Mon. Not. R. astr. Soc. (1987) **229,** 73–102

The effects of caustics on scintillating radio sources

Jeremy J. Goodman *Institute for Advanced Study, Princeton, NJ 08540, USA*

Roger W. Romani and Roger D. Blandford *Theoretical Astrophysics, California Institute of Technology, Pasadena, CA 91124, USA*

Ramesh Narayan *Steward Observatory, University of Arizona, Tucson, AZ 85721, USA*

Accepted 1987 May 28. Received 1987 March 24

Summary. We consider the scintillation properties of compact radio sources for a spectrum of interstellar electron density fluctuations which is a power law over a finite range of spatial frequencies. In particular, if the power law is truncated at an inner scale intermediate between the diffractive scale and the refractive scale, we find that there is additional power in the spectrum of the intensity variations at these scales. This power is associated with strong focusing events, or *caustics*. These events are best described as simple diffraction catastrophes, which are classified and analysed on the basis of geometrical and wave optics, taking account of the strong dispersion of the ISM scattering, which introduces important frequency dependencies. It is shown that such caustics can enhance the flux modulation and may produce patterns in dynamic spectra similar to those observed. Caustics are best observed by monitoring pulsar fluxes through a range of frequencies and specific observations are suggested which should either detect their effects or delimit their importance. We conclude that caustic effects are only likely to dominate the flux variations when the inner scale is ~10 per cent of the refractive scale. For somewhat larger inner scales individual caustics could be resolved. Whether such conditions obtain in the ISM is still uncertain; if they do not, a power-spectrum description may not be adequate and focusing events from single scatterers (i.e. rare high-density clouds instead of an incoherent superposition of small-amplitude density fluctuations along the line-of-sight) may dominate the refractive variations.

1 Introduction

While the possibility of a slow component of the flux scintillation of radio pulsars by the interstellar medium (ISM) has been discussed for some time (e.g. Salpeter 1967; Lovelace 1970; Shishov 1974; Shapirovskaya 1978; Sieber 1982), more recent work (Rickett, Coles & Bourgois

1984; Blandford & Narayan 1985, hereafter BN) has emphasized the importance of such refractive (i.e. geometrical optics) fluctuations and shown that they may account for the low-frequency variability of pulsars and extragalactic radio sources. In these studies the spectrum of electron density fluctuations in the ISM giving rise to the scintillations has been most frequently characterized as an extended power law in spatial wavenumber, $Q(q) \sim q^{-\beta}$ where β is close to the 'Kolmogorov' value, $\beta = 11/3$ Rickett (1977). It has been shown (BN; Romani, Narayan & Blandford 1986, hereafter RNB) that refractive scintillation should cause a number of observables other than the flux to fluctuate, as well, and that here should be correlations among these fluctuations. Comparison of the variations computed for a $\beta = 11/3$ spectrum with the observations has, however, met with several difficulties: (i) the observed fluctuations (i.e. modulation indices) exceed those computed for the assumed spectrum of density perturbations, (ii) strongly non-Gaussian 'spikes', inconsistent with the simple theory, are sometimes seen in long-term studies of pulsar flux (e.g. Cole, Hesse & Page 1970; Helfand, Fowler & Kuhlman 1977), (iii) quasi-periodic bands are sometimes observed in the 'dynamic scintillation spectra' (i.e. contour plots of flux density in the frequency–time plane) of many pulsars (e.g. Roberts & Ables 1982; Cordes, Pidwerbetsky & Lovelace 1987), again not explained by the standard theory.

These three observations argue for increased power in the low-frequency refractive fluctuations relative to the normal 'diffractive' scintillation. There are three ways in which the spectrum of electron density perturbations could be modified to achieve this: (i) the power law index of the density perturbations could be increased to $\beta \gtrsim 4$, (ii) the power spectrum could be truncated at an inner scale, (iii) an extra component may be added to the perturbation spectrum at low frequencies. The first approach has been suggested by Roberts & Ables (1982) and Hewish, Wolszczan & Graham (1985) based on observations of dynamic scintillation spectra. BN, Goodman & Narayan (1985; henceforth GN) and RNB have calculated fluctuations based on $\beta > 4$ spectra, finding that the modulation indices etc. do increase although some difficulties remain in interpreting the observations. Truncation of the power spectrum has been considered by Coles *et al.* (1987) and is the subject of further investigation in this work. We find that, in addition to the enhancement of the refractive low-frequency fluctuations discussed by these authors, there should be an additional contribution to the flux modulation from intermediate scales in the form of strong focusing events or *caustics*. The third approach, superposition of extra fluctuation power at a preferred scale has been advocated by Cordes & Wolszczan (1986). Although this third possibility generates clear predictions in terms of preferred time-scales for variability and organization of dynamic spectra, we do not discuss it in detail. However, we note that this extra fluctuation power may be characterized by a random modulation at large scales or, as a more radical alternative, the distributed line-of-sight scattering may be dominated by single large fluctuations (Romani, Blandford & Cordes 1987). In either case, strong focusing can arise from scattering on these large scales and many of our qualitative results are applicable to the resulting modulations.

In Section 2 we compute the autocorrelation function of flux variations for a power law spectrum of phase fluctuations with an inner scale and show that there is excess power on intermediate scales. We identify conditions under which this contribution to the flux variations is dominant. In Section 3, we review the general appearance of simple diffraction catastrophes (i.e. caustics) under conditions appropriate to interstellar scintillation and show that they are the cause of the new component in the intensity spectrum in the presence of an inner scale. In Section 4 we apply our results to the case of radio scintillation in the interstellar medium. We describe the variation of the diffraction catastrophes as a function of observation frequency and discuss strategies for the detection of individual caustics. In Section 5 the predictions of the caustic picture for the appearance of dynamic scintillation spectra are described and compared with other scenarios. We conclude in Section 6 by suggesting specific observational tests which could help us to gauge the importance of caustic events.

2 Flux variations due to power law phase fluctuations with an inner scale

In this section we present a formal treatment of the power spectrum of flux variations arising from a power law spectrum of phase fluctuations with an inner scale cut-off. We find that the cut-off introduces extra power at scales between those normally associated with diffraction and refraction. We show in Section 3.3 that this power is due to the effect of caustics. For readers wishing to avoid technical complications, the principal results are displayed in Figs 2 and 3.

2.1 THE THIN SCREEN

To simplify the analysis, we adopt the thin-screen model throughout this paper, i.e. we replace the electron-density fluctuations of the interstellar medium by a thin scattering layer ('screen') at distances z_{os} and z_{sp} respectively from the observer's plane and from the pulsar (Fig. 1). We use the notations $\mathbf{R} = (X, Y)$ and $\mathbf{r} = (x, y)$ for positions in the observer's plane and in the thin screen,

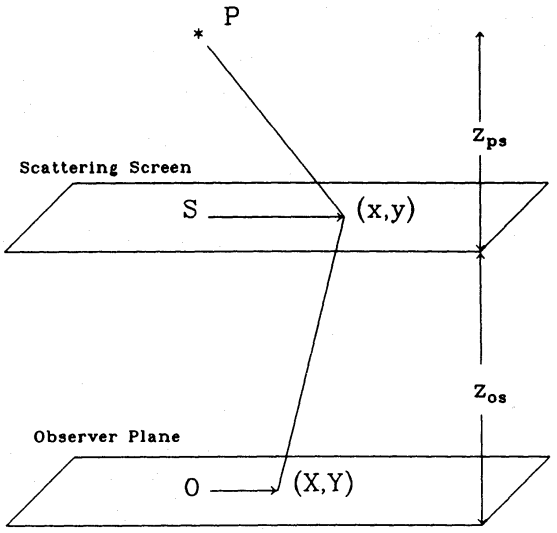

Figure 1. Scattering geometry. The scattering medium is idealized as a thin screen at a distance z_{os}, with the source at distance z_{ps} beyond the screen. The transverse coordinates on the observer plane and scattering screen are denoted (X, Y) and (x, y), respectively, with the coordinate origin taken to be the screen intersection with the optical axis *PSO*.

respectively, and take the origins for these two coordinate systems at the intersections with the optical axis *OSP*. To make matters even simpler, we replace the spherical wavefront emanating from the pulsar and impinging on the far side of the screen by a plane wave of the same flux, thus in effect letting $z_{sp} \to \infty$. In order to preserve the pattern of fluxes on the observer's plane, we change the distance between the screen and the observer to

$$z' \equiv \left(\frac{1}{z_{os}} + \frac{1}{z_{sp}}\right)^{-1}, \tag{2.1.1}$$

and use the rescaled coordinates

$$\mathbf{R}' \equiv \left(\frac{z_{sp}}{z_{os} + z_{sp}}\right) \mathbf{R} \tag{2.1.2}$$

in the observer's plane. These scalings ensure that the variation of path length with $\mathbf{r}$, $\mathbf{R}$ is preserved. Henceforth we drop the primes. We further assume that $r \ll z$ and $R \ll z$ (the 'paraxial approximation'), which is reasonable since the scattering angles in our problem are measured in mas, at most.

The advantage of the thin-screen model is that it permits us to describe a particular ray path between the pulsar and the observer by **r** and **R** alone (for a given z), whereas with an extended scattering medium an infinite number of parameters would be required. Predictions of the thin-, screen and extended-medium models have been compared by Tatarskii & Zavorotnyi (1980). In general the agreement between the two models is very good (but see RNB; Frehlich 1986). In particular, Coles *et al.* have evaluated an approximate computation of the refractive fluctuations due to an extended scattering medium with an inner-scale cut-off and find an effect comparable to the thin-screen case. It should be noted that, in so far as the extended medium is assumed to be statistically homogeneous, it is as much an idealization as the thin-screen model. Indeed, the thin-screen model is to be preferred when the scattering is dominated by a relatively small segment of the line-of-sight, e.g. an H II region.

2.2 THE DIFFRACTION INTEGRAL

If we ignore the very small depolarizing effects of the scattering, then the electric field on the observer's plane can be represented by a scalar $\psi(\mathbf{R})$ whose absolute square is proportional to the flux $F(\mathbf{R})$, i.e.

$$F(\mathbf{R}) = F_0 |\psi(\mathbf{R})|^2, \tag{2.2.1}$$

where $F_0 \propto (z_{os} + z_{sp})^{-2}$ is the average flux. If the flux is measured in a very narrow band centred on wavelength $\lambda \equiv 2\pi/k$, and the Fresnel scale r_F is defined by

$$r_F = \sqrt{z/k} \tag{2.2.2}$$

then, in the thin-screen model described above,

$$\psi(\mathbf{R}) = \int \frac{d^2\mathbf{r}}{2\pi r_F^2} \exp\,[i\zeta(\mathbf{r}, \mathbf{R})], \tag{2.2.3}$$

where

$$\zeta(\mathbf{r}, \mathbf{R}) = \phi(\mathbf{r}) + \frac{(\mathbf{r}^2 - 2\mathbf{r}\cdot\mathbf{R})}{2r_F^2} \tag{2.2.4}$$

is the total phase along the path that passes through **r** and **R**, apart from an irrelevant term that depends on **R** only. The quantity $\phi(r)$ is the excess phase at **r** introduced by the scattering material, while the second term is a geometrical contribution due to path-length variations.

2.3 PHASE STATISTICS

We are interested in evaluating certain moments of ψ. We regard the excess phase $\phi(\mathbf{r})$ to be a random function of **r**, and we assume that it has zero mean, which is allowed since the mean can be subtracted from ζ without affecting the flux. Because the relation between ψ and ϕ is non-linear, the problem is tractable only if we make the usual assumption that ϕ is a stationary Gaussian process with correlation function

$$C(\mathbf{r}) \equiv \langle \phi(\mathbf{r}' + \mathbf{r})\, \phi(\mathbf{r}') \rangle. \tag{2.3.1}$$

Since the electron column density must be positive, the Gaussian approximation requires at the very least that the fluctuations in column density be small compared to the mean. In relating our thin-screen results to the extended medium, $C(\mathbf{r})$ should be interpreted as the mean-square difference in integrated phase along two *parallel* lines-of-sight with linear separation **r** and common length z.

$C(\mathbf{r})$ is uniquely determined by its Fourier transform $k^{-2}Q(q)$, the power spectrum of ϕ:

$$Q(q) = k^2 \int d^2\mathbf{r} \exp(i\mathbf{q}\cdot\mathbf{r})\, C(\mathbf{r})$$

$$= k^2 \langle \tilde{\phi}(\mathbf{q})\, \tilde{\phi}^*(\mathbf{q}) \rangle / A, \tag{2.3.2}$$

where $\tilde{\phi}(\mathbf{q})$ is the Fourier transform of $\phi(x, y)$, calculated over an area A on the screen. The extra factor of k^2 in the definition makes Q dimensionless. For simplicity we assume that the scattering is isotropic, so that C and Q depend only on $r \equiv |\mathbf{r}|$ and $q \equiv |\mathbf{q}|$, respectively. We can interpret $q^2Q(q)/k^2$ as the mean-square amplitude of the phase fluctuations on the scale $r = 2\pi/q$.

We are particularly interested in spectra that are well approximated by power laws in some range:

$$Q(q) = Q_0 q^{-\beta}, \quad 2\pi/r_{\text{out}} \equiv q_{\text{out}} < q < q_c \equiv 2\pi/r_c. \tag{2.3.3}$$

We assume that Q is negligibly small outside the range, $q_{\text{out}} < q < q_c$. Thus, the largest scales in the medium are $\sim r_{\text{out}}$, the 'outer' scale, and the smallest scales are $\sim r_c$, the 'cut-off' or 'inner' scale.

The correlation function corresponding to (2.3.3) for $r_c \ll r \ll r_{\text{out}}$ and $2 < \beta < 6$ (e.g. GN) is

$$C(r) \approx \frac{Q_0}{2\pi k^2(\beta-2)} \left[\frac{2^{3-\beta}\Gamma[(6-\beta)/2]}{(\beta-4)\,\Gamma(\beta/2)}\, r^{\beta-2} + q_{\text{out}}^{2-\beta} \right] \qquad (\beta < 4)$$

$$\approx \frac{Q_0}{2\pi k^2(\beta-2)} \left[\frac{2^{3-\beta}\Gamma[(6-\beta)/2]}{(\beta-4)\,\Gamma(\beta/2)}\, r^{\beta-2} + q_{\text{out}}^{2-\beta} + \frac{q_{\text{out}}^{4-\beta} r^2}{2(\beta-4)} \right] \qquad (\beta > 4). \tag{2.3.4}$$

For $r \gg r_{\text{out}}$, $C(r)$ is negligibly small. On the other hand, for $r \leqslant r_c$, $C(r)$ can be expressed as a convergent series in r^2 with the term in $q_{\text{out}}^{2-\beta}$ as the leading term. We have omitted terms from (2.3.4) that vanish as $q_{\text{out}} \to 0$ and $q_c \to \infty$.

2.4 CORRELATION FUNCTION AND POWER SPECTRUM OF THE FLUX

The normalized correlation function of the flux,

$$W(\mathbf{R}) \equiv \langle F(0)\, F(\mathbf{r}) \rangle / \langle F \rangle^2, \tag{2.4.1}$$

contains information about the typical amplitude of flux variations and the time-scales on which they occur. Its two-dimensional Fourier transform can be obtained from equation (2.2.1)–(2.2.4) (Rumsey 1975):

$$\tilde{W}(\mathbf{q}) \equiv \int d^2\mathbf{R} \exp(i\mathbf{q}\cdot\mathbf{r}) W(\mathbf{R})$$

$$= \int d^2\mathbf{r} \exp(i\mathbf{q}\cdot\mathbf{r}) \exp\{\tfrac{1}{2}[D(\mathbf{r} + r_F^2\mathbf{q}) + D(\mathbf{r} - r_F^2\mathbf{q}) - 2D(\mathbf{r}) - 2D(r_F^2\mathbf{q})]\}, \tag{2.4.2}$$

where

$$D(\mathbf{r}) \equiv \langle [\phi(\mathbf{r}) - \phi(0)]^2 \rangle = 2[C(0) - C(\mathbf{r})] \tag{2.4.3}$$

is the structure function of ϕ. One can interpret $q^2\tilde{W}(q)$ as the mean-square amplitude of flux variations $\Delta F/F$ per octave on the length-scale $R = 2\pi/q$ in the observer's plane. If v is the relative velocity of the Earth and the pulsar transverse to the line-of-sight, then R corresponds to a time-scale $t = R/v$.

The second and third terms in equations (2.3.4) express the large contribution of scales near r_{out} to the variance of the phase and, when $\beta>4$, to the phase gradient as well. They *cancel* from (2.4.2) because only second and higher derivatives of ϕ are responsible for focusing and defocusing the rays. However, these terms can contribute importantly to timing noise (Lovelace 1970; Armstrong 1984; Blandford, Narayan & Romani 1984). For questions concerning the flux alone, the precise value of r_{out} is unimportant provided that it is larger than all relevant length scales in the screen: in particular, r_{out} should be larger than the refractive scale r_{ref}, defined by

$$r_{ref} = \theta_{rms} z \equiv 2\pi/q_{ref}, \tag{2.4.4}$$

where θ_{rms} instantaneous size of the scatter-broadened image. We shall assume henceforth that $r_{out} \gg r_{ref}$.

For the purposes of evaluating (2.4.2), it is convenient to make the dependence of D on r_c simple and explicit. We assume

$$D(\mathbf{r}) = -\frac{Q_0'}{2}\left[(r^2+r_c^2)^{(\beta-2)/2} - r_c^{\beta-2}\right] \qquad (\beta<4)$$

$$D(\mathbf{r}) = -\frac{Q_0'}{2}\left\{(r^2+r_c^2)^{(\beta-2)/2} - r_c^{\beta-2} + \frac{q_{out}^{4-\beta}\Gamma(\beta/2)}{2^{4-\beta}\Gamma[(6-\beta)/2]}\, r^2\right\} \qquad (\beta>4) \tag{2.4.5}$$

where Q_0' is the coefficient of $r^{\beta-2}$ in (2.3.4), and has the sign of $(\beta-4)$. This form of $D(\mathbf{r})$ has the right properties to describe the power spectrum (2.3.3), and corresponds to a particular choice for the functional form of the cut-off of $Q(q)$ outside the power law range.

GN have discussed the case $r_c = 0$ under the assumption that $|A| \gg 1$, where

$$A \equiv Q_0' r_F^{\beta-2}. \tag{2.4.6}$$

This is equivalent to assuming strong scintillation since $|A|$ is essentially the typical phase difference across the Fresnel length r_F. Their results for $q^2\tilde{W}(q)$ for the cases $(A, \beta) = (-10^3, 11/3)$ and $(10^3, 4.3)$ are shown by the solid lines in Figs 2 and 3, where lengths have been normalized to r_F (i.e. $\eta = qr_F$). We see that, when $r_c = 0$ and $|A| \gg 1$, almost all of the power in $\Delta F/F$ resides in two prominent peaks.

The high spatial frequency peak in Figs 2 and 3 reflects the power in *diffractive* scintillation. Rays from different parts of the scatter-broadened image interfere to form chaotic fringes on the observer's plane with characteristic linear separation $r_{dif} \sim 1/k\theta_{rms}$. Comparing with (2.4.4) and (2.2.2), we find

$$r_{dif} = r_F^2/r_{ref} \equiv 2\pi/q_{dif}. \tag{2.4.7}$$

The total power in diffractive scintillation is always ~ 1 for $|A| \gg 1$.

The low spatial frequency peak in Figs 2 and 3 is due to *refractive* effective. Phase fluctuations with wavelength $\sim r_{ref}$ on the screen focus or defocus rays toward the observer, and give rise to flux variations on the observer's plane at the same scale. GN show that the power in refractive fluctuations decreases with $|A|$ if $\beta<4$ but depends only on β if $4<\beta<6$. In general, refractive flux variations are more pronounced for $\beta>4$ than for $\beta<4$, as shown by Figs 2 and 3.

2.5 POWER IN FLUX VARIATIONS IN THE PRESENCE OF AN INNER SCALE

Using techniques similar to those described by GN, the spectrum of flux variations $\tilde{W}(q)$ can be calculated for the case $r_c \neq 0$. The details are given in Appendix A. The two dashed lines in Figs 2 and 3 show $q^2\tilde{W}(q)$ for a large inner scale, $r_c = 0.1r_{ref}$, and for a smaller cut-off, $0.01r_{ref}$. In Fig. 2, it was necessary to normalize Q_0 in proportion to $(r_c/r_{dif})^{4-\beta}$ in order to maintain the positions of

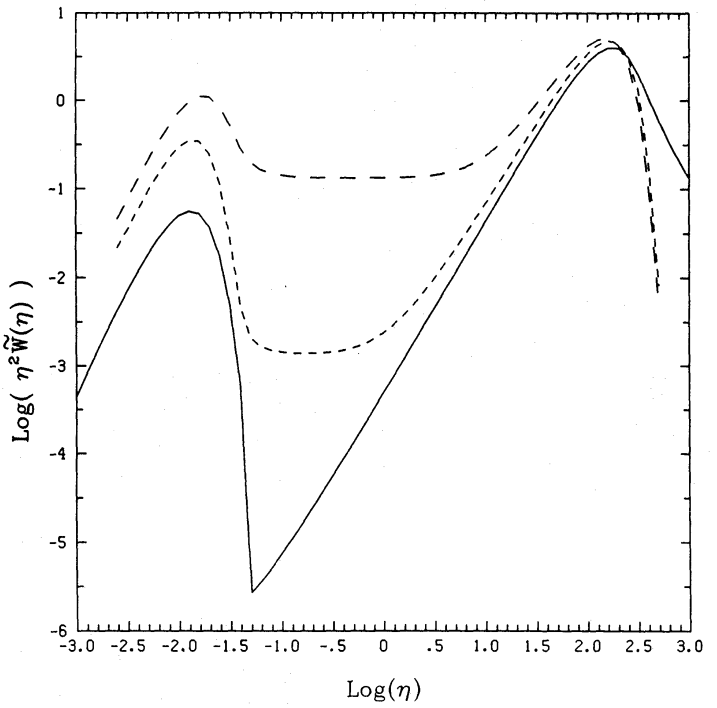

Figure 2. Flux variation power spectrum for $\beta=11/3$ and $r_c=0.0$ (solid line), $r_c=0.01r_{ref}$ (short dashes) and $r_c=0.1r_{ref}$ (long dashes). Spatial frequency is in units of the Fresnel scale $\eta=qr_F$. Note that, for $r_c\neq0$, to maintain the separation of η_{ref} and η_{dif}, the strength of the scattering $\propto|A|$ has been renormalized and this has enhanced the refractive (low q) peak (see Section 2.5).

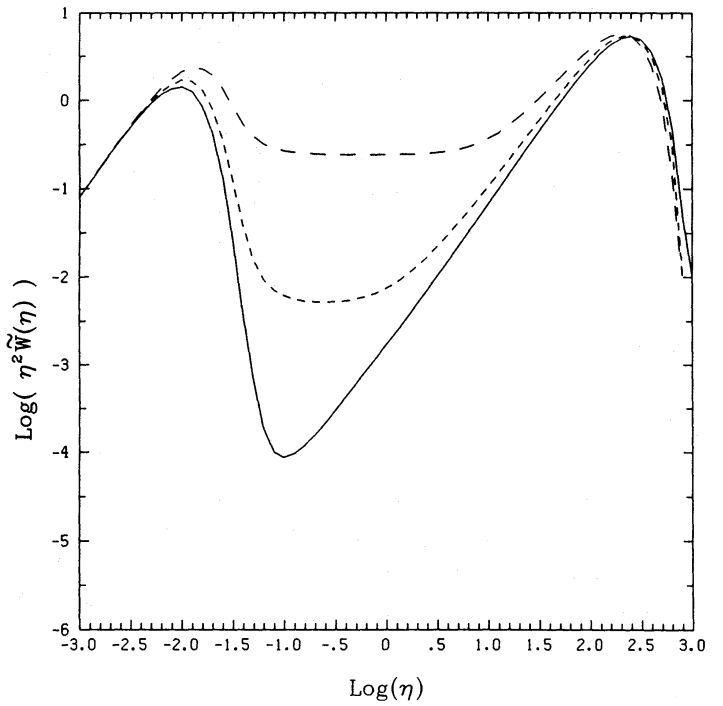

Figure 3. Flux variation power spectrum for $\beta=4.3$ (as in Fig. 2). Note that, even for $r_c=0$, the refractive fluctuations are markedly stronger than in the $\beta<4$ case. Since for $\beta>4$ large scales dominate the scattering, it has not been necessary to rescale $|A|$ for $r_c\neq0$.

the diffractive and refractive peaks. This is because for $\beta<4$ the image size $\theta_{\rm rms}$ is determined by the phase fluctuations on the scale r_c, so that $\theta_{\rm rms}\sim k^{-1}[D(r_c)]^{1/2}/r_c \propto r_c^{(\beta-4)/2}$ from equation (2.4.5). An important consequence of the renormalization of Q_0 is that the refractive peak increases in height, an effect that has been noted by Coles *et al.* (1987). For $\beta>4$, the phase gradient is dominated by the largest scales, hence by $r_{\rm out}$; but since the rms phase *curvature* decreases with scale ($\propto r^{(\beta-6)/2}$), the effect of the power at $r_{\rm out}$ is mainly to shift the image without distorting it, provided that $r_{\rm out}\gg r_{\rm ref}$. The scale that controls $\theta_{\rm rms}$ is then the one whose focal length $\approx z$, which implies that its linear size must be $\approx z\theta_{\rm rms}=r_{\rm ref}$. (These arguments have been given by BN and are further developed in Section 3.) From this observation one can determine $\theta_{\rm rms}$ self-consistently in terms of z and Q_0 (*cf.* BN), but for now it is sufficient to note that varying r_c does not affect $\theta_{\rm rms}$ for $\beta>4$, unless $r_c\geq r_{\rm ref}$.

A dramatic effect of the finite inner scale r_c is that the 'valley' in $q^2\tilde{W}(q)$ between the refractive and diffractive peaks is filled in with a region of constant power per logarithmic interval. According to the results of Appendix A, the filled region should extend over the range of q from

$$q_{\rm int}\equiv r_{\rm F}^{-2}|Q_0'|^{-1/2}r_c^{(4-\beta)/2}\equiv 2\pi/r_{\rm int} \tag{2.5.1}$$

at the left to

$$q_{\rm cdif}\equiv 2\pi r_c/r_{\rm F}^2\equiv 2\pi/r_{\rm cdif} \tag{2.5.2}$$

at the right. The 'intermediate' length scale $r_{\rm int}$ can be expressed in terms of the other scales as

$$r_{\rm int}=r_{\rm ref}, \qquad \beta<4 \tag{2.5.3}$$

$$r_{\rm int}=r_{\rm ref}(r_c/r_{\rm ref})^{(\beta-4)/2}, \qquad \beta>4. \tag{2.5.4}$$

The length scale $r_{\rm cdif}$ is the diffractive scale $(1/k\theta)$ associated with an image of projected size r_c on the thin screen. The height of the filled region in Figs 2, 3 is determined by the height of the $r_c=0$ spectrum at $q=q_{\rm cdif}$. Hence the total power contributed by the filled region is approximately

$$\left[\frac{\Delta F}{F}\right]_c^2=\int_{q_{\rm int}}^{q_{\rm cdif}} q\,dq\tilde{W}(q)=q_{\rm cdif}^2\tilde{W}(q_{\rm cdif})\ln(q_{\rm cdif}/q_{\rm int}). \tag{2.5.5}$$

We can estimate the importance of the contribution (2.5.5) by noting that if $r_c=0$, the total power in diffractive scintillations is always unity and that in the range $r_{\rm F}^{-1}\leq q\leq q_{\rm dif}$,

$$\tilde{W}(q)\equiv\begin{cases}\tilde{W}(q_{\rm dif}) & 2<\beta<4\\ \tilde{W}(q_{\rm dif})(q_{\rm dif}/q)^{\beta-4} & 4<\beta<6,\end{cases} \tag{2.5.6}$$

(*cf.* GN). Hence, since $r_{\rm dif}r_{\rm ref}=r_c r_{\rm cdif}=r_{\rm F}^2$,

$$\left[\frac{\Delta F}{F}\right]_c^2\sim 2(r_c/r_{\rm int})^2\ln(r_c r_{\rm int}/r_{\rm F}^2). \tag{2.5.7}$$

Note that $[\Delta F/F]_c$ increases rapidly as $r_c\to r_{\rm ref}$.

Our results seem paradoxical, particularly for $\beta>4$, where we have no need to renormalize $|A|$: by *decreasing* the power in ϕ, we have *increased* the power in $\Delta F/F$. The explanation lies in the development of caustics, about which we have more to say in Section 3.4.

3 Physics of caustics

In the previous section we found that the presence of an inner scale in the phase fluctuations introduces a new component in the flux spectrum. We associated this fluctuation power with caustics, which we study in greater detail here. In Section 3.1 we give a geometrical optics

description of caustics, borrowing freely from the notions of catastrophe theory. We classify the kinds of caustics that are allowed and identify the ones that could be relevant for interstellar radio scintillation. We find that the pure geometrical optics approach leads to infinite flux for an observer located exactly on the caustic, whereas in reality this divergence would be controlled by the finite wavelength of the radiation. This leads us to investigate the effect of diffraction on caustics. Section 3.2 introduces certain useful physical concepts about the wavefront and the nature of images. In particular, we describe a way of estimating the flux due to an image. Using these ideas, we discuss in Section 3.3 the effect of wave optics on caustics, considering in detail the fold caustic. A higher order caustic, the cusp, is described in Appendix B. Finally, we show in Section 3.4 that caustics do indeed produce the new component in the flux variation spectrum.

3.1 CAUSTICS AND CATASTROPHE THEORY – GEOMETRICAL OPTICS

As in Section 2, we consider a point source at infinity and a thin phase-changing screen at a distance z from the observer. The wave amplitude $\psi(X, Y)$ received by an observer at $\mathbf{R}=(X, Y)$ is given by equations (2.2.3) and (2.2.4). Note that the phase $\phi(x, y) \propto k^{-1}$ because of dispersion in the interstellar medium, and the second geometrical term is $\propto k$. The integral (2.2.3) is dominated by points of stationary phase on the screen that satisfy

$$\frac{\partial \zeta}{\partial x}=\frac{\partial \phi}{\partial x}+\frac{(x-X)}{r_F^2}=0,$$

$$\frac{\partial \zeta}{\partial y}=\frac{\partial \phi}{\partial y}+\frac{(y-Y)}{r_F^2}=0. \tag{3.1.1}$$

For the moment we will ignore the dispersion in ϕ and consider the geometrical optics limit $k \rightarrow \infty$ where we can talk of 'rays'. In this limit, we can say that the only rays that an observer at (X, Y) receives are those that emerge from points on the screen satisfying equation (3.1.1). Each such ray represents an image of the point source. In fact (3.1.1) provides a many-to-one mapping between each point (x, y) on the screen at which a ray from the source is incident, and the point (X, Y) on the observer plane at which the ray is received. Keeping this mapping in mind, we consider below the trajectories of a bundle of rays emanating from the point source.

Rays emanating from a point source can cross, so that an observer located at the crossing point may see two or more images of the source. As the observer moves, or as the conditions of observation change, two or more rays may coalesce. The point, line or surface where this occurs is a *caustic*. It is an envelope formed by the family of rays from the point source. In the geometrical optics approximation, a point source observed from a caustic will subtend a finite angular size and therefore will be infinitely amplified. Although there are an infinite number of ways that rays can coalesce on a caustic, they can be classified into a few elementary forms or *catastrophes* whose important features and scaling laws are structurally stable to small perturbations. In general there is a hierarchy of catastrophes corresponding to the local phase being flat to successively higher order in a polynomial expansion of the excess phase about points satisfying (3.1.1). The classification of structurally stable catastrophes according to such a hierarchy was developed by Thom (1975), Arnol'd (1975) and others, and has found application in a number of fields (e.g. Poston & Stewart 1982). Berry & Upsill (1980) have summarized the application to wave propagation in random media. We present here a simplified discussion of the main features, using the concepts introduced in the previous sections.

One way to trace rays is to evaluate the optical distance or phase $\zeta(s_i)=\omega \int n\, dr/c$ where $n(\mathbf{r})$ is the refractive index along a virtual path parameterized by a set of *state variables* s_i (e.g. coordinates x, y on the scattering screen). Actual rays correspond to extrema of $\zeta(s_i)$ (equation 3.1.1),

according to Fermat's principle, and moving the observer represents changing *control variables* C_i (e.g. observer coordinates X, Y). Finding the ray(s) that pass through some point defines a gradient map from C_i to s_i and caustics are singularities of this map, where the Hessian $|\partial^2\zeta/\partial s_i \partial s_j|$ vanishes. In the case of equation 2.2.4, at least one of the principal curvatures of the phase surface $\zeta(x, y; X, y)$ passes through zero at a caustic.

If we confine attention to catastrophes with up to five rays there are seven elementary forms. These are conventionally described analytically by expanding the phase in a power series around the caustic using the minimum number of state variables (one or two) and control variables (up to four) needed to exhibit the catastrophe. In mathematical language, two catastrophes are said to be equivalent if one can be transformed into the other using a diffeomorphism of the control variables and, at each point in control space, a diffeomorphism of the state variables.

The classification of elementary catastrophes can be understood (though not properly derived) from basic considerations. In the simplest case (the fold) just two rays coalesce. If we regard the angle between the rays that represent the two images as being a function of the C_i, then just one control variable need be varied in the generic case to cause this angle to vanish. When n images coalesce, $n-1$ angles must vanish and we must search for the catastrophe in a control space of dimensionality (codimension) $n-1$. That is to say, we need $n-1$ control variables to describe an elementary catastrophe involving n rays.*

Catastrophes that can be described using one state variable are called cuspoids and are known successively as the fold ($n=2$), cusp ($n=3$), and swallowtail ($n=4$). The fold involves two images and so the phase must have two turning points, i.e. it must be cubic in the state variable (see Fig. 4). It is conventional to choose the coefficient of the linear term to be the single control variable, and as we are using the simplest analytical description of the catastrophe we remove the quadratic term by an elementary change of variable and drop the constant. Hence the phase can be expanded locally as a standard polynomial.

$$\bar{\zeta}_{\text{fold}} = s^3/3 + c_s \tag{3.1.2}$$

for a suitably defined state variable s and control parameter C. Similarly, the standard form of the cusp is

$$\bar{\zeta}_{\text{cusp}} = s^4/4 + C_2 s^2/2 + C_1 s \tag{3.1.3}$$

and so on.

One state variable is not always adequate as we demonstrate in Fig. 5. Elementary catastrophes requiring two state variables are known as umbilics; by sketching topologically distinct nestings of contours it is easy to verify that a minimum of four rays must be involved for such events. There are two distinct possibilities when $n=4$, the hyperbolic and the elliptic umbilic (see Fig. 5). The character of the remaining two elementary five-ray catastrophes is similarly derived.

Although this examination of state variable space has enabled us to classify the elementary catastrophes, we are more interested in the behaviour of the observed images as we vary the control parameters. Referring to equation (3.1.2) for the fold we see two images when $C<0$, merging at $C=0$ at the caustic and vanishing for $C>0$. In the case of a cusp, three images can become one image as the two control parameters C_1, C_2 are varied (Fig. 4). In this way we can subdivide the control space into distinct regions identified by the number of images of a point source observable from within them. As an observer crosses the boundary from one region to another, he will generally encounter a fold catastrophe and the number of images will change by ± 2. Occasionally, however, he may encounter a higher order catastrophe, when the phase

*In fact, we don't necessarily need to vary $n-1$ control variables to make n images coalesce. In the case of the swallowtail catastrophe, discussed in Fig. 5, four images can be made to vanish by varying only two control parameters. However, we do need to vary $n-1=3$ independent parameters to find the full singularity.

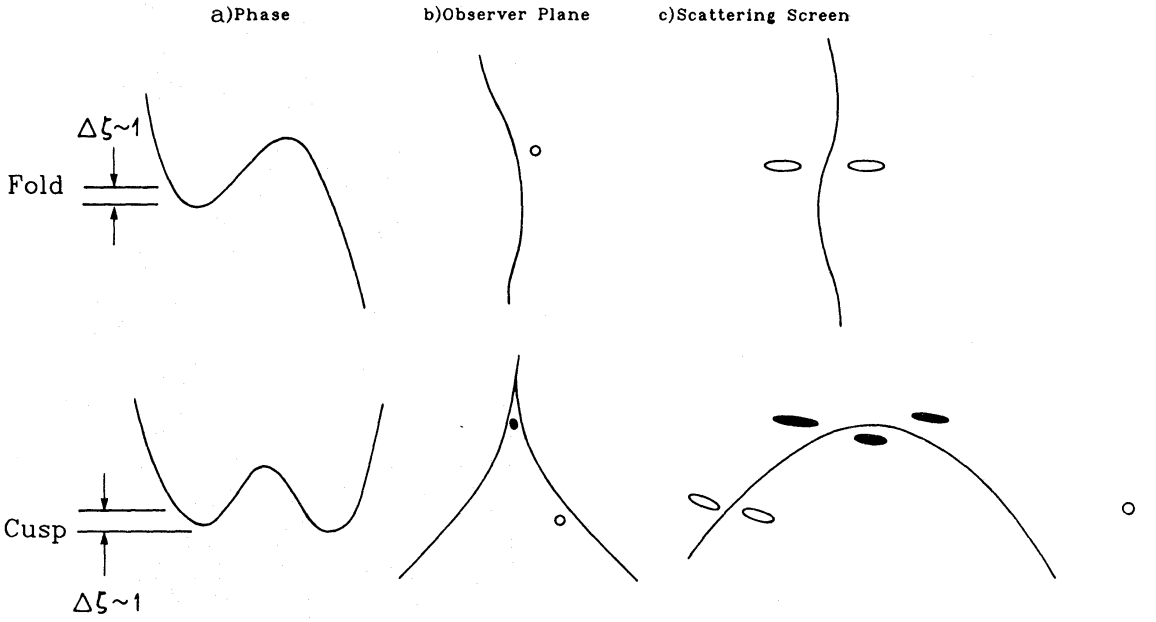

Figure 4. (a) Fold and cusp catastrophes exhibited in one dimension as a function of the state variable s. The vertical coordinate denotes excess phase ζ and sub-images are found at the extrema of ζ. The width of an extremum at a depth ~ 1 describes the sub-image patch size and, hence, intensity. As the control parameter X is varied approaching the singularity, the extrema (sub-images) broaden and merge. Crossing the singularity, the three cusp images merge, leaving a single extremum; there are no extrema on the dark side of a fold. (b) Loci of the fold and cusp caustic on the observer plane. For the cusp, the filled circle represents an observer near the cusp point, while the open circles represent observers approaching fold caustics from the bright side. (c) Loci of the fold and cusp caustics on the scattering screen. The open circles in (b) form two images each, elongated and hence magnified. At the cusp point three images merge tangentially to the line of the fold asymptotes. For a source inside the cusp there will be (globally) three images on the scattering screen; a source encountering the cusp from the outside will produce a single bright image.

surface will be flatter to higher order. The images will appear to brighten and, in the case of the swallowtail and the hyperbolic umbilic, the number of images may change by ± 4.

It is in determining the types of caustic which will be encountered that the power of catastrophe theory becomes apparent. If, as is the case for a radio telescope passing through the wave field formed by interstellar irregularities, there is only one control variable (time or equivalently one observer coordinate, say X), a single control variable is all that can be explored. We therefore only expect to encounter folds (though we may pass close to a cusp). On the other hand, if we were able to explore two observer coordinates, e.g. by taking a photograph or, as in the present application, by varying the wavelength of observation, then we can also expect to encounter cusps. In order to encounter higher order catastrophes, we should have to explore three or more coordinates, e.g. X, Y and z. These conclusions are independent of the complexity of the ray paths (i.e. the number of state variables necessary to specify them.)

Powerful as these results may be, they do not allow us to quantify the contribution of caustics to the flux variations. To do this, we must revert to a specific model of the propagation and the conditions of observations and we use the thin screen model of Section 2.1. This also allows us to illustrate the general principles we have just developed. We first consider the fold catastrophe.

Our procedure is to expand $\zeta(x, y; X, Y)$ in a Taylor series around the singularity in the mapping, translating and rotating the coordinate system for convenience and retaining only those terms that can dominate the total phase and its first two derivatives over the region of interest. We fix the origin of the (x, y) coordinates on the fold and orientate the axes along the principal axes of the phase curvature tensor with the curvature vanishing along the x direction. The (X, Y) axes are

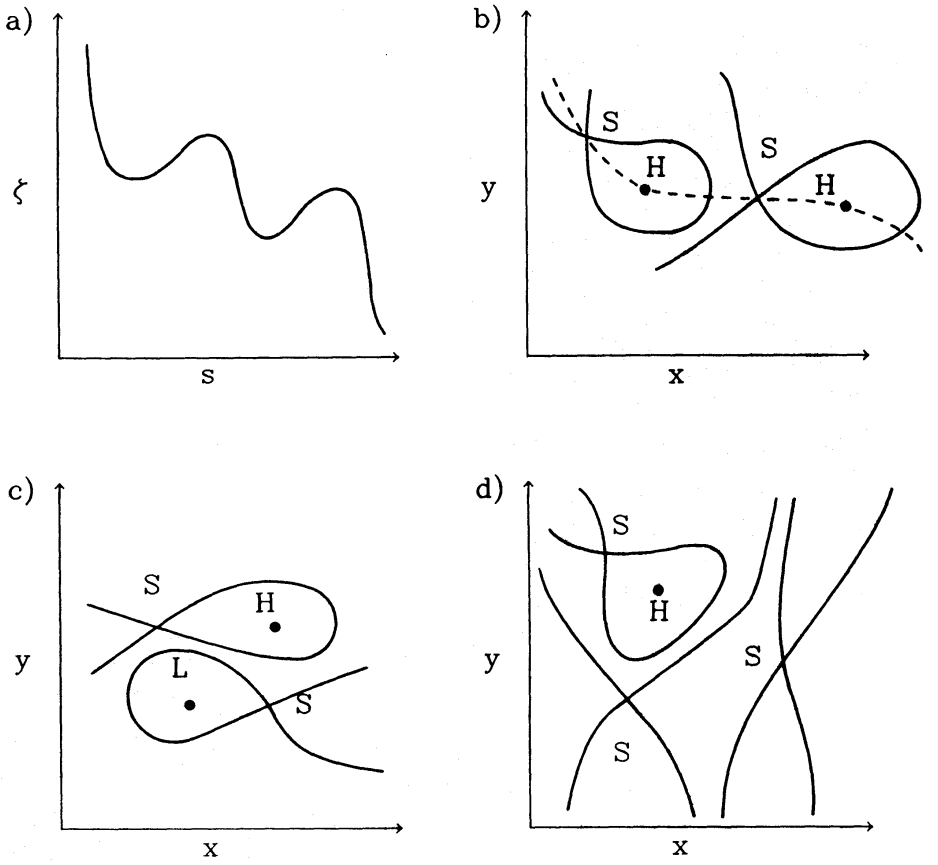

Figure 5. Catastrophes of codimension 3. (a) Swallowtail exhibited in one dimension. There are four turning points corresponding to four rays which can be made to merge simultaneously. (b) When there are two state variables (i.e. coordinates x, y on the screen), we can plot contours of constant phase ζ to exhibit, in this example of a swallowtail, two maxima (H) and two saddles (S). Note that only one variable (which can be chosen to advance along the dashed line) is necessary to exhibit the catastrophe. (c) Contours of constant phase $\zeta(x, y)$ near a hyperbolic umbilic catastrophe. Just as was the case with the swallowtail, there are four stationary points (two saddles S, a maximum H and a minimum L). This catastrophe is nevertheless distinct from the swallowtail because two state variables must be used to describe it. (d) Contours of constant phase $\zeta(x, y)$ near an elliptic umbilic catastrophe. This nesting of contours cannot be continuously deformed into either the swallowtail or the hyperbolic umbilic and therefore represents a topologically distinct catastrophe.

orientated parallel to the (x, y) axes so that the caustic lies along the Y axis. The phase $\zeta(x, y; X, Y)$ then takes the form

$$\zeta(x, y; X, Y)=\tfrac{1}{3}ax_3+\tfrac{1}{2}bx_2y+\tfrac{1}{2}cy^2-ax-\beta y, \qquad a>0, \tag{3.1.4}$$

where $a=\phi_{xxx}/2$, $b=\phi_{xxy}$, $c=\phi_{yy}-k/z$. The state variables are x, y, and the control variables are $\alpha=kX/z, \beta=kY/z$. The term in x^2 vanishes with the curvature, the axis rotation has removed the xy term and higher order terms, xy^2, y^3 are demonstrably subdominant in the vicinity of the stationary points.

The rays are located at the stationary points of ζ, i.e.

$$x\sim\pm\left(\frac{\alpha}{a}\right)^{1/2}+\frac{\alpha b^2}{4a^2c}-\frac{\beta b}{2ac}$$

$$y\sim\frac{\beta}{c}-\frac{\alpha b}{2ac} \tag{3.1.5}$$

plus higher order terms. Two images can be seen when $\alpha>0$. The Hessian is

$$\begin{vmatrix} \partial^2\zeta/\partial x^2 & \partial^2\zeta/\partial x\partial y \\ \partial^2\zeta/\partial x\partial y & \partial^2\zeta/\partial y^2 \end{vmatrix} = (2ax+by)c-b^2x^2 \sim \pm 2ac(\alpha/a)^{1/2}$$

retaining only the leading terms. As the magnification is the reciprocal of the Hessian, it diverges near the origin on the critical line $y=-2ax/b$ which (since a and b are of the same order) has an arbitrary orientation with respect to the principal axes of the curvature tensor.

We wish to relate the specific form of the fold catastrophe produced by a single screen (equation 3.1.4) to the standard polynomial (equation 3.1.2). To do this, we must first augment the polynomial with the simplest terms describing the extra state variable and control variable C_2, i.e.

$$\bar{\xi}=s^3/3+C_1S+t^2/2+C_2t. \tag{3.1.7}$$

The state and control variables are then related, to lowest order, through

$$s=a^{1/3}x$$

$$t=c^{1/2}y$$

$$C_1=-a^{-1/3}\alpha$$

$$C_2=\alpha b/(2ac^{1/2})-c^{-1/2}\beta. \tag{3.1.8}$$

In other words, the observer coordinate system must be sheared to bring the catastrophe into standard form. In Appendix B, we repeat this exercise for the cusp.

3.2 WAVEFRONTS, IMAGES, AND POWER SPECTRA

We begin our discussion of diffraction in caustics by introducing in this section certain key concepts related to the diffraction integral (2.2.3) and the power spectrum (2.3.3). We will make frequent use of the length scales r_F, r_{ref}, r_{dif}, r_c, r_{int} and r_{cdif} introduced in Section 2 and listed in Table 1.

A crucial concept in what follows is that of the 'image' of a point source as projected on the thin scattering screen. Consider first the case when the excess phase $\phi(x, y)$ vanishes. The diffraction

Table 1. Important length-scales for refractive scintillation from a density perturbation power spectrum with an inner scale. Scales are introduced and/or defined by the indicated equation in the text and are listed in order of size (for $r_F<r_c<r_{int}$).

Name	Symbol	Definition		Equation
diffractive	r_{dif}	$= r_F^2/r_{ref} \sim 1/k\theta_{rms}$		(2.4.7)
cut-off diffractive	r_{cdif}	$= 2\pi/q_{cdif} = r_F^2/r_c$		(2.5.2)
Fresnel	r_F	$= (z/k)^{1/2}$		(2.2.2)
inner or cut-off	r_c	$= 2\pi/q_c$		(2.3.3)
intermediate	r_{int}	$= \begin{cases} r_{ref}, \\ r_{ref}(r_c/r_{ref})^{(\beta-4)/2}, \end{cases}$	$\beta<4$ $\beta>4$	(2.5.3),(2.5.4)
refractive	r_{ref}	$= z\theta_{rms}$		(2.4.4)
outer	r_{out}	$= 2\pi/q_{out}$		(2.3.3)

integral (2.2.3) involves only the geometrical contribution to $\zeta(x, y, X, Y)$ which has a stationary point at $[x(X, Y), y(X, Y)]$. The dominant contribution to $\psi(X, Y)$ arises from a coherent region of radius $\sim r_F$ around this point, and the contribution from the rest of the screen cancels out due to the oscillating phase. We then say that the image of the source is centred on (x, y) and has a coherent 'size' $\sim r_F$. A more detailed discussion of the physics of the diffraction integral can be found in standard texts (e.g. Born & Wolf 1980; Hecht & Zajac 1974), which formalize the idea of the 'coherent patch' by means of a geometrical construction in the complex plane using the Cornu spiral.

Consider now the case of non-vanishing $\phi(x, y)$. Provided a particular restriction that is discussed below is satisfied, it is still correct to say that the diffraction integral (2.2.3) is dominated by regions of stationary phase, i.e. regions in the vicinity of points (x, y) that satisfy equations (3.1.1). For the strong scintillation regime in which we are interested, there will in general be a large number of points on the screen that satisfy these conditions. We will therefore call the set of such points the *image* of the source, and each individual point a *sub-image*. To calculate the flux contributed by a single sub-image located at (x_I, y_I), we make a local quadratic expansion of $\zeta(x, y)$. Let us assume that the x, y axes are aligned with the principal axes of the curvature tensor of ζ. We then have

$$\zeta(x, y) \sim \zeta(x_I, y_I) + \frac{1}{2}\frac{\partial^2 \zeta}{\partial x^2}(x - x_I)^2 + \frac{1}{2}\frac{\partial^2 \zeta}{\partial y^2}(y - y_I)^2. \tag{3.2.1}$$

As before, we can say that the dominant contribution to the flux arises from a coherent region around (x_I, y_I), but now the coherent patch is in general an ellipse, whose semi-axes l_x, l_y are given by the condition that ζ changes by ~ 1 rad, i.e.

$$l_x \sim \left(\frac{\partial^2 \zeta}{\partial x^2}\right)^{-1/2}, \qquad l_y \sim \left(\frac{\partial^2 \zeta}{\partial y^2}\right)^{-1/2}. \tag{3.2.2}$$

The flux F_s contributed by the sub-image is the square of the wave-amplitude, which leads to

$$F_s \sim l_x^2 l_y^2 / r_F^4. \tag{3.2.3}$$

This discussion in terms of individual sub-images is valid so long as the quadratic expansion (3.2.1) extends over an area much larger than the coherent central patch. In other words, the power series of ζ must be valid for changes in $\zeta \gg 1$. As we show below this condition holds in all of the situations considered in this paper because of the presence of the cut-off in the spectrum of phase fluctuations. In the case of an extended spectrum with no cut-off, more care is needed. One should identify a wavevector q_{max} such that the rms phase fluctuations due to $q > q_{max}$ is $\lesssim 1$ rad (e.g. Gapper & Hewish 1981; RNB), and the power from these large wavevectors should be filtered out. The residual $\zeta(x, y)$ can then be approximately analysed as above.

We now consider the effect of the power spectrum $Q(q)$ given in equation (2.3.3). For the particular power-law form assumed, the mean-square phase fluctuation varies with q as $\sim k^{-2} Q_0 q^{2-\beta}\, d(\ln q)$. Let us define ϕ_c to be the rms phase fluctuation due to scales near the cut-off scale r_c, i.e.

$$\phi_c \equiv k^{-1} Q_0^{1/2} r_c^{(\beta-2)/2}. \tag{3.2.4}$$

Hills and valleys in $\phi(x, y)$ tend to have linear sizes $\sim r_c$, and ϕ_c measures the typical phase difference between a neighbouring hill and valley.

The geometrical scattering angle at any point on the screen is given by $k^{-2}\nabla\phi$, and so the rms scattering angle due to scales with wavevector q is $\theta(q) \sim k^{-1} Q_0^{1/2} q^{(4-\beta)/2}$. We then see that the intermediate scale r_{int} defined in (2.5.1) is related to the scattering angle at the cut-off scale by

$$r_{int} \sim z\theta(q_c) \sim r_F^2 \phi_c / r_c. \tag{3.2.5}$$

The total mean square scattering angle from all scales is

$$\langle\theta^2\rangle\sim\frac{k^{-4}}{(2\pi)^2}\int\int d^2q q^2 Q(q)=Q_0\frac{k^{-4}}{2\pi}\int_{q_{\rm out}}^{q_c} d(\ln q)q^{4-\beta}. \tag{3.2.6}$$

The behaviour of this integral depends critically on the sign of $4-\beta$. When $\beta<4$, $q^{4-\beta}$ is maximal at the upper limit q_c; hence

$$\theta_{\rm rms}\sim\theta(q_c)\sim(Q_0/k^4 r_c^{4-\beta})^{1/2}, \qquad \beta<4. \tag{3.2.7}$$

This is the angular size of the image, and associated with it is the refractive scale (*cf.* equation 2.4.4)

$$r_{\rm ref}=r_{\rm int}\sim r_{\rm F}^2\phi_c/r_c, \qquad \beta<4. \tag{3.2.8}$$

When $\beta>4$, the refractive scale is no longer given by equation (3.2.8) since the scattering angle is dominated by small q, i.e. large length-scales. In fact, the scattering angle diverges unless there is an 'outer scale' to the spectrum. However, the divergence does not concern us since the largest scales only steer the mean position of the image without contributing to its broadening. The image broadening is in fact determined by the largest scale which can focus radiation at the observer, i.e. which satisfies the condition $r_{\rm F}^2\nabla^2\phi\sim1$. Since the contribution to $\partial^2\phi/\partial\mathbf{r}^2$ from wavevectors $\sim q$ varies as $(Q_0k^{-2}q^{6-\beta})^{1/2}$, therefore we obtain

$$r_{\rm ref}\sim(r_{\rm F}^2\phi_c)^{2/(6-\beta)}r_c^{-(\beta-2)/(6-\beta)}, \qquad \beta>4. \tag{3.2.9}$$

Note that equations (3.2.8) and (3.2.9) agree for the marginal $\beta=4$ case. Note also that (3.2.5) and (3.2.9) are consistent with (2.5.4), which gives the relation between $r_{\rm int}$ and $r_{\rm ref}$ for $\beta>4$ (the two scales are equal for $\beta<4$).

The theory of Section 2 and Appendix A requires three conditions to be valid, which can be written as

$$r_{\rm ref}\gg r_{\rm F}, \tag{3.2.10}$$

$$r_{\rm ref}\gg r_c, \tag{3.2.11}$$

$$r_c\gg r_{\rm F}^2/r_{\rm int}. \tag{3.2.12}$$

The first condition ensures strong scattering, and the second ensures multi-path propagation, both of which are crucial for the occurrence of strong scintillation. The third condition can be rewritten with the help of equation (3.2.5) as

$$\phi_c\gg1. \tag{3.2.13}$$

But this is precisely the condition we required earlier for the concept of sub-image to make sense. We thus see that the filling up of the valley in $q^2\tilde{W}(q)$ that we discussed in Section 2 occurs only when sub-images can be distinguished individually.

Let us now estimate the mean flux of sub-images. From equation (2.2.4), we have

$$l_x,\ l_y\sim\left(\frac{\partial^2\zeta}{\partial\mathbf{r}^2}\right)^{-1/2}\sim\left[\left(\frac{\partial^2\phi}{\partial\mathbf{r}^2}\right)_{\rm rms}-\frac{1}{r_{\rm F}^2}\right]^{-1/2}. \tag{3.2.14}$$

Applying equation (3.2.11) we see that

$$r_{\rm F}^2\left(\frac{\partial^2\phi}{\partial\mathbf{r}^2}\right)_{\rm rms}\sim\left(\frac{r_{\rm ref}}{r_c}\right)\gg1, \qquad \beta<4,$$

$$\sim\left(\frac{r_{\rm ref}}{r_c}\right)^{(6-\beta)/2}\gg1, \qquad \beta>4. \tag{3.2.15}$$

Thus $(\partial^2\phi/\partial\mathbf{r}^2)_{\rm rms} \gg 1/r_{\rm F}^2$ in all regimes of interest, and so the contribution of the latter to l_x, l_y can be neglected. Equation (3.2.3) then gives the flux from each sub-image as

$$F_{\rm s} \sim (r_{\rm c}/r_{\rm ref})^2, \qquad \beta<4, \tag{3.2.16}$$

$$F_{\rm s} \sim (r_{\rm c}/r_{\rm ref})^{(6-\beta)}, \qquad \beta>4. \tag{3.2.17}$$

Since the mean total flux is conserved, this implies that the mean number of sub-images in the image is

$$N \sim (r_{\rm ref}/r_{\rm c})^2, \qquad \beta<4, \tag{3.2.18}$$

$$N \sim (r_{\rm ref}/r_{\rm c})^{(6-\beta)}, \qquad \beta>4. \tag{3.2.19}$$

Equation (3.2.18) has a transparent interpretation; it says that the image of size $r_{\rm ref}$ is uniformly covered with sub-images with mean separation $\sim r_{\rm c}$. Equation (3.2.19) is more interesting. The wavefront still has hills and valleys of typical linear size $r_{\rm c}$, but the number of sub-images is now $<(r_{\rm ref}/r_{\rm c})^2$. The reason is that, since the rms scattering angle due to perturbations on any given scale is larger than the broadening due to scattering by all smaller scales, the sub-images are clustered at each scale. This results in a reduction of the number of images within $r_{\rm ref}^2$ and gives the image a hierarchical fractal structure (e.g. GN).

For the discussion of caustics in Section 3.3 we need estimates of higher derivatives of ζ. The geometrical term does not contribute, and so

$$\left(\frac{\partial^3\zeta}{\partial\mathbf{r}^3}\right)_{\rm rms} \sim \frac{\phi_{\rm c}}{r_{\rm c}^3}, \tag{3.2.20}$$

$$\left(\frac{\partial^4\zeta}{\partial\mathbf{r}^4}\right)_{\rm rms} \sim \frac{\phi_{\rm c}}{r_{\rm c}^4}. \tag{3.2.21}$$

It is assumed that the cut-off spectrum beyond $q_{\rm c}$ is at least as steep as q^{-10} for these estimates to be accurate; weaker cut-offs will give somewhat different coefficients.

3.3 DIFFRACTION EFFECTS IN CAUSTICS

As we discussed in Section 3.1, a caustic is obtained when one or both of the principal curvatures of $\zeta(x, y)$ vanishes at a sub-image. Equations (3.2.2) and (3.2.3) show that the flux then diverges within the quadratic approximation. Indeed, in the limit of pure geometrical optics, the image becomes infinitely bright, although, for a finite wavelength, diffraction controls this divergence and there is a finite magnification of each sub-image. To treat this case, we will need to generalize the discussion of Section 3.2 by introducing the effects of higher order terms in the polynomial expansion of ζ. The idea of the coherent patch is still correct and so the flux is determined by the coherent contribution from the region inside the $\Delta\zeta\sim1$ contour around the sub-image. However, when the curvature in a particular direction vanishes, as it does at a caustic, the dimension of the coherent patch along that direction is no longer determined by the curvature but instead by the first non-vanishing term in the Taylor series.

Let us consider the fold caustic. From equation (3.1.4), the phase ζ takes the form

$$\zeta(x, y, X, Y) = \tfrac{1}{3}ax^3 + \tfrac{1}{2}bx^2y + \tfrac{1}{2}cy^2 - \frac{(Xx+Yy)}{r_{\rm F}^2}, \qquad |x|, |y| \ll r_{\rm c}, \tag{3.3.1}$$

$$|a| \sim |b| \sim \phi_{\rm c}/r_{\rm c}^3, \qquad |c| \sim \phi_{\rm c}/r_{\rm c}^2. \tag{3.3.2}$$

In the discussion that follows, we restrict our attention to the sub-images found in the small area,

$|x|, |y| \lesssim r_c$. There are, of course, $\sim N$ other sub-images in the image, but they do not influence the behaviour of the caustic. The locations of sub-images in this 'local' area for a given observer position (X, Y) are described by (3.1.5) which, to lowest order, gives

$$X = r_F^2(ax^2) \sim r_{int}(x/r_c)^2, \tag{3.3.3}$$

$$Y = r_F^2(cy + bx^2/2) \sim r_{int}[(y/r_c) + (x/r_c)^2/2], \tag{3.3.4}$$

where we have used equations (3.3.2) and (3.2.5). The caustic is defined by the vanishing of the determinant of the curvature tensor (3.1.6), which gives a locally straight line at arbitrary orientation on the (x, y) screen and, to lowest order, the line $X=0$ on the ground. This line in the observer plane is perpendicular to the x axis, namely the direction on the screen along which the curvature vanishes. Globally, however, the quadratic term in equation (3.1.6) indicates that the caustic is curved, with a radius of curvature $\sim r_{int}$. In fact, we can state on the basis of this result that each patch on the scattering screen of size $\sim r_c \times r_c$ produces a fold caustic *loop* on the observer plane of 'radius' $\sim r_{int}$, as shown in Fig. 6. Note that, since $r_{int} \propto z$, the length of these caustic loops (and hence the frequency of fold crossings) will increase with z while the number of loops remains constant. This is a crucial difference between one- and two-dimensional screens. In the former case the mean fold spacing is just $\sim r_c$, independent of z.

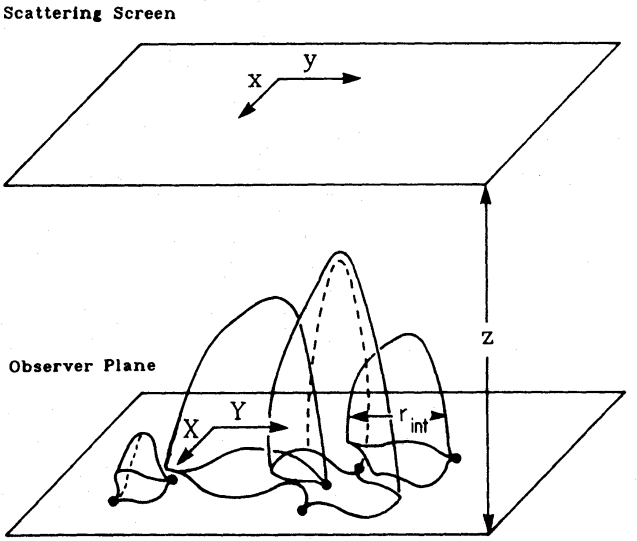

Figure 6. Caustic sheets. Beyond a focal length from the scattering screen, rays can merge and caustics will form. The surface on which two rays coalesce describes the fold sheets, whose intersection with the observer plane forms fold loops. The lines on these sheets where three rays merge intersect the plane in cusp points.

We now investigate the variation of received flux for observer positions in the vicinity of the caustic. The qualitative features are shown in Fig. 4. It is sufficient to consider the line $Y=0$ on the observer plane, locally perpendicular to the caustic. The behaviour we determine here can be applied elsewhere on the caustic by simple translation. To the order at which we are working, $Y=0$ implies $y=0$ (equation 3.3.4), and so equation (3.3.1) simplifies to the standard form (*cf.* equation 3.1.7)

$$\zeta(x, y, X) = \tfrac{1}{3}ax^3 + \tfrac{1}{2}cy^2 - \frac{Xx}{r_F^2}, \qquad a > 0. \tag{3.3.5}$$

The maximum flux is clearly received at $X=0$, and the corresponding image is centred on $x=y=0$. At this point, the curvature of ζ parallel to x vanishes and so the size of the coherent patch along x

is determined from the third derivative of ζ. The area of the coherent patch on the screen is thus $\sim(a^{-1/3})\times(c^{-1/2})\sim r_c^2/\phi_c^{5/6}$. The received flux is thus

$$F_{f,\max}\sim\frac{r_c^4}{\phi_c^{5/3}r_F^4}\sim\phi_c^{1/3}\bar{F}_s\sim\phi_c^{1/3}/N. \tag{3.3.6}$$

We see that the maximum flux is greater than the mean flux of a sub-image by the factor $\phi_c^{1/3}$. For $X\neq0$, equation (3.3.3) with $y=0$ can be written as

$$x=\pm\left(\frac{X}{ar_F^2}\right)^{1/2}\sim\pm\left(\frac{X}{r_{int}}\right)^{1/2}r_c, \qquad y=0. \tag{3.3.7}$$

Equation (3.3.7) reveals the characteristic feature of the fold caustic, namely that there are two sub-images for observer position on one side of the caustic (the side $X>0$ when a is positive), and no sub-images on the other side ($X<0$). Note also the square-root mapping between the image and observer positions.

For non-zero positive X, the extremum at $x=+(X/ar_F^2)^{1/2}$ is at a minimum of ζ with respect to variations in x, while the other extremum is at a maximum. The phase difference between the two extrema is

$$\Delta\zeta=\tfrac{4}{3}ax^3\sim\phi_c\left(\frac{X}{r_{int}}\right)^{3/2}. \tag{3.3.8}$$

When $\Delta\zeta<1$, the two extrema cannot be distinguished as different sub-images, and we must think of the image as a single coherent patch with flux $\sim F_{f,\max}$. For $\Delta\zeta>1$, the two images separate and can produce interference fringes. Thus, the width X_f of the central fringe in the fold caustic is given by the condition $\Delta\zeta\sim1$, which leads to

$$X_f\sim r_{int}/\phi_c^{2/3}. \tag{3.3.9}$$

Note that this is smaller than the size of the caustic loop, which is $\sim r_{int}$.

For $X>X_f$, we can compute the fluxes of the two images through equations (3.2.2) and (3.2.3). We find

$$l_x\sim(ax)^{-1/2}\sim\phi_c^{-1/2}(X/r_{int})^{-1/4}r_c,$$
$$l_y\sim c\sim\phi_c^{-1/2}r_c, \tag{3.3.10}$$

and this leads to an observed flux variation of the form

$$F(X)\sim\left(\frac{X_f}{X}\right)^{1/2}F_{f,\max}. \tag{3.3.11}$$

Thus the flux falls off as $X^{-1/2}$ away from the caustic, in agreement with the geometrical optics result. However, equation (3.3.11) only determines the *mean* flux. The two images will in fact interfere with each other, leading to flux oscillations with separation $S(X)$ varying as

$$S(X)\sim(d\Delta\zeta/dX)^{-1}\sim\left(\frac{X_f}{X}\right)^{1/2}X_f. \tag{3.3.12}$$

The fringe spacing decreases away from the caustic. For negative X, there is no extremum in the phase function; hence there are no images and the flux falls off rapidly on a scale $\sim X_f$.

The entire discussion above was for the case of the fold caustic. A similar analysis can be made for higher order caustics. Appendix B gives the results for the cusp caustic.

3.4 SPECTRUM OF FLUX VARIATIONS DUE TO CAUSTICS

We now make the connection between the results of Section 2 and what we have discussed so far in Section 3. In the asymptotic regime ($\phi_c \to \infty$) considered in Section 2, the fold caustic will dominate over the cusp caustic, and so we limit ourselves to estimating the contribution of folds to the flux spectrum.

We have seen that folds generally form loops in the observer plane of size $\sim r_{\rm int}$; hence the smallest q at which the fold can contribute is $\sim q_{\rm int}$. The smallest length-scales in the flux variation arise from the fine interference fringe structure in the fold. From equation (3.3.14) we see that for $X \sim r_{\rm int}$, this gives $q_{\rm max} \sim r_{\rm int}^{1/2}/X_{\rm f}^{3/2} \sim \phi_c/r_{\rm int} \sim q_{\rm cdif}$. These are precisely the two limits between which the new component in the flux spectrum occurs (see equations 2.5.1, 2.5.2). Let us now consider the shape of the spectrum. We substitute equation (3.3.8) into (2.2.3) and do the integral over y to obtain the wave-amplitude at $(X, 0)$.

$$\psi(X, 0) = \frac{1}{\sqrt{c r_{\rm F}^2}} \int dx \exp\left(\frac{i}{3} a x^3 - i \frac{k}{z} X x\right). \tag{3.4.1}$$

The Fourier transform of the one-dimensional flux pattern along X is

$$\tilde{I}(q) \equiv \int_{-\infty}^{\infty} dX\, \psi^*(X, 0)\psi(X, 0) \exp(iqX)$$

$$= \frac{1}{\sqrt{c}} \int_{-\infty}^{\infty} dX \exp\left[\frac{i}{3} a \left(\frac{3z}{k} q X^2 + \frac{3z^2 q^2}{k^2} X + \frac{z^3 q^3}{k^3}\right)\right]. \tag{3.4.2}$$

This Gaussian integral can be done exactly and gives a power-law scaling in one direction

$$|\tilde{I}(q)|^2 \propto q^{-1}. \tag{3.4.3}$$

In two dimensions, we get $\tilde{I}(\mathbf{q}) \propto q_x^{-1} \delta(q_y)$ if the fringes are parallel to the Y direction, as in our calculations. Averaging over all orientations of the fold, we then get

$$|\tilde{I}(q)|^2 \propto q^{-2}, \tag{3.4.4}$$

which is identical to the form in Section 2 and Appendix A for the extra contribution due to the cut-off.

It has been shown by Berry (1977) that for light rays reflecting off a randomly corrugated surface, the flux variations vary as

$$(\Delta F/F)^2 \sim \ln(k) \tag{3.4.5}$$

as the limit of geometrical optics is approached. Berry further shows that the logarithmic divergence is due specifically to fold caustics. To approach this limit in our problem, we cannot simply take $k \to \infty$, because in view of the cold-plasma dispersion relation, this would entirely suppress scintillation. Instead, we must hold $\theta_{\rm rms}$ fixed as $k \to \infty$, which translates to fixing $k^{-2}C(r)$; in other words, we maintain a constant amplitude in the fluctuations of the optical path length. When the limit is taken in this way the intensity modulation due to the intermediate length-scales, $(\Delta F/F)_c^2$ from equation (2.5.7), obeys the scaling (3.4.5).

This completes our proof that the extra flux variation power introduced on scales between $r_{\rm dif}$ and $r_{\rm ref}$ by a cut-off in the phase spectrum is due to the effect of caustics, primarily fold caustics.

4 Observations of individual caustics

In Section 2 we discussed the contribution of the ensemble of caustics to the intensity fluctuation

spectrum. In this section and the next, we apply the results of the previous sections to the interstellar medium. In particular, we estimate the required magnitude of r_c if individual caustics are to have observable effects. Unless otherwise stated, we assume $\beta<4$.

We first estimate the consequences of imposing an inner scale r_c on the density perturbation spectrum for the flux variation power spectrum described in Section 2. We have a typical observing wavelength $\lambda_m \equiv (\lambda/1\,\mathrm{m}) \sim 1$ and a distance to the effective scattering screen of $D_{kpc} \equiv (D/1\,\mathrm{kpc}) \sim 1$, giving a Fresnel scale $r_F \sim 2\times10^{11}(\lambda_m D_{kpc})^{1/2}$ cm. We can express the strength of the scattering Q_0 in terms of a conventional C_{-4} (Armstrong, Cordes & Rickett 1981; BN) where $Q_0 \propto C_{-4}$ and $C_{-4}\sim1$ for nearby pulsars. If there were no inner scale, then from the measured scattering angles the refractive scale would be given by

$$r_{ref} \sim 3\times10^{13}(C_{-4}D_{kpc}^{\beta-1}\lambda_m^{\beta})^{1/(\beta-2)}\,\mathrm{cm} \qquad \beta<4$$

$$r_{ref} \sim 3\times10^{13}(C_{-4}D_{kpc}^{3}\lambda_m^{4})^{1/(6-\beta)}\,\mathrm{cm} \qquad \beta>4 \tag{4.1}$$

(*cf.* RNB). Imposition of a cut-off scale at r_c implies that for a given scattering strength (C_{-4}) the refractive scale would be reduced (for $\beta<4$). Thus, as discussed in Section 3, we must increase our estimate of C_{-4} by a factor $(r_c/r_{dif})^{4-\beta}$, ($\beta<4$), in order to keep fixed the measured scattering size. For $\beta>4$, the large scales dominate, so r_{ref} is not affected. RNB also show that for $\beta=11/3$ and no inner scale we have $(\Delta F/F)^2_{ref}=0.015C_{-4}^{-2/5}D_{kpc}^{-11/15}\lambda_m^{-17/15}$, while for $\beta>4$, GN show that $(\Delta F/F)^2_{ref}=\beta-4$. With these numbers, for $D_{kpc}=\lambda_m=C_{-4}=1$, $\beta=11/3$, we have $(\Delta F/F)^2_{dif}\sim1$ and $(\Delta F/F)^2_{ref}\sim0.015$. As an example, if we now posit an inner scale $r_c=7\times10^{12}$ (i.e. adequate to give prominent individual folds, see below) then, from the discussion in Section 2.5, $(\Delta F/F)^2_{ref}$ is increased to ~0.26 and from equation (2.5.7) the caustics contribute $(\Delta F/F)^2_c\sim0.93$. Adding the non-diffractive contributions in quadrature we get $\Delta F/F|_{rms,\,non\text{-}dif}\sim1.1$, sufficient to explain even the most strongly varying pulsars.

A somewhat larger inner scale may be required to resolve the individual diffraction catastrophes. Nevertheless, the presence of large non-Gaussian fluctuations in the slow modulation of pulsar flux suggests that large, detectable events may be occasionally present. To estimate this possibility, a reasonable criterion for detectability is that, at maximum, an individual catastrophe provide as much flux as from the time-averaged image as a whole. Suppose, more generally, we require a fold to have a peak flux α times the mean value; then from equations (3.3.6) and (3.2.5), in order for the merging images to be observable, we require that

$$r_c \gtrsim 7\times10^{12}\alpha^{3/7}(r_{int}/3\times10^{13}\,\mathrm{cm})^{5/7}(r_F/2\times10^{11}\,\mathrm{cm})^{2/7}\,\mathrm{cm}. \tag{4.2}$$

For a cusp, a similar criterion on the peak flux (equation B8) gives

$$r_c \gtrsim 4\times10^{12}\alpha^{2/5}(r_{int}/3\times10^{13}\,\mathrm{cm})^{3/5}(r_F/2\times10^{11}\,\mathrm{cm})^{2/5}\,\mathrm{cm}. \tag{4.3}$$

From these expressions it is clear that the inner scale should be substantially larger than the Fresnel scale for individual catastrophes to be prominent. Equations (4.2), (4.3) are valid for all values of β. When $\beta<4$, $r_{int}=r_{ref}$ but when $\beta>4$, r_{int}/r_{ref} is given by equation (2.5.4). Thus, the required value of r_c in (4.2) or (4.3) is slightly less in this case – but only very slightly, since observations indicate $|\beta-4|\leq0.3$ (GN).

Equations (4.2) and (4.3), together with the condition that some diffractive scintillation be present (namely $r_c<r_{ref}$), imply that caustics can only cause a strong over-brightening (i.e. $\alpha\gtrsim3$) for a restricted range of cut-off scales and observing frequencies. As shown in Fig. 7, there is in fact a minimum wavelength below which (for a given α, distance and scattering strength) caustics will not be observed; since at this wavelength the cut-off scale will be roughly the same as the

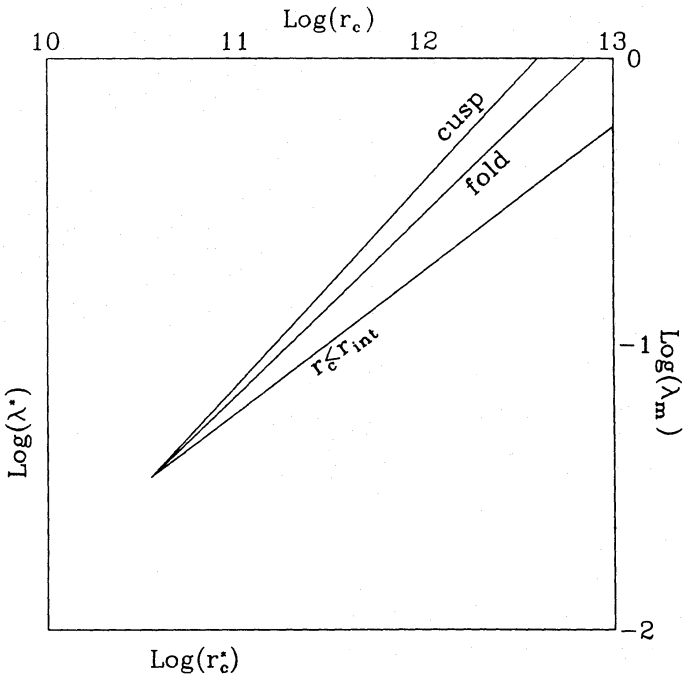

Figure 7. For a given over-brightening α, strong caustics can be found only for a restricted range of cut-off scale and observation wavelength. In particular, for any r_c, λ must be greater than λ_{min}, shown by the line $r_c<r_{int}$ (see Section 4). The maximum wavelengths are denoted by the lines 'fold' and 'cusp'; together these lines give the minimum wavelength λ^* and cut-off scale r_c^*, here for $\alpha=1$. The top and right-hand axes give numerical values for r_c in cm, λ_m in m, and $C_{-4}=D_{kpc}=\alpha=1$.

scattering disc, there is an associated minimum inner scale. For $\beta\sim4$, these are

$$\lambda^*\sim0.03(C_{-4}D_{kpc}^2)^{-1/3}\times\begin{cases}\alpha\ \text{(fold)}\\ \alpha^{2/3}\ \text{(cusp)}\end{cases}\ \text{m}$$

$$r_c^*\sim3.5\times10^{10}(C_{-4}D_{kpc})^{-1/6}\times\begin{cases}\alpha^2\ \text{(fold)}\\ \alpha^{4/3}\ \text{(cusp)}\end{cases}\ \text{cm}. \qquad (4.4)$$

For longer observing wavelengths and larger cut-off scales, there will be a region over which caustics of a given strength may be observed, shown in Fig. 7 as the regions between $\lambda_{min}=(r_c/r^*)^{1/2}\lambda^*$ and $\lambda_{max}=(r_c/r_c^*)^{7/11}\lambda^*$ (fold), $(r_c/r_c^*)^{5/7}\lambda^*$ (cusp).

Coles *et al.* have analysed the scintillation data for a number of pulsars and have estimated bounds on an inner scale, based on purely 'refractive' modulation indices and the prevalence of scintillation over a range of observation frequencies. They conclude, on quite conservative grounds, that the average r_c for the various lines-of-sight can be $\gtrsim10^{11}$ cm, although a very large scatter makes estimates difficult. This is substantially smaller than our fiducial value, but, even for such r_c, caustic behaviour may be significant for certain lines-of-sight and observing frequencies, though they will not, in general, dominate the fluctuation statistics. We therefore concentrate on cut-offs of order several times 10^{12} cm.

If the distance to the screen is greater than the focal length for perturbations on the scale r_c then each r_c^2 scale patch should produce, on the average, one caustic. This caustic will form a fold loop in the observer plane of length $2\pi r_{int}$ with, in general, of order one associated cusp. Thus an observer moving across this plane will encounter $\sim r_{int}/r_c^2$ folds cm^{-1}. If this motion is due to a typical pulsar velocity of $10^7 v_7$ cm s^{-1} and we use the standard values for the ISM scales given

above, then the mean time between fold crossings is

$$\sim 2(r_c/7\times10^{12}\,\mathrm{cm})^2(r_{\mathrm{int}}/3\times10^{13}\,\mathrm{cm})^{-1}v_7^{-1}\,\mathrm{day}. \tag{4.5}$$

Taking the size of a cusp to be that of the diffraction-broadened maximum (of width $r_{\mathrm{int}}/\phi_c^{3/4}$, equation B9), we can estimate the frequency of cusp encounters to be

$$\frac{r_{\mathrm{int}}[r_F^2/(r_c r_{\mathrm{int}})]^{3/4}}{r_c^2}\,v\ \mathrm{s}^{-1}\sim\left(\frac{r_{\mathrm{int}}r_F^6}{r_c^7}\right)\frac{v}{r_c}\ \mathrm{s}^{-1} \tag{4.6}$$

where we have, on the average, one cusp per r_c^2. With our standard values, this gives ~4 yr between cusp encounters. If cusps become fairly bright ($\alpha\gtrsim3$), then it should be possible to detect several diffraction fringes in the X direction.

While it is the second and higher local derivatives of phase on the screen which determines the sub-image intensity, to detect a given area on the screen as a sub-image there is also the condition that the phase gradient have the appropriate value to steer the rays towards the observer. For $\beta<4$, this typical bending is the mean scattering angle, θ_{rms}, meaning that the mean position shift of the observer from the spot directly below the caustic is $\theta_{\mathrm{rms}}D=r_{\mathrm{ref}}$. [For $\beta>4$, the bending could in principle be $\gg\theta_{\mathrm{rms}}$ if the outer scales were sufficiently large, but dynamic scintillation spectra place important limits on r_{out} in this case (GN).] Since $\theta_{\mathrm{rms}}\propto\lambda^2$, this shift is a rather strong function of the observation frequency. This has two consequences. First, in order to resolve a small-scale structure, such as the peak of a fold $\Delta X\sim r_F^{4/3}r_{\mathrm{ref}}^{1/3}r_c^{-2/3}$, one must observe over a narrow band to avoid image blurring from chromatic aberration. Since $\Delta X_{\mathrm{rms}}/r_{\mathrm{ref}}\sim2\Delta\lambda/\lambda$ this condition is (for $\beta\sim4$)

$$\frac{\Delta\lambda}{\lambda}>\frac{r_F^{4/3}}{r_{\mathrm{ref}}^{2/3}r_c^{2/3}} \tag{4.7}$$

With the numbers above and $r_c=8\times10^{12}$ cm this is $\Delta\lambda/\lambda>3\times10^{-3}$. Since a typical decorrelation bandwidth for diffractive scintillation is $\sim50C_{-4}^{-1}\lambda_m^{-4}D_{\mathrm{kpc}}^{-2}$ kHz (RNB), i.e. $\Delta\lambda/\lambda\sim10^{-4}$, it should be possible to filter out the scintillations (and improve the signal-to-noise ratio) without blurring the caustics. The second consequence of the frequency dependence is that one may use the motion of the caustics to scan the observer plane. To find caustics, one might wish to observe for $t_{\mathrm{ref}}\sim r_{\mathrm{ref}}/v\sim30$ day at ~meter wavelengths, monitoring continuously at a single frequency. Alternatively, one could observe 10 times at three-day intervals, each time scanning a range $\Delta\lambda/\lambda\sim0.05$. With our standard numbers, equation (4.5) predicts ~10 folds should be found by either of these procedures.

This frequency dependence coupled with the fall-off of the caustic's flux as a function of distance [equation (3.3.11) for the fold, equation (B10) for the cusp] suggests that the best way to search for caustics may be via matched filters in λ, t space. For folds, for example, one would monitor for both image creation events $I\sim t^{-1/2}$ and image destruction events $I\sim(-t)^{-1/2}$. Light curves for the fold and the cusp are shown schematically in Fig. 8. The flux maxima should lie along slopes in the λ, t plane (like dynamic scintillation spectra) with $\Delta\lambda/\Delta t\sim\lambda v/r_{\mathrm{ref}}$.

Existing time series of compact, refractively scintillating radio sources show examples of fluctuations that can be interpreted as caustics. The long-term pulsar intensity records of Cole *et al.* (1970) and Helfand *et al.* (1977) show several fluctuations apparently much larger than the rms deviation which might be ascribable to caustic events. A particular example of compact extragalactic radio source fluctuations apparently due to strong refractive focusing can be seen in the striking events observed for three radio quasars by Fiedler *et al.* (1987). The quasar light curves of these events show two resolved, but sharply peaked maxima bracketing a period of suppressed flux, and in one instance four strong intensity spikes at a second frequency. The observations

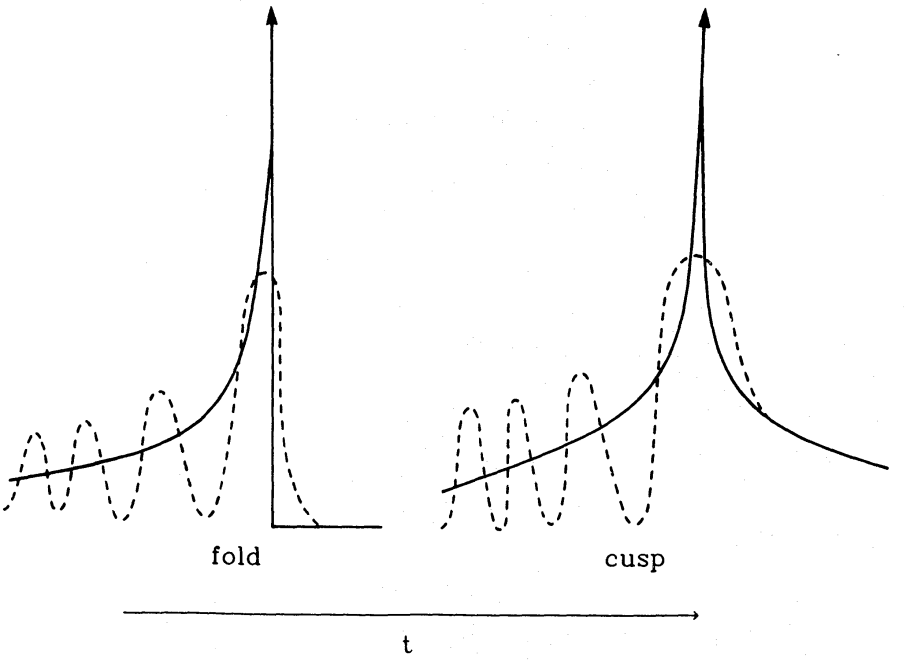

Figure 8. Light curve of a point source near fold and cusp catastrophes as a function of the single control parameter, t. The solid line depicts the geometrical optics variation of the mean flux: $|t|^{-1/2}$ for the fold, $|t|^{-1/3}$ for the cusp. The dashed lines depict the fringe behaviour for finite wavelength observations; fringes are found on the bright side of the fold while the flux drops exponentially to zero after the two sub-images annihilate. For the cusp, fringes can only be seen by observers within the apex of the cusp. In the specific example shown, as the observer passes outside the caustic near the cusp point, two sub-images annihilate and the remaining sub-image fades smoothly, generating no fringes.

suggest an interstellar propagation effect and, with typical galactic distances and velocities, a focusing perturbation of size $\sim 10^{13}$–10^{14} cm. This focusing can be interpreted as caustic events caused by excess power at large scales [option (iii) in the Introduction] in the form of a single ~Gaussian profile scatterer producing two sets of two fold catastrophes, viewed from a distance of several focal lengths (Romani *et al.* 1987). Since the density perturbation δn_e inferred by Fiedler *et al.* is rather large, the single high-density lensing perturbation should be in a special high-pressure region (e.g. an old supernova remnant) or be magnetically confined. The condition that a single cloud of size $\sim r_{ref}$ dominate the accumulated small-scale scattering along the line-of-sight can be estimated from equation (4.1) to be

$$n_e \gtrsim 20 C_{-4}^{1/2} D_{kpc}^{1/2} \text{ cm}^{-3} \tag{4.8}$$

for $\beta \sim 4$ (lower electron densities are adequate if the density perturbation takes the form of a filament or a sheet). These observations argue strongly for the presence of 10^{13}–10^{14} cm-scale, high-density focusing plasma structures which can occasionally dominate the distributed scattering of the general ISM. For this component, the statistics will be non-Gaussian and a power-law description will not be adequate.

5 Dynamic scintillation spectra

The patterns of flux maxima in the (ν, t) plane known as dynamic scintillation spectra have been recognized as important probes of scales larger than r_{dif} for some time (e.g. Hewish 1980). The presence of systematic patterns and periodicities imposed on the mean drift expected for the effect of a random collection of larger scales was clearly described by Roberts & Ables (1982) and

has been recently studied further by Cordes *et al.* (1987). Pulsars are often found to exhibit frequency drifts with bands modulated on a scale several times smaller than the typical band spacing. Occasionally, two or even more periodicities are seen, forming sets of periodic flux modulations at different angles to the mean drift (crossed, or herringbone patterns) as portrayed in Fig. 9. One interpretation is that these are the interference patterns formed when there are two (simple periodicities) or three (crossed patterns) bundles of rays interfering at the observer. When there is a single thin screen, these ray bundles must arise from different directions and so the image of the source must be subdivided into a small (~2–3) number of bright sub-images. In general the interference pattern on the ground is a measure of the separation of the sub-images on the sky, although Cordes *et al.* (1986) have noted that, for an extended medium, the correspondence may not be direct. In fact, Cordes & Wolszczan (1986) find the number of coherent fringes to be quite high, typically indicating a sub-image spacing 3–5 times the sub-image size.

The situation would seem difficult to realize in an extended $\beta<4$ spectrum since the rms scattering angle is dominated by perturbations on a scale $r_{\rm dif}$; large scales of the order of the sub-image separation would not have enough power to divert the sub-images by an angle of order their own angular size. Conversely, an extended $\beta>4$ power law would have difficulty producing such well-separated spots because intermediate scales would have sufficient power to break up the sub-images. Indeed, as described in GN, the fractal dimension of the image is $6-\beta<2$. Hence one expects a hierarchy of scales within the image, and the separation of patches on each scale should be less than or of order the patch size. In addition, such a spectrum must, of course, be truncated at long wavelengths to avoid arbitrarily large drift slopes and non-linear density perturbations.

To account for the form of the periodicities, it again seems necessary to postulate either (i) a cut-off at small scales or (ii) extra power at a large scale. While in some cases the latter picture seems most applicable, at high frequencies, with $r_{\rm F}\sim r_{\rm c}$, the inner scale effects can dominate.

If there is an inner scale $r_{\rm c}\sim r_{\rm ref}$ then there should be only a few sub-images per image spot. In this case the interference patterns could arise from the members of a single caustic; a merging pair

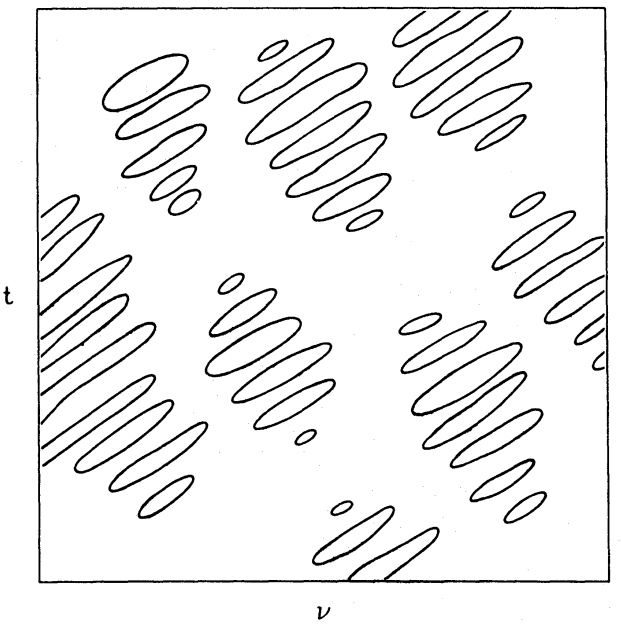

Figure 9. Crossed patterns in dynamic scintillation spectra generated by interfering sub-images. Here we show schematically in the ν, t plane maxima of flux lying on striated bands caused by the interference of two bright sub-images, which drift in ν due to the varying refraction. Nearly orthogonal, finer scale fringes are superposed due to the regular interference pattern of a more distant third sub-image.

(fold) for the case of single periodicities, or a merging triple (cusp) for herringbone patterns. Since r_c is a specific scale in the ISM this will only work for a limited range of frequencies for which the total image size is close to but slightly larger than this scale. Such a picture may be too restricted to account for the apparently common occurrence of such patterns.

A smaller inner scale is possible if each of the bright sub-images contributing to the fringe pattern is itself a fold caustic. In this model, the *mean* number of sub-images is $\gg 1$, but a few are highly-magnified pairs. The typical separation between two pairs is $\sim r_{\rm int}$, since the pairs are independent. Also, since the minimal separation between the components of a merging pair is $\sim r_{\rm int}/\phi_c^{2/3} \sim r_{\rm int}(r_F^2/r_c r_{\rm int})^{2/3} \sim 3\times 10^{-3} r_{\rm int}$ for the standard numbers above, the number of coherent fringes is potentially very large. Let us consider the case of a herringbone pattern produced by the interference of three folds. The fringes will be obvious and striking if the peak flux of the interference pattern is twice the mean flux from the rest of the image. Let the mean number of normal sub-images be N, which is $\sim(r_{\rm int}/r_c)^2$ according to equation (3.2.18), and let the flux of each of the three folds be F_c, with an average total flux from all sub-images normalized to unity. Then, since the amplitudes add coherently in the fringes, we require $3^2 F_c > 2(N-3)/N$, i.e. $F_c \geqslant 0.2$. Now it follows from equations (3.3.11), (3.3.9) and (3.3.6) that for a large number of sub-images

$$F_c(X) \sim \left(\frac{r_{\rm int}}{X}\right)^{1/2} \bar{F}_s, \tag{5.1}$$

where X is the distance to the fold line in the observer's plane, and $\bar{F}_s \sim 1/N$ is the mean sub-image flux. If the interference pattern is to be frequently visible, X must be no larger than the average distance to the nearest fold line, $r_{\rm int}/N$ [*cf.* the derivation of equation (4.5)]. Hence $N \lesssim 25$. Putting $X = r_{\rm int}/N$ in equation (3.3.7), we find that the projected separation between the components of the pair is $\sim r_c/\sqrt{N} \sim r_{\rm int}/N$. So $r_c \sim r_{\rm int}/5 \sim 6\times 10^{12}$ cm for our standard numbers and the angular separation between bright folds should be of order 25 times the 'size' of each fold.

In this picture one expects patterns only for observing frequencies and screen distances such that there are the appropriate number N of sub-images within the scattering disc. Moreover, the mean flux of the pulsar should be brighter when drifting patterns are observed, typically 20 per cent above normal for periodic patterns and 40 per cent above normal for herringbone patterns. Finally, although patterns can be observed over typically an octave in observation frequency, the form of any individual pattern should be correlated only over a limited range of t and ν. In fact a positional shift of $\sim r_{\rm int}/N$ (the distance to the nearest fold line) should destroy the arrangement of the caustics; this corresponds to $t \sim 1\, v_7^{-1}$ day or from the frequency dependence above $\Delta\lambda/\lambda \sim 10^{-2}$. Thus herringbone patterns should rearrange from day to day. We note that Hewish *et al.* (1985) found that dynamic scintillation spectra not only rearranged on ~day time-scales, but that the coherence bandwidth of the periodicities was quite small ($\Delta\lambda/\lambda \lesssim 10^{-3}$ for the case of PSR 1642−03).

Cordes & Wolszczan (1986) have proposed that the subdivision of the image is effected by extra power superposed on a Kolmogorov spectrum at a large scale. The size $r_{\rm int}$ of each sub-image is determined by the $\beta = 11/3$ scattering while the extra scale at $\sim 5 r_{\rm int}$ has sufficient power to steer images separated by this scale to a common focus. As in the multiple caustic picture, periodic drifts should only occur for a special octave of observing frequency for which there are a few well-separated images. However, here one expects the patterns to be much more persistent in both ν and t. The patterns should be similar over an octave in ν while the variation time-scale should be several times the refractive time-scale of order 30 day. Observationally, since drifting patterns can change quite rapidly ($\lesssim 1$ day) but then may persist for months, it is not clear whether such slow variation of the rapid caustic behaviour is indicated. The mean flux of the image in the extra scale picture also varies differently from that in the multiple caustic model. Here all of the flux

comes from a few sub-images, so when an interference pattern is set up the increase over the mean flux level should be of order unity. In addition, there should be caustics associated with the large extra scale fluctuations, albeit degraded by the Kolmogorov fluctuations. In this model, when sub-images approach one another the spacing of the periodic interference pattern increases and the sub-images brighten as they approach the fuzzy, large-scale caustic. In contrast, for the multiple caustic picture, merger events are independent and so fringe spacing and overall flux are not correlated in this fashion. Again, for a complete description of the observed dynamic spectra it may be necessary to invoke the existence of suitably-sized high-density clouds as deterministic scatterers.

6 Conclusions

In summary, if there is an inner scale in the density perturbation spectrum of the ISM comparable to the size of the scattering disc of pulsars and compact extragalactic radio sources, then caustics can have a dramatic effect on the flux modulations and individual folds and cusps may even be resolvable. Since the size of the scattering disc $r_{\rm ref} \propto \lambda^2$, this condition can only be satisfied over a narrow range of frequencies for an individual pulsar. Nevertheless, despite these restrictive conditions, there are indications from the observations of non-Gaussian spikes in the light curves and quasi-periodic patterns in the dynamic scintillation spectra that caustics are, in some cases, important. It is likely that a more complex model than a truncated power-law density fluctuation spectrum in a thin screen will be necessary to reconcile theory and observation. A possibly relevant modification would involve non-power-law density fluctuations; i.e. a highly non-Gaussian distribution of density perturbations. In particular, if the scattering is relatively localized, a few large clouds, or even a single one, with an enhanced density fluctuation can dominate the scattering. Such enhanced scattering from clumps of scale $\gtrsim r_{\rm int}$ might be found in H II regions, supernova remnants, magnetically confined plasma clouds, etc. Numerical simulation may provide the best means of studying such localized scatterers. Such simulations could probe the effects of relaxing the thin screen assumption and the statistics of higher-order catastrophes, as well. (It can be difficult to realize a *sharp* cut-off in the angular spectrum of density perturbations from a specific physical scale in an extended medium; caustics should be correspondingly weakened.) The computation of higher flux moments and comparison with the behaviour of pulsar flux records would also be useful in determining the importance of higher order caustic events.

From an observational standpoint, it will be important to determine if large intensity spikes and periodic dynamic scintillation spectra occur often over a restricted range of frequencies for each specific pulsar as suggested above, or if such phenomena are less common, but present over a wide range of observation frequency for most pulsars. The latter case would require a more complicated model of the spectrum of ISM density perturbations than a truncated power law. Further, as described by Cordes *et al.* (1987), dynamic scintillation spectra are important tools for determining when pulsars are subject to multiple imaging. Sources displaying strong, varying periodicities are clearly good candidates for bright caustics. Monitoring of such sources over a range of time lags would allow determination of the variation time-scale for the drifting bands and detection of the flux correlation with drift parameters described above. Data from Cordes & Wolszczan (1986) are useful as a first attempt to quantify these phenomena; such data show that in a number of cases dramatic periodicities are present, providing an important arena for testing the caustic model.

Acknowledgments

We thank M. Berry for several helpful discussions. Support for this work was provided by the

National Science Foundation under grants AST 84-15355 (Caltech), AST 86-11121 (U. Arizona), and PHY 82-17352(IAS). We are also happy to acknowledge support for a W. M. Keck Foundation Fellowship (JJG) and from a Fannie and John Hertz Foundation Fellowship (RWR).

References

Armstrong, J. W., 1984. *Nature,* **307,** 527.
Armstrong, J. W., Cordes, J. M. & Rickett, B. J., 1981. *Nature,* **291,** 561.
Arnol'd, V. I., 1975. *Russ. math. Survs.*, **30,** 1.
Berry, M. V., 1977. *J. Phys. A: Math. Gen.*, **10,** 2061.
Berry, M. V. & Upstill, C., 1980. *Progress in Optics,* vol. 18, p. 257, ed. E. Wolf, North Holland, Amsterdam.
Blandford, R. & Narayan, R., 1985. *Mon. Not. R. astr. Soc.*, **213,** 591 (BN).
Blandford, R., Narayan, R. & Romani, R. W., 1984. *J. Astrophys. Astr.*, **5,** 369.
Born, M. & Wolf, E., 1980. *Principle of Optics,* Pergamon Press, New York.
Cole, T. W., Hesse, H. K. & Page, C. G., 1970. *Nature,* **225,** 712.
Coles, W. A., Frehlich, R. G., Rickett, B. J. & Codona, J. L., 1987. *Astrophys. J.,* submitted.
Cordes, J. M. & Wolszczan, A., 1986. NAIC preprint 219.
Cordes, J. M., Pidwerbetsky, A. & Lovelace, R. V. E., 1987. *Astrophys. J.,* in press.
Fiedler, R. L., Dennison, B., Johnston, K. J. & Hewish, A., 1987. *Nature,* submitted.
Frehlich, R. G., 1986. Preprint.
Gapper, G. R. & Hewish, A., 1981. *Mon. Not. R. astr. Soc.*, **197,** 209.
Goodman, J. & Narayan, R., 1985. *Mon. Not. R. astr. Soc.*, **214,** 519 (GN).
Hecht, E. & Zajac, A., 1974. *Optics,* Addison-Wesley, Reading, Massachusetts.
Helfand, D. J., Fowler, L. A. & Kuhlman, J. V., 1977. *Astr. J.*, **82,** 701.
Hewish, A., 1980. *Mon. Not. R. astr. Soc.*, **192,** 799.
Hewish, A., Wolszczan, A. & Graham, D. A., 1985. *Mon. Not. R. astr. Soc.*, **213,** 167.
Lovelace, R. V. E., 1970. *PhD thesis,* Cornell University.
Poston, T. & Stewart, I. N., 1982. *Catastrophe Theory and its Applications,* Pitman, London.
Rickett, B. J., 1977. *Ann. Rev. Astr. Astrophys.*, **15,** 479.
Rickett, B. J., Coles, W. A. & Bourgois, G., 1984. *Astr. Astrophys.*, **134,** 390.
Roberts, J. A. & Ables, J. G., 1982. *Mon. Not. R. astr. Soc.*, **201,** 1119.
Romani, R. W., Blandford, R. & Cordes, J., 1987. *Nature,* submitted.
Romani, R. W., Narayan, R. & Blandford, R., 1986. *Mon. Not. R. astr. Soc.*, **220,** 19 (RNB).
Rumsey, V. H., 1975. *Radio Sci.*, **10,** 107.
Salpeter, E. E., 1967. *Astrophys. J.*, **147,** 433.
Shapirovskaya, N. Ya., 1978. *Soviet Astr.*, **22,** 544.
Shishov, V. I., 1974. *Soviet. Astro.*, **17,** 598.
Sieber, W., 1982. *Astr. Astrophys.*, **113,** 311.
Tatarskii, V. I. & Zavorotnyi, V. U., 1980. *Prog. Opt.*, **XVIII,** 204.
Thom, R., 1975. *Structural Stability and Morphogenesis,* Benjamin, Reading, Massachusetts.

Appendix A

We consider a phase structure function of the form (2.4.5), $D(r) \propto (|r^2-r_c^2|^{\alpha/2}-r_c^\alpha)$, $\alpha=\beta-2$. The limiting forms of this expression are

$$D(r) \propto |r^2|^{\alpha/2} \qquad r \gg r_c$$

$$D(r) \propto r_c^{\alpha-2} r^2 \qquad r \ll r_c \tag{A1}$$

as appropriate for a structure function $\langle[\phi(r)-\phi(0)]^2\rangle$ calculated from equation (2.3.2) and a fluctuation spectrum with a Gaussian cut-off $Q_0 q^{-\beta} \exp[-(qs_0)^2]$, $\beta<4$. This choice of $D(r)$ yields

$$\tilde{W}(\mathbf{q}) = \int d^2\mathbf{r} \exp(i\mathbf{q}\cdot\mathbf{r}) \exp\{Ar_F^\alpha[2|(qr_F^2)^2+r_c^2|^{\alpha/2}+2|r^2+r_c^2|^{\alpha/2} - |(r+qr_F^2)^2+r_c^2|^{\alpha/2}-|(r-qr_F^2)^2+r_c^2|^{\alpha/2}-2r_c^\alpha]\} \tag{A2}$$

where A is (see GN)

$$A=\frac{2^{2-\beta}}{\pi}\frac{\Gamma[(6-\beta)/2]}{(\beta-4)(\beta-2)\Gamma(\beta/2)}[Q_0 r_F^{\beta-2}]. \tag{A3}$$

Note that, for $2<\beta<4$, $A<0$ and that the quantity in brackets is essentially the amplitude of the phase fluctuations on the Fresnel length. As described in Section 2, when $r_c>r_{dif}$, A must be replaced by $A(r_F/r_c)^{\beta-4}$ for $\beta<4$ to maintain the strength of the scattering.

For strong scattering and fair separation of the various length scales, we have three conditions

$$|A|\gg 1 \qquad \text{i.e. } \phi\gg 1, \tag{A4}$$

$$|A|(r_c/r_F)^{\alpha}\gg 1 \qquad \text{i.e. } r_c\gg r_{dif}, \tag{A5}$$

$$|A|(r_c/r_F)^{\alpha-4}\gg 1 \qquad \text{i.e. } r_c\ll r_{ref}. \tag{A6}$$

Under these assumptions, we wish to investigate the form of $\tilde{W}$ (A2) in the regime $q_{int}\ll q\ll q_{cdif}=r_c/r_F^2$. The integral will be dominated by $|r|\ll r_c$ so we can expand the second exponential to lowest order in r, giving

$$\tilde{W}(q)=\int d^2r \exp(iq\cdot r)\exp[-\alpha(2-\alpha)|A|(r_c/r_F)^{\alpha-4}q^2r^2]. \tag{A7}$$

The exponent in square brackets will cut off the integral when $r\gtrsim[|A|(r_c/r_F)^{\alpha-4}]^{-1/2}q^{-1}$, which, by equation (A6), gives $r\ll q^{-1}$. Accordingly the first exponential can be replaced by one and the Gaussian integration carried out, yielding

$$\tilde{W}(q)=\frac{\pi}{2^{1/2}\alpha(2-\alpha)|A|(r_c/r_F)^{\alpha-4}}q^{-2} \tag{A8}$$

which gives the constant power at intermediate scales in Figs 2 and 3.

Similar analyses can be conducted for the other regimes, giving

$$r_F^{-2}\tilde{W}(q)=\begin{cases}2^{\alpha}\pi\alpha(2-\alpha)(\Gamma[(\alpha+2)/2]/\Gamma[(4-\alpha)/2])|A|(qr_F)^{2-\alpha}\exp[-(q/q_{ref})^2/2] & q<q_{int}\\ \pi/[\alpha|A|(r_c/r_F)^{\alpha-2}]\exp[-(q/q_{dif})^2/2] & q\gg q_{cdif}\end{cases} \tag{A9}$$

where $q_{dif}\equiv q_{ref}^{-1}r_F^{-2}\equiv[2\alpha|A|(r_c/r_F)^{\alpha-2}]^{1/2}$. The scalings for the extended power-law spectra (solid lines in Figs 2 and 3) have been given by GN.

Appendix B

In this appendix we show how to relate the specific expression describing the local phase variation from a thin screen observed at position (X, Y) in the vicinity of a cusp to the standard polynomial equation (3.1.3). Again, we must enlarge the standard form to admit two state variables by writing

$$\bar{\zeta}=\tfrac{1}{4}s^4+\tfrac{1}{2}C_2s^2+C_1s+\tfrac{1}{2}t^2. \tag{B1}$$

We follow the same procedure that we described in Section 3.3 for the fold. We allow ourselves to translate and rotate the (x, y) and (X, Y) coordinate systems and only retain the terms that can be dominant in the vicinity of the stationary points. We obtain

$$\zeta(x, y; \alpha, \beta)=\tfrac{1}{4}ax^4+\tfrac{1}{2}bx^2y+\tfrac{1}{2}cy^2-\alpha x-\beta y; \qquad a>0 \tag{B2}$$

where $a=\phi_{xxxx}/6, b=\phi_{xxy}, c=\phi_{yy}. \alpha=kX/z, \beta=kY/z$ are control variables. Again, this expansion is

only valid for $|x|, |y| \ll r_c, r_{int}$. The stationary points of equation (B2) are given by the real roots of the equation

$$\left(a-\frac{b^2}{2c}\right)x_i^3+\frac{b\beta}{c}x_i-\alpha=0 \tag{B3}$$

together with the associated y coordinate

$$y_i=-\frac{(bx_i^2/2-\beta)}{c}. \tag{B4}$$

There are three roots corresponding to three rays when

$$f(\alpha,\beta)=\left(\frac{\alpha}{2}\right)^2\left(a-\frac{b^2}{2c}\right)+\left(\frac{b\beta}{3c}\right)^3<0, \tag{B5}$$

and one ray when the inequality is reversed. $f(\alpha,\beta)=0$ is the equation of the caustic in the observer plane. The image of the caustic on the screen, i.e. the location of the coalescing images, is the parabola

$$H(x,y)=bcy+(3ac-b^2)x^2=0. \tag{B6}$$

$H(x,y)$, the Hessian, is the reciprocal of the sub-image amplification.

The procedure that we follow in relating the specific expression for the phase variation, equation (B2), to the standard form, equation (B1), is to assume that there exists the phase relation $\zeta=\bar{\zeta}+g(\alpha,\beta)$ where $g(\alpha,\beta)$ is a function to be determined. The three functions $C_1(a,b,c;\alpha,\beta)$, $C_2(a,b,c;\alpha,\beta)$ can be determined by comparing the phases at the three stationary points (a one-to-one mapping is guaranteed) and solve for $s(x,y)$, $t(x,y)$. We obtain

$$\begin{aligned}
C_1&=\alpha(a-b^2/2c)^{-1/4}\\
C_2&=\beta bc^{-1}(a-b^2/2c)^{-1/2}\\
g&=-\beta^2/2c\\
s&=-(a-b^2/2c)^{1/4}x\\
t&=c^{1/2}(y-\beta/c+bx^2/2c).
\end{aligned} \tag{B7}$$

Equations (B7) allow us to map a cusp created by a thin phase screen on to the standard polynomial equation. They therefore can be used to compute the diffraction pattern from an individual cusp, given the low-order derivatives of the phase ϕ at the cusp.

The intensity behaviour and fringe scales can be estimated for a cusp, following Section (3.3), although the complete description of the fringes given by the Pearcey function (e.g. Berry & Upstil 1980) is rather complex. At the cusp, $x=y=0$, we have from equation (B2) the area of the first coherent zone is $\sim(a^{-1/4})\times(c^{-1/2})\sim r_c^2/\phi_c^{3/4}$ and so, in analogy to equation (3.3.6), the maximum flux is

$$F_{c,\max}\sim\frac{r_c^4}{\phi_c^{3/2}r_F^4}\sim\phi_c^{1/2}\bar{F}_s\sim\phi_c^{1/2}/N. \tag{B8}$$

From (B3), the sub-image separation for an observer at $Y=0$, $X\neq 0$ is $\delta x\approx\pm(X/r_{int})^{1/3}r_c$. Thus the x-width of the central fringe of a cusp caustic, given by the distance over which the phase varies by order unity is

$$X_c\sim r_{int}/\phi_c^{3/4}. \tag{B9}$$

From equations (3.2.2), (3.2.3) and (B2) we can estimate the variation in the *mean* flux with distance X from the cusp point as

$$F_c(X)\sim\left(\frac{\partial^2\zeta}{\partial x^2}\right)^{-1}\left(\frac{\partial^2\zeta}{\partial y^2}\right)^{-1} r_F^{-4}\sim\left(\frac{\phi_c(X/r_{int})^{2/3}}{r_c^2}\right)^{-1}\left(\frac{\phi_c}{r_c^2}\right)^{-1} r_F^{-4}\sim\left(\frac{X_f}{X}\right)^{2/3} F_{c,\,max}. \tag{B10}$$

An approximate scale for the decrease in fringe spacing within this envelope for observer shifts in the X direction, following equation (3.3.12) is

$$S_c(X)\sim(d\Delta\zeta/dX)^{-1}\sim\left(\frac{X_c}{X}\right)^{1/3} X_c. \tag{B11}$$

There should be no coherent fringes in the Y direction beyond the central maximum of width $\sim r_F^2/l_y$ (3.2.2).