu. A., 1964a, *Radiofizika* **7**, 664-673.
- KRAVTSOV, Yu. A., 1964b, *Radiofizika* **7**, 1049-1056.
- KRAVTSOV, Yu. A., 1968, *Soviet Phys. Acoust.* **14**, 1-17.
- LANCZOS, C., 1966, *The variational principles of mechanics*, 3rd ed. (University of Toronto Press, Toronto).
- LARMOR, J., 1891, *Proc. Camb. Phil. Soc.* **7**, 131-137.
- LENZ, F., 1956, in: *Proc. 3rd Int. Conf. on electron microscopy*, London 1954 (Royal Microscopical Society, London) pp. 83-85.
- LENZ, F. and H. MECKING, 1961, in: *Proc. Eur. Reg. Conf. on electron microscopy*, Delft 1960, Vol. I, eds. A. L. Houwink and B. J. Spit (Nederlandse Vereniging voor Elektronenmicroscopie, Delft) pp. 41-43.
- LEVIT, S. and U. SMILANSKY, 1977, *Ann. Phys. New York* **108**, 165-197.
- LIESEGANG, S., 1953, *Optik* **10**, 5-14.
- LONGUET-HIGGINS, M. S., 1956, *Proc. Camb. Phil. Soc.* **52**, 234-245.
- LONGUET-HIGGINS, M. S., 1957, *Phil. Trans. R. Soc. Lond.* **A249**, 321-387.
- LONGUET-HIGGINS, M. S., 1958, *Proc. Camb. Phil. Soc.* **54**, 439-453.
- LONGUET-HIGGINS, M. S., 1959, *Proc. Camb. Phil. Soc.* **55**, 91-100.
- LONGUET-HIGGINS, M. S., 1960a, *J. Opt. Soc. Am.* **50**, 838-856.
- LONGUET-HIGGINS, M. S., 1960b, *Proc. Camb. Phil. Soc.* **56**, 27-40.
- LUDWIG, D., 1966, *Commun. Pure Appl. Math.* **19**, 215-250.
- MANDELBROT, B. B., 1977, *Fractals* (W. H. Freeman, San Francisco).
- MARCUSE, D., 1972, *Light transmission optics* (Van Nostrand, New York).
- MASLIN, N. M., 1976a, *J. Atmos. Terr. Phys.* **38**, 239-250.
- MASLIN, N. M., 1976b, *J. Atmos. Terr. Phys.* **38**, 1251-1263.

- MERCIER, R. P., 1962, Proc. Camb. Phil. Soc. **58**, 382–400.
- MINNAERT, M., 1954, The nature of light and colour in the open air (Dover, New York).
- MASLOV, V. P., 1965 (in Russian) [French translation: 1972. Théorie des perturbations et des méthodes asymptotiques (Dunod, Paris)].
- MCDONALD, J. E., 1963, Am. J. Phys. **31**, 282–284.
- NUSSENZVEIG, H. M., 1969a, J. Math. Phys. **10**, 82–124.
- NUSSENZVEIG, H. M., 1969b, J. Math. Phys. **10**, 125–176.
- NUSSENZVEIG, H. M., 1977, Sci. Am. **236**, 116–127.
- NYE, J. F., 1978, Proc. R. Soc. Lond. **A361**, 21–41.
- NYE, J. F., 1979, Phil. Trans. R. Soc. Lond. **A292**, 25–44.
- NYE, J. F. and M. V. BERRY, 1974, Proc. R. Soc. Lond. **A336**, 165–190.
- PEARCEY, T., 1946, Phil. Mag. **37**, 311–317.
- PORTEOUS, I. R., 1971, J. Diff. Geom. **5**, 543–564.
- POSTON, T., 1980, Catastrophe theory in physics, Physics Reports, to be published.
- POSTON, T. and I. N. STEWART, 1978, Catastrophe theory and its applications (Pitman, London).
- RICE, S. O., 1944, Bell Syst. Tech. J. **23**, 282–332.
- RICE, S. O., 1945, Bell Syst. Tech. J. **24**, 46–156.
- RIESS, J., 1970a, Ann. Phys. New York **57**, 301–321.
- RIESS, J., 1970b, Phys. Rev. **D2**, 647–653.
- RIESS, J., 1976, Phys. Rev. **B13**, 3862–3869.
- ROMERO, M. del C., 1978, On the Berry index, M.Sc. Thesis, Mathematics Institute, University of Warwick, U.K.
- SALPETER, E. E., 1967, Astrophys. J. **147**, 433–448.
- SCHEFFELS, W., M. HAHN and F. LENZ, 1953, Optik **10**, 455–458.
- SCHLEICH, F., D. HOFFMEISTER, H. KOOPS and F. LENZ, 1968, Optik **27**, 305–316.
- SCHULMAN, L. S., 1975, Caustics and multivaluedness, in: Functional integration and its applications, ed. A. M. Arthurs (Clarendon Press, Oxford) pp. 144–156.
- SEWELL, M. J., 1977, Math. Proc. Camb. Phil. Soc. **82**, 147–163.
- SEWELL, M. J., 1978, Math. Proc. Camb. Phil. Soc. **83**, 273–288.
- SHISHOV, V. I., 1971, Izv. Vuz. Radiofiz. **14**, 85–92.
- SPIVAK, M., 1975, A comprehensive introduction to differential geometry, Vol. 3 (Publish or Perish, Boston) Chaps. 4 and 6.
- STEVENS, P. S., 1976, Patterns in nature (Penguin, Harmondsworth, U.K.).
- TANNER, L. H., 1978, Opt. Laser Technol. **10**, 125–128.
- THOM, R., 1972, Stabilité structurelle et morphogénèse (Benjamin, Reading, Mass.) [English translation, updated: 1975, Structural stability and morphogenesis (Benjamin, Reading, Mass.)].
- TRICKER, R. A. R., 1970, Introduction to meteorological optics (Mills and Boon, London).
- TRINKAUS, H. and F. DREPPER, 1977, J. Phys. **A10**, L11–L16.
- UPSTILL, C., 1979, Proc. R. Soc. Lond. **A365**, 95–104.
- USCINSKI, B. J., 1977, The elements of wave propagation in random media (McGraw-Hill, New York).
- VARCHENKO, A. N., 1976, Funct. Anal. Appl. **10**, 13–38.
- WOLF, E., 1959, Proc. Phys. Soc. **74**, 269–280.
- WOLF, E., 1960, Some aspects of a rigorous scalar treatment of electromagnetic wave propagation, in: Electromagnetic wave propagation (Academic Press, New York) pp. 119–125.
- WRIGHT, F. J., 1977, Wavefield singularities, Ph.D. Thesis, H. H. Wills Physics Laboratory, University of Bristol, U.K.

- WRIGHT, F. J., 1979, Wavefront dislocations and their analysis using catastrophe theory, in: Structural stability in physics. eds. W. Göttinger and H. Eikemeier (Springer, Berlin) pp. 141-156.
- WRIGHT, F. J., 1980, J. Phys. A. to be published.
- ZAHLER, R. S. and H. J. SUSSMAN, 1977, Nature **269**, 759-763 [see also correspondence, **270**, 382-384 and 658].
- ZEEMAN, E. C., 1977, Catastrophe theory: selected papers 1972-1977 (Addison-Wesley, Reading, Mass.).

the “lollipop” (Fig. 6.8), the elliptic umbilic shrinks to its singular point and unfolds again, one or more of its cusps pierce the “foliage”, and the unfolded elliptic umbilic and folds interact via beak-to-beak events and swallowtails, leaving three line pairs ending in inward pointing cusps which recede from the field of view.

These triple junctions link up to form a network resembling chain mail. The existence of such networks can be confirmed experimentally by producing caustics of refraction from a type of bathroom window glass (Pilkington’s “Atlantic”), as shown in Fig. 6.9. However, it is impossible for all the umbilic catastrophes produced by an undulating wavefront to be of elliptic type; this follows from the arguments outlined in § 4.3, according to which all elliptic umbilics have index $-\frac{1}{2}$, whereas the surface has net index zero. Therefore there must be hyperbolic umbilic points, of index $+\frac{1}{2}$, on the surface: where do they focus? Recall that the caustics have continuity through the far field (§ 5.2). This implies that, provided there are no hyperbolic umbilics in the real field, the missing ones must lie in the virtual field. This possibility is exemplified by “Atlantic” glass,

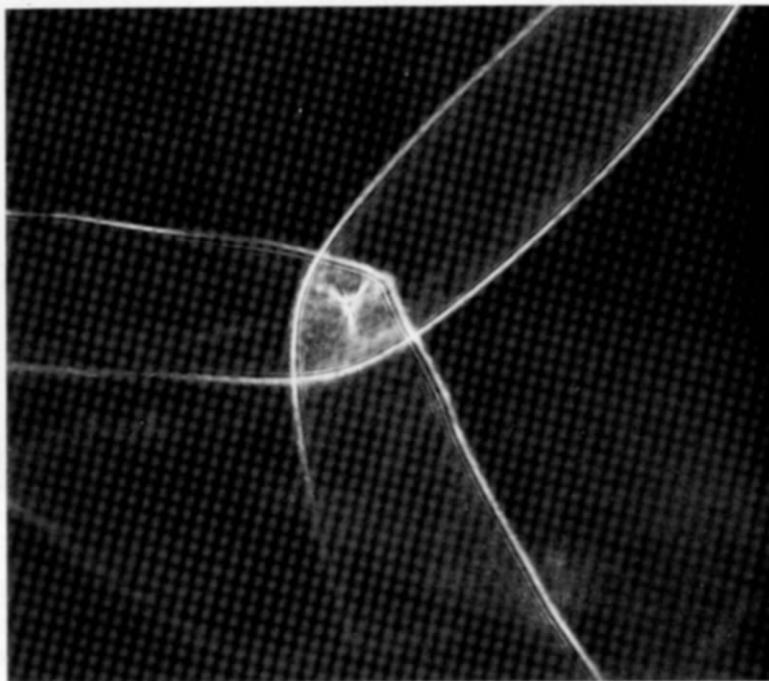


Fig. 6.7. An elliptic umbilic centred caustic triple junction resolved (courtesy of J. F. Nye).