Caustics of weak shock waves

Rodolfo R. Rosales
Department of Mathematics Room 2-337, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Esteban G. Tabak
Courant Institute of Mathematics, New York University, 251 Mercer Street, New York, New York 10012

(Received 30 April 1997; accepted 29 August 1997)

The caustics of weak shock waves are studied through matched asymptotic expansions. It is shown that these caustics are thinner and more intense than those of smooth waves with a comparable amplitude. This difference in scalings solves a paradox that would have the caustics of weak shock waves behave linearly, even though linear theory for discontinuous fronts predicts infinite amplitudes near the caustic and in the reflected wave. With the new scalings, the behavior of shocks both near the caustic and in the far field is described by nonlinear equations. The new scales are robust, in the sense that they survive the addition of a small amount of viscosity to the equations. As a viscous shock approaches the caustic, its intensity amplifies and its width decreases in such a way that the new scalings are actually reinforced. A new paradox arises, however: The nonlinear Tricomi equation which describes the behavior of the fronts near caustics does not appear to admit the triple shock intersections which have been observed experimentally. This new open problem is closely related to the von Neumann paradox of oblique shock reflection. © *1998 American Institute of Physics.* [S1070-6631(97)03912-3]

I. INTRODUCTION

High frequency waves can be described in the language of geometrical optics, with fronts propagating along one-dimensional rays and energy being conserved along ray tubes. However, this description fails in the neighborhood of caustics, the envelopes of the rays, where the effects of the variations along fronts and along rays become comparable. Thus a two-dimensional theory is required near caustics, a situation similar to the one taking place near the singular rays of diffraction theory. A good description of the caustics is particularly important, since there the wave intensity amplifies significantly, due to the collapse of infinitesimal ray-tubes into points. Thus the caustics are layers with strong energy concentration, often requiring careful control.

The caustics of linear waves have been well understood for some time, since the pioneering work of Buchal and Keller[1] and Ludwig.[2] Their analysis involves a local multiple-scale expansion near the caustics, where not only the fast scale transversal to the fronts, but also an intermediate scale along the fronts contribute to the leading order behavior of the waves. This analysis has been extended by Hunter and Keller to the caustics of smooth weakly nonlinear waves.[3] Their surprising result is that even waves strong enough to require a weakly nonlinear treatment away from the caustics behave linearly at the caustics. The reason is that, even though the waves amplify near the caustics, they spend too little time there for nonlinear effects to accumulate.

This result leads to contradiction, however, if naively applied to the caustics of weak shock waves. On the one hand, these caustics have been observed by Sturtevant and Kulkarny to behave nonlinearly even for waves of very small amplitude.[4] The main manifestations of this nonlinearity are the occurrence of triple shock intersections and the displacement of the locus of the caustic from its linear position. On the other hand, the linear theory of caustics involves a Hilbert transform which, if applied to a discontinuous incident profile, yields a logarithmic singularity across the reflected wavefront. Such singularities concentrated along lines are incompatible with the hypothesis behind linear theory.

In this paper, we develop a way out of this contradictory situation. We find that the caustics of weak shock waves are thinner and more intense than those of smooth waves of comparable amplitude, and they behave nonlinearly. These results may have important practical implications for the growing number of devises which utilize the amplification of focusing sound waves for medical and engineering purposes.

The plan of the paper is the following. In section II, we review the linear theory of caustics. We introduce a number of coordinate systems useful for the description of waves near a caustic, develop the caustic expansion for linear isentropic gas dynamics, and match it with the outer expansion of linear geometrical optics. Out of this matching, a Hilbert transform appears which relates the waves before focusing to those past the caustic, denoted from now on as incident and reflected waves, respectively. Then we apply this theory to discontinuous incident fronts, and show that it predicts reflected waves with infinite amplitude. These asymptotic results are verified with the exact solution corresponding to a circular caustic.

In section III, we describe the weakly nonlinear theory of waves away from focus and near caustics. The resulting equations in the caustic layer are equivalent to a nonlinear Tricomi equation. We then consider matching the inner and outer expansions, and describe the Hunter and Keller's argument, which shows that the caustics of smooth weakly nonlinear waves should behave linearly. This result, however, is inconsistent for weak shock waves.

In section IV, we solve this apparent contradiction. We

 1070-6631/98/10(1)/206/17/$10.00

find that the caustics of weak shock waves are thinner and more intense than those of smooth waves, and behave nonlinearly. A clue to this realization is the fact that shock waves do not have a natural length scale, so matching the inner and outer layers should be based on amplitude and not on wavelength, as is normally done for smooth waves. We study the more general scenario of a shock within a smooth wave, and describe a setting that allows us to combine the two different scalings in a single framework. Then we consider the focusing of viscous shocks. One would naively expect that viscosity would restore the scaling for smooth waves, since it does provide a typical width for the shocks. We show, however, that this is not the case: As a shock wave amplifies near a caustic, its width decreases in such a way that it allows the nonlinear scale of inviscid shocks to take over.

In section V, we briefly discuss an open problem: The nonlinear equations for the caustic layer do not allow triple shock intersections, one of the most prominent nonlinear features of the caustics of weak shock waves. We comment on the relation between this problem and the von Neumann paradox of oblique shock reflection. The elucidation of this paradox is the subject of much present work (see for instance Refs. 5, 6, 7, 8, and 9), and is not pursued any further here. Finally, we put forward some concluding remarks.

II. CAUSTICS OF LINEAR WAVES

In this section, we review the Linear Theory of Caustics, as it applies to high frequency waves. Many of the arguments in the following sections derive from particular features of this theory, so it is important to summarize it here in a unified way. Although we shall attempt to make this section as self-contained and clear as possible, many details will be omitted for the sake of brevity. We refer the interested reader to the original papers where the theory of caustics was established in the framework of Linear Geometrical Optics, particularly Refs. 1, 2, and the review article,[10] which has an extensive list of references.

The plan of this section is the following. After some general remarks, we introduce in subsection II A three coordinate systems useful for the description of waves near caustics: Two Caustic Polar Coordinate Systems, associated with the incident and reflected waves, and a Normal Caustic Coordinate System which serves as a unifying framework for both. In this latter system, we propose a coordinate stretching which magnifies the vicinity of the caustic. In subsection II B, we develop the caustic expansion for the equations of linear isentropic gas dynamics, which we proceed to match with the outer Linear Geometrical Optics expansion in subsection II C. In subsection II D, we discuss the singular behavior of discontinuous fronts near caustics predicted by linear theory, a behavior which is confirmed in subsection II E by the exact solution to the linear equations corresponding to a circular caustic.

Consider a Linear Geometrical Optics (LGO) expansion in two space dimensions in an homogeneous isotropic medium. Assume a nondimensionalization such that the sound speed c is one and the radius of curvature of the wave fronts is generically $O(1)$. Since we consider high frequency waves, we let the wavelength (and period) of the waves be $2\pi\epsilon/k$, with $0<\epsilon\ll 1$ and $k=O(1)$ The wave fronts are described by the *phase*,

$$\theta=\frac{\Phi(x,y)-t}{\epsilon}, \tag{1}$$

where the *action variable* $\Phi(x,y)$ satisfies the **Eikonal equation**,

$$(\nabla\Phi)^2=1. \tag{2}$$

This is equivalent to the statement that the wave fronts ($\theta=$const) move normal to themselves at speed 1, along straight lines denoted *rays*.

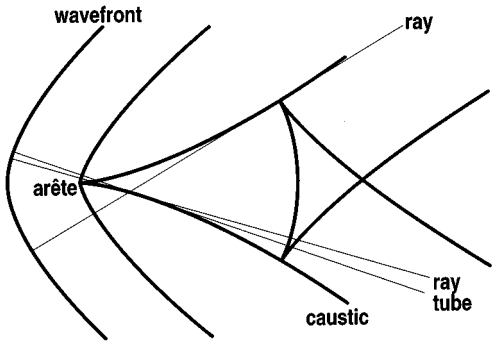

FIG. 1. Typical focusing wave front propagating to the right. The front propagates normal to itself along the rays, at unit speed. The rays starting near the point of maximum curvature cross first, at the arête. Other rays cross later, forming the caustic (envelope of the rays). As the front propagates, it folds on itself along the caustic. The front is shown at three different times: before any focusing occurs, precisely when the first focusing occurs (at the arête), and at a later time, with a fold along each caustic branch. Also displayed is a ray tube collapsing at the caustic.

A typical focusing wave front is shown on the left in Fig. 1. As it moves normal to itself along the rays, it develops folds. The envelopes of rays where this folding occurs are the *caustics*. Typically, caustics begin at an *arête*, corresponding to the point with maximal curvature on the original front (whose neighborhood is the first to focus and fold). The location of the caustics, as well as the detailed behavior of the state variables close to them, are particularly important, since at the caustics the wave intensity amplifies enormously. This occurs because, away from the caustics, the wave energy is conserved along ray tubes. Thus formally, in this LGO *outer expansion*, the energy density becomes infinite when the ray tubes collapse into points at a caustic. In order to determine the behavior near caustics more precisely, an *inner expansion* becomes necessary; we describe such an expansion below. In this work, we shall concentrate on the behavior at caustics far from the arête; the more complicated structure near an arête requires additional care.

A. Caustic coordinate systems

Next we shall develop various systems of coordinates valid near a caustic, which are useful in describing the local behavior of the waves. To this end, consider a smooth convex caustic Γ given parametrically by

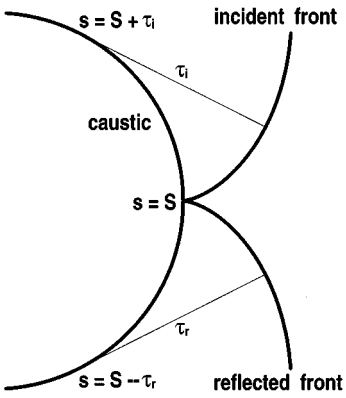

FIG. 2. Parametric description of the incident (before folding) and reflected (after folding) fronts near the caustic. For the incident front, τ_i is the distance along the ray to the caustic, while $S+\tau_i$ is the value of the arclength parameter s on the caustic at the tangency point of the ray. A similar description holds for the reflected front. The figure is for a circular caustic.

$$x=X(s) \quad \text{and} \quad y=Y(s).$$

Here s is the caustic arclength, which can be identified with the value of the action variable Φ along it:

$$\Phi(X(s),Y(s))=s.$$

[Note that the wavefronts—as given by (1) and (2)—are normal to the caustic, where they fold defining an *incident* front and a *reflected* front. The contact point then moves along the caustic at unit speed.] Of course, Φ is defined—and two valued—only on one side of Γ. Let now $\kappa=\kappa(s)$ denote the curvature of Γ, which we shall assume positive without loss of generality. We denote by

$$\hat{\mathbf{t}}(s)=(\dot{X}(s),\dot{Y}(s))^t \quad \text{and} \quad \hat{\mathbf{n}}(s)=(-\dot{Y}(s),\dot{X}(s))^t,$$

the unit vectors respectively tangent and normal to the caustic Γ at s, where the dot indicates a derivative with respect to s. We have chosen them so that the tangent vector points in the direction of propagation and the normal vector points towards the region without waves.

We can now describe the incident wave front $\Phi_i\equiv S$ by the time τ_i it will take each point along the front to reach the caustic, i.e. the distance to the caustic along the normal to the front (see Fig. 2). A similar construction applies to the reflected wave front $\Phi_r\equiv S$. Let $\vec{\mathbf{R}}=\vec{\mathbf{R}}(s)=(X(s),Y(s))^t$ denote a point on the caustic, and $\vec{\mathbf{r}}=(x,y)^t$ a generic point in space. Then we have, for the incident front $\Phi_i\equiv S$, the parametric representation

$$\vec{\mathbf{r}}_i=\vec{\mathbf{R}}(S+\tau_i)-\tau_i\hat{\mathbf{t}}(S+\tau_i). \tag{3}$$

Similarly, for the reflected front $\Phi_r\equiv S$,

$$\vec{\mathbf{r}}_r=\vec{\mathbf{R}}(S-\tau_r)+\tau_r\hat{\mathbf{t}}(S-\tau_r). \tag{4}$$

(Note that we are taking both τ_i and τ_r positive.) This amounts to introducing two **Caustic Polar Coordinate Systems (CPCS)**, both valid on the side of the caustic where Φ is defined (see Fig. 3): an **advanced CPCS** (s_i,τ_i) and a **retarded CPCS** (s_r,τ_r). These systems are the natural

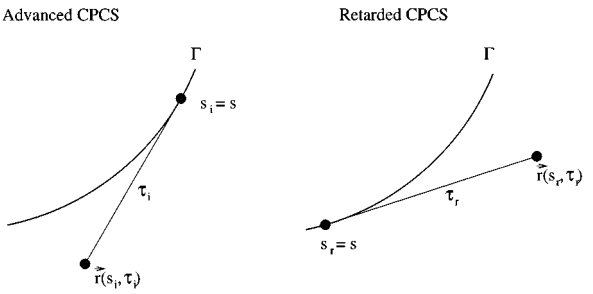

FIG. 3. Advanced and Retarded Caustic Polar Coordinate Systems (CPCS), with coordinates (s_i,τ_i) and (s_r,τ_r), respectively, defined on the wavefront side of the caustic. Both τ_i and τ_r are non-negative. The systems are singular at the caustic $\tau_i=\tau_r=0$, where they fold into each other with a square root type of singularity.

choice for describing the incident and reflected fronts, respectively. Clearly, the following simple formulas apply:

$$\Phi_i=s_i-\tau_i \quad \text{and} \quad \Phi_r=s_r+\tau_r. \tag{5}$$

However, they are not good for describing a neighborhood of the caustic, since there they become singular and fold into each other. Therefore, we introduce a third coordinate system, the **Normal Caustic Coordinate System (NCCS)**, shown in Fig. 4. The coordinates in NCCS are the arclength along the caustic s and the signed distance to the caustic d. Thus

$$\vec{\mathbf{r}}(s,d)=(x,y)^t=\vec{\mathbf{R}}(s)-d\vec{\mathbf{n}}(s). \tag{6}$$

This system does not degenerate at the caustic. The wavefront region is $d>0$ while $d<0$ corresponds to the region in the shade. The incident and reflected wavefronts $\Phi_i\equiv S$ and $\Phi_r\equiv S$ do not have a closed form in this new coordinate system, but they can be expanded in powers of $d^{1/2}$ near the caustic, as follows from the exact expressions in (3)–(5) upon expansion of the coordinate change from the CPCS's to NCCS. The incident front in NCCS is given by

$$s=S+\tfrac{2}{3}\sqrt{2\kappa(S)}d^{3/2}+O(d^2), \tag{7}$$

and the reflected front by

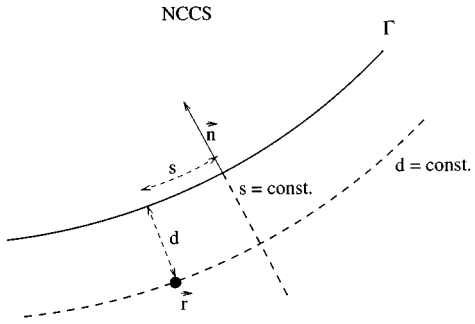

FIG. 4. Normal Caustic Coordinate System (NCCS). The coordinates are s, the distance along the caustic, and d, the signed distance to the caustic, with $d>0$ on the wavefront side of the caustic. This system is nonsingular in a neighborhood of the caustic.

$$s = S - \tfrac{2}{3}\sqrt{2\kappa(S)}\,d^{3/2} + O(d^2). \tag{8}$$

In view of these expansions, it appears natural to introduce two fast variables near the caustic, where Geometrical Optics fails: An *imperfect phase* ψ, to take care of the fast dependence normal to the fronts,

$$\psi = \frac{s-t}{\epsilon}, \tag{9}$$

and a transversal intermediate variable η, to incorporate the fast variation along the fronts near the caustic, where their curvature becomes infinite:

$$\eta = \frac{d}{\epsilon^{2/3}}. \tag{10}$$

Then the phase $\theta = (\Phi - t)/\epsilon$ has the expansion, near the caustic,

$$\theta = \psi \mp \tfrac{2}{3}\sqrt{2\kappa(s)}\,\eta^{3/2} + O(\epsilon^{1/3}\eta^2), \tag{11}$$

where the upper (lower) sign corresponds to the incident (reflected) phase. This expansion for the phase applies when $0 \leqslant \eta \ll \epsilon^{-2/3}$ $(0 \leqslant d \ll 1)$. The higher order terms are uniformly small provided that $0 \leqslant \eta \ll \epsilon^{-1/6}$ $(0 \leqslant d \ll \sqrt{\epsilon})$. On the other hand, it can be shown that the LGO expansion is valid near the caustic for $d \gg \epsilon^{2/3}$ (that is, $\eta \gg 1$). Thus we can use the two term expansion for the phase above as long as $1 \ll \eta \ll \epsilon^{-1/6}$.

B. Example of an expansion near the caustic

The equations we shall consider are

$$\rho_t + u_x + v_y = 0, \quad u_t + \rho_x = 0, \quad v_t + \rho_y = 0, \tag{12}$$

corresponding to **Linear Isentropic Gas Dynamics**, nondimensionalized so that the unperturbed density ρ_0 and the sound speed c are both equal to 1. Transforming these equations into the NCCS introduced above, we obtain

$$\begin{aligned} &(1+\kappa(s)d)\rho_t + \tilde{u}_s - ((1+\kappa(s)d)\tilde{v})_d = 0, \\ &(1+\kappa(s)d)\tilde{u}_t + \rho_s = 0, \\ &\tilde{v}_t - \rho_d = 0, \end{aligned} \tag{13}$$

where $\tilde{u}$ and $\tilde{v}$ are the components of the velocity respectively parallel and normal to the caustic. [Namely, if $\varphi = \varphi(s)$ is the caustic angle: $\dot{X} = \cos\varphi$, $\dot{Y} = \sin\varphi$, so that $\kappa = \dot{\varphi}$; then $\tilde{u} = u\cos\varphi + v\sin\varphi$ and $\tilde{v} = -u\sin\varphi + v\cos\varphi$.]

We expand in terms of the variables ψ and η introduced above, with expansion parameter $\epsilon^{1/3}$:

$$\begin{aligned} &\rho \sim \rho_0(\psi,\eta,s) + \epsilon^{1/3}\rho_1(\psi,\eta,s) + \cdots, \\ &\tilde{u} \sim \tilde{u}_0(\psi,\eta,s) + \epsilon^{1/3}\tilde{u}_1(\psi,\eta,s) + \cdots, \\ &\tilde{v} \sim \tilde{v}_0(\psi,\eta,s) + \epsilon^{1/3}\tilde{v}_1(\psi,\eta,s) + \cdots, \end{aligned} \tag{14}$$

and obtain at the leading orders that (i) $\tilde{v}_0$ does not depend on the fast variables ψ and η, (ii) $\tilde{u}_0$ and ρ_0 have the same dependence on ψ [that is, $(\partial/\partial\psi)(\tilde{u}_0 - \rho_0) = 0$], and (iii) ρ_0 and $\tilde{v}_1$ satisfy the following system of partial differential equations (equivalent to the Tricomi equation):

$$2\kappa(s)\,\eta\rho_{0\psi} + \tilde{v}_{1\eta} = 0, \tag{15}$$

$$\tilde{v}_{1\psi} + \rho_{0\eta} = 0. \tag{16}$$

It can be shown that this expansion is valid for $|d| \ll \sqrt{\epsilon}$, or $|\eta| \ll \epsilon^{-1/6}$.

Since the problem is linear, we can separate variables. In particular, since the coefficients of the equations are independent of ψ, the dependence on ψ must be exponential. Thus we can write

$$\rho_j = \hat{\rho}_j(\eta,s)e^{ik\psi}, \tag{17}$$

with similar expressions for the other variables—where k is an $O(1)$ nonzero real constant. Then we obtain that $\tilde{u}_0 = \rho_0$, and $\tilde{v}_0 = 0$. (Note that $\tilde{v}_0$ vanishing is consistent with the fact that the wave motion near the caustic is mainly parallel to the caustic.) The equations (15)–(16) reduce to an Airy equation that we can solve as

$$\rho_0 = f(s)\mathrm{Ai}(-\alpha\eta)e^{ik\psi}, \tag{18}$$

$$\tilde{v}_1 = \frac{1}{ik}\,\alpha f(s)\mathrm{Ai}'(-\alpha\eta)e^{ik\psi}, \tag{19}$$

where $\alpha = (2\kappa(s)k^2)^{1/3} > 0$ and $\mathrm{Ai} = \mathrm{Ai}(z)$ is the Airy function which decays as $z \to \infty$.

In the wavefront region outside the caustic layer, where the expansion is valid, i.e. $1 \ll \eta \ll \epsilon^{-1/6}$, we can use the asymptotic behavior of the Airy function for large values of the argument, to obtain

$$\tilde{u}_0 = \rho_0 \sim 2\epsilon^{1/6}d^{-1/4}F(s)\cos\left(\frac{2}{3}\sqrt{2\kappa(s)}|k|\,\eta^{3/2} - \frac{\pi}{4}\right)e^{ik\psi}, \tag{20}$$

where $F(s) = (1/2\sqrt{\pi})f(s)(2\kappa(s)k^2)^{-1/12}$. In these formulas $f(s)$ is an arbitrary function that must be determined by matching with the incoming LGO wave into the caustic region, as explained next.

C. Linear geometrical optics and matching

In this subsection, we match the inner solution (20) with the outer solution provided by Linear Geometrical Optics (LGO). This matching provides the connection between the incident and reflected waves at the caustic. The main result is that these two waves are connected by a Hilbert Transform, a critical fact for the analysis of shock waves in later sections.

The LGO expansion for a single mode solution to (12) has the form

$$\begin{aligned} &\rho \sim [a_0(x,y) + \epsilon a_1(x,y) + \cdots]e^{ik\theta}, \\ &u \sim [u_0(x,y) + \epsilon u_1(x,y) + \cdots]e^{ik\theta}, \\ &v \sim [v_0(x,y) + \epsilon v_1(x,y) + \cdots]e^{ik\theta}, \end{aligned} \tag{21}$$

where $\theta = [\Phi(x,y) - t]/\epsilon$, $u_0 = \Phi_x a_0$, and $v_0 = \Phi_y a_0$. The action variable Φ satisfies the Eikonal equation (2) and the leading order amplitude a_0 satisfies the transport equation

$$2(\nabla\Phi)\cdot\nabla a_0 + (\Delta\Phi)a_0 = 0. \tag{22}$$

In order to solve (22), it is best to work with the characteristics for the Eikonal equation (2)—which corresponds

to writing the equations in the advanced and retarded CPCS's. Let $k_1=\Phi_x$ and $k_2=\Phi_y$. Then the rays are given by

$$\frac{dx}{d\tau}=k_1 \quad \text{and} \quad \frac{dy}{d\tau}=k_2. \tag{23}$$

Along them the Eikonal equation takes the form

$$\frac{dk_1}{d\tau}=\frac{dk_2}{d\tau}=0 \quad \text{and} \quad \frac{d\Phi}{d\tau}=1. \tag{24}$$

Clearly, since from the Eikonal equation $k_1^2+k_2^2=1$, the parameter τ is the arclength along rays (increasing in the direction of propagation). For each ray, we set $\tau=0$ at the caustic, so that the incident wave corresponds to $\tau<0$ and the reflected wave to $\tau>0$. Then we obtain the following representation for the rays and the action variable Φ:

$$\begin{aligned} &k_1=X'(s), \quad k_2=Y'(s), \quad \Phi=s+\tau, \\ &x=X(s)+k_1\tau, \quad \text{and} \quad y=Y(s)+k_2\tau. \end{aligned} \tag{25}$$

It is also easy to see that $\Delta\Phi=1/\tau$. It follows that the characteristic form of the transport equation (22) is

$$2\frac{da_0}{d\tau}+\frac{1}{\tau}a_0=0. \tag{26}$$

Thus

$$a_0=\frac{1}{\sqrt{-\tau}}\hat{I}(s) \quad \text{for the incident wave and}$$

$$a_0=\frac{1}{\sqrt{\tau}}\hat{R}(s) \quad \text{for the reflected wave,} \tag{27}$$

where the function $\hat{I}(s)$ is determined by the wave focusing at the caustic (as produced by initial or boundary conditions) and $\hat{R}(s)$ must be determined from $\hat{I}$ by matching across the caustic layer.

Thus, along any ray tangent to the caustic,

$$\rho\sim\frac{1}{\sqrt{|\tau|}}\begin{bmatrix}\hat{I}\\ \hat{R}\end{bmatrix}\exp\left(\frac{ik}{\epsilon}(s+\tau-t)\right), \tag{28}$$

with $u\sim k_1\rho$ and $v\sim k_2\rho$.

The actual solution ρ we are after is the sum of the incident and the reflected waves. These were computed for ease in characteristic coordinates, i.e. the advanced and retarded CPCS described above. However, to do the matching, it is convenient to have all solutions written down in a unified framework, specifically the NCCS [with coordinates (s,d)]. In these coordinates, the incident and reflected waves take the form

$$\rho_i\sim\left(\frac{2d}{\kappa(s)}\right)^{-1/4}\hat{I}(s)\exp\left[ik\left(\psi-\frac{2}{3}\sqrt{2\kappa(s)}\,\eta^{3/2}\right)\right], \tag{29}$$

$$\rho_r\sim\left(\frac{2d}{\kappa(s)}\right)^{-1/4}\hat{R}(s)\exp\left[ik\left(\psi+\frac{2}{3}\sqrt{2\kappa(s)}\,\eta^{3/2}\right)\right]. \tag{30}$$

The full solution, valid for $1\ll\eta\ll\epsilon^{-1/6}$, is the sum of these two waves. The function $\hat{I}(s)$ is determined by the data away from the caustic; the function $\hat{R}(s)$, on the other hand, has to be connected to $\hat{I}(s)$ through matching with the inner layer, i.e. with (20). Notice that both (29), (30) and (20) are valid in the same range $1\ll\eta\ll\epsilon^{-1/6}$ and matching is possible. Clearly we must have

$$\left(\frac{2}{\kappa(s)}\right)^{-1/4}\hat{R}(s)=\epsilon^{1/6}F(s)\exp\left(-i\,\frac{\pi}{4}\,\text{sign}(k)\right), \tag{31}$$

$$\left(\frac{2}{\kappa(s)}\right)^{-1/4}\hat{I}(s)=\epsilon^{1/6}F(s)\exp\left(i\,\frac{\pi}{4}\,\text{sign}(k)\right). \tag{32}$$

Thus the desired connection formula is

$$\hat{R}(s,k)=-i\ \text{sign}(k)\hat{I}(s,k). \tag{33}$$

Notice the amplification factor $\epsilon^{-1/6}$ between the outer LGO solution and the inner caustic expansion.

So far, we have considered a single mode solution, with frequency k/ϵ. Clearly, the whole procedure generalizes to waves which do not have a sinusoidal profile. In fact, these waves can be obtained from the single mode solution above through Fourier Analysis. The incident and reflected waves are now given, in the outer expansion, by

$$\rho_i\sim\frac{1}{\sqrt{|\tau|}}I(s,\theta) \quad \text{and} \quad \rho_r\sim\frac{1}{\sqrt{|\tau|}}R(s,\theta). \tag{34}$$

The functions I and R admit the Fourier representations,

$$I(s,\theta)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\hat{I}(s,k)e^{ik\theta}dk, \tag{35}$$

$$R(s,\theta)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\hat{R}(s,k)e^{ik\theta}dk, \tag{36}$$

where $\hat{I}$ and $\hat{R}$ satisfy the connection formula (33). It follows then that the function

$$h=I+iR=\frac{2}{\sqrt{2\pi}}\int_{0}^{\infty}\hat{I}(s,k)e^{ik\theta}dk, \tag{37}$$

is analytic in the upper half-plane $\text{Im}(\theta)>0$ and decays as $\text{Im}(\theta)\to\infty$. Thus I and R are related by the Hilbert Transform:

$$R(s,\theta)=-\frac{1}{\pi}P.V.\int_{-\infty}^{\infty}\frac{I(s,\zeta)}{\zeta-\theta}\,d\zeta=H(I), \tag{38}$$

with $I=-H(R)$.

In summary, we have shown the following: (a) The amplification factor for the amplitude near a caustic is given by $\epsilon^{-1/6}$; (b) The width of the caustic layer is $\epsilon^{2/3}$; and (c) The incident and reflected wave profiles are connected by the Hilbert Transform. These facts play a crucial role in the arguments of the following sections.

D. Linear theory for discontinuous fronts

Although we have concentrated so far on linear waves, our purpose in this work is to study caustics for nonlinear waves. Given enough time, nonlinear hyperbolic waves will generally break into shocks; therefore it is important to un-

derstand how a shock wave behaves at a caustic. As a first step in this direction, we study now the linear caustics of discontinuous fronts.

From the results of the previous subsection, we know that the incoming and reflected waves are related by the Hilbert Transform. But the Hilbert Transform. But the Hilbert Transform of a function with a discontinuity has a logarithmic singularity. Therefore, if the incoming wave is discontinuous, the amplitude of the reflected wave will be unbounded. This shows that there is a fundamental difficulty with the theory when the waves are discontinuous. No matter how small the wave amplitude, linear theory in this case appears to eventually ''self-destruct,'' since the linearizing assumption must fail sufficiently close to the caustic for a discontinuous wave. But the way nonlinearity plays a role in resolving this difficulty involves a rather subtle effect: a ''naive'' approach to the problem quickly leads to a contradiction—as shown below in Section III.

There are many similar situations, where linear theory fails for discontinuous wavefronts across some layer or interface due to a Hilbert Transform type of connection formula. In most cases an appropriate nonlinear theory is unknown. The first example of such a situation (as far as we know) was pointed out by Lighthill[11] in 1950.

As a prototypical example of the singularity predicted by the connection formula (38), consider an incoming wave consisting of a periodic saw-tooth (in this case we must use Fourier Series instead of Transforms),

$$I(\theta)=\tfrac{1}{2}(\theta-\pi), \quad \text{for } 0<\theta<2\pi, \quad \text{extended periodically.}$$

This is the real part of the analytic function $h=-i\log(1-e^{i\theta})$, which decays as $\mathrm{Im}(\theta)$ approaches $+\infty$. Thus the reflected wave is given by the imaginary part of this function, i.e.

$$R(\theta)=-\log\left|2\sin\left(\frac{\theta}{2}\right)\right|.$$

These incoming and reflected wave profiles are represented in Fig. 5. We see that, at the points where $I(\theta)$ is discontinuous, $R(\theta)$ develops a logarithmic singularity. This behavior, shown here in an example, is completely general. In fact, it is easy to show that if $I=I(\theta)$ is piecewise smooth and has a simple discontinuity at some $\theta=\theta_0$, with jump $\mu=I(\theta_0+0)-I(\theta_0-0)$, then near θ_0,

$$R(\theta)=\frac{\mu}{\pi}\log|\theta-\theta_0|+O(1). \tag{39}$$

It is not surprising that a focusing front with a discontinuity should produce an infinity at the caustic. After all, the linear amplification factor at the caustic goes like $(k/\epsilon)^{1/6}$. Since discontinuities are dominated by the energy in the large k's, an infinite amplification at the caustic is to be expected. What is surprising though, is the fact that the infinity occurs not only within the caustic layer, but propagates outside this layer with the reflected wave! To confirm that this result is not an artifact of the asymptotics, we carry out in the next subsection an example where we can solve the equations exactly: the case where the caustic is a circle.

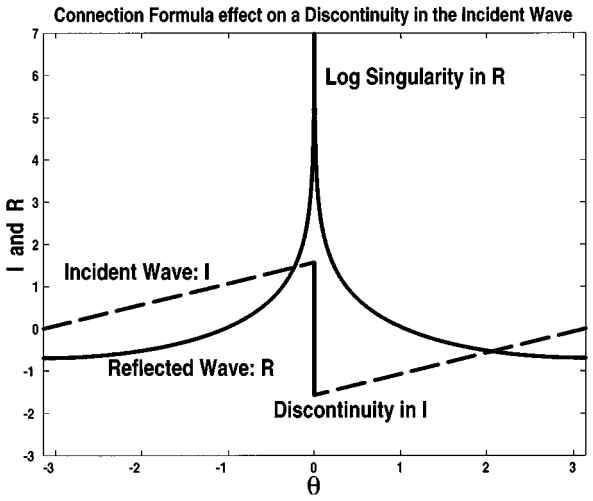

FIG. 5. Logarithmic singularity produced by the Hilbert Transform in the connection formula (38) when the incident wave has a discontinuous profile. The example here is for a saw-tooth incident wave I, but the effect is very general. The spike in R is exponentially narrow, but it reaches ∞.

E. A simple example: Circular caustics

When the caustic is a circle, a number of simplifications occur: The incoming and reflected fronts, which can be constructed using the approach in subsection II A, become spirals with rather simple formulas. In fact both the advanced and the retarded Caustic Polar Coordinate Systems (CPCS) admit explicit and rather simple representations in terms of the Standard Polar Coordinate System (SPCS). Furthermore, the Normal Caustic Coordinate System (NCCS) is in this case essentially the SPCS, in which the solutions of the wave equation can be written explicitly by the separation of variables.

To see this, consider a circular caustic of radius one, which we parametrize by $x=X(s)=\cos(s)$ and $y=Y(s)=\sin(s)$-as in subsection II A. Then, if (r,ϑ) is the SPCS, it is easy to see that

$$\begin{aligned}
&r^2=1+\tau_i^2 \quad \text{and} \quad \tan(s_i-\vartheta)=\tau_i \quad (\text{with } 0<s_i-\vartheta<\tfrac{1}{2}\pi),\\
&r^2=1+\tau_r^2 \quad \text{and} \quad \tan(\vartheta-s_r)=\tau_r \quad (\text{with } 0<\vartheta-s_r<\tfrac{1}{2}\pi),\\
&r=d+1 \quad \text{and} \quad \vartheta=s.
\end{aligned} \tag{40}$$

So, if we write $r=\sec\beta$ (with $0<\beta<1/2\pi$), for $r>1$ we obtain the simple relations

$$\begin{aligned}
&\tau_i=\tan\beta \quad \text{and} \quad s_i=\vartheta+\beta,\\
&\tau_r=\tan\beta \quad \text{and} \quad s_r=\vartheta-\beta.
\end{aligned} \tag{41}$$

Thus a circular caustic provides an ideal scenario to check the validity of the various conclusions of the asymptotic analysis.

Consider now the wave equation (in SPCS)

$$\Psi_{tt}=\frac{1}{r}(r\Psi_r)_r+\frac{1}{r^2}\Psi_{\vartheta\vartheta}. \tag{42}$$

Now we can write the following special exact solutions obtained by separation of variables:

$$\Psi=\sum_{n=1}^{\infty} a_n J_n(nr)e^{in(\vartheta-t)}+\text{c.c.}, \tag{43}$$

where the J_n's are the Bessel Functions of the first kind,[12] the a_n's are arbitrary complex constants, and c.c. stands for complex conjugate.

There is no ϵ in equation (43), as we have not yet introduced any high frequency assumption. Thus, consider now the limit in which the only contributions that matter for Ψ in (43) arise from $n\gg 1$. Then, outside the caustic [for $r=\sec\beta>1$, as in (41)] we can use the asymptotic formula,[12]

$$J_n(nr)\sim\sqrt{\frac{2}{n\pi\tan\beta}}\cos(n\tan\beta-n\beta-\tfrac{1}{4}\pi), \tag{44}$$

valid for $n\gg 1$. This and (41) yield

$$\Psi\sim\frac{1}{\sqrt{\tau_i}}I(\Phi_i-t)+\frac{1}{\sqrt{\tau_r}}R(\Phi_r-t), \tag{45}$$

where Φ_i and Φ_r are as in (5), $\Theta_i=\Phi_i-t$ is the incident phase, $\Theta_r=\Phi_r-t$ is the reflected phase, and the functions I and R are given by

$$\begin{aligned} I(\theta)&=e^{+i\frac{1}{4}\pi}\sum_{n=1}^{\infty}\frac{1}{\sqrt{2\pi n}}a_n e^{in\theta}+\text{c.c.},\\ R(\theta)&=e^{-i\frac{1}{4}\pi}\sum_{n=1}^{\infty}\frac{1}{\sqrt{2\pi n}}a_n e^{in\theta}+\text{c.c.} \end{aligned} \tag{46}$$

Clearly, both $I(\theta)$ and $R(\theta)$ are 2π-periodic and their Fourier Series coefficients are related by equation (33), so that R is the Hilbert Transform of I—exactly as in subsection II C. Thus, if the a_n's are chosen so that $I(\theta)$ has a discontinuity, $R(\theta)$ will have a logarithmic singularity.

We note that the calculation above shows that, in general, the convergence properties of (43) outside the caustic ($r>1$) are characterized by the series with general term $(a_n/\sqrt{n})e^{in\zeta}\cos(n\xi)$.

On the other hand, inside the caustic we can write $r=\operatorname{sech}\alpha$, with $\alpha>0$. Then, for $n\gg 1$, we have[12] $J_n(nr)\sim(2n\pi\tanh\alpha)^{-1/2}\exp[-n(\alpha-\tanh\alpha)]$, where $\alpha>\tanh\alpha$. Thus, the convergence properties of (43) are exponentially accelerated and the solution inside the caustic will generally be very smooth.

Finally, the convergence near the caustic can be obtained from the asymptotic behavior of $J_n(nr)$ for $n\gg 1$ and r near 1, which involves an $O(1/\sqrt{n})$ behavior times an Airy function of r. This also can be used to obtain the expansions in subsection II B for this case.

One can also consider equation (43) for $r\gg 1$, without the explicit high frequency assumption that lead us to use formula (44). In this case we must use the asymptotic formula[12]

$$J_n(z)\sim\sqrt{\frac{2}{\pi z}}\cos(z-\tfrac{1}{2}n\pi-\tfrac{1}{4}\pi),$$

valid for $z\gg 1$. We then obtain exactly the same result as (45). The small parameter here arises from the quotient of the wavelength of the incoming and reflected waves—Archimedian spirals centered at the origin of step 2π—far away from the caustic (which is 2π) and their radius of curvature $r\gg 1$.

III. WEAKLY NONLINEAR ASYMPTOTICS

In this section, we develop inner and outer expansions near a caustic similar to those in section II, but including nonlinear effects. Nonlinearity brings in two new important ingredients: the existence of critical amplitudes required for the nonlinear corrections to appear at the same order as the linear terms; and the lack of exact solutions for the asymptotic equations, which makes matching the inner and outer expansions less straightforward than for the linear case. The section that follows after this is devoted to solving these difficulties, bypassing the need for exact solutions and introducing a new scale, totally absent in the linear case.

A. Caustic expansion

The material in this subsection follows very closely that in subsection II B, but for a weakly nonlinear example. Again, we consider, for the equations of two-dimensional **Isentropic Gas Dynamics**, a small perturbation from an equilibrium state. We assume that the equations have been nondimensionalized so that the equilibrium state has density and sound speed equal to one and flow velocity equal to zero. We further assume that the length scales are such that typical curvatures for the wavefronts and caustics are order one.

Writing the density as $1+\rho$ and the flow velocity as $(u,v)^t$, so that ρ, u, and v are small, the equations are

$$\begin{aligned} &\rho_t+u_x+v_y+(\rho u)_x+(\rho v)_y=0,\\ &u_t+q_x+uu_x+vu_y=0,\\ &v_t+q_y+uv_x+vv_y=0, \end{aligned} \tag{47}$$

where $q=q(\rho)$ is defined in terms of the pressure $P=P(\rho)$ (or $c=\sqrt{dP/d\rho}$, the sound speed) by

$$\frac{dq}{d\rho}=\frac{1}{1+\rho}\frac{dP}{d\rho}=\frac{c^2}{1+\rho}.$$

The variables q and c have expansions in powers of ρ of the form

$$\begin{aligned} q(\rho)&=\rho+(\beta-0.5)\rho^2+\cdots,\\ c(\rho)&=1+\beta\rho+\cdots, \end{aligned} \tag{48}$$

where β is a positive constant. Thus these equations reduce to those in (12) when ρ, u, and v are infinitesimal,

Next we consider a neighborhood of a caustic like the one in subsection II A, and rewrite the equations above in the NCCS. As in subsection II B we let $\tilde{u}$ and $\tilde{v}$ be the components of the flow velocity respectively parallel and normal to the caustic. Then we obtain the following extension of (13):

$$\begin{aligned} &(1+\kappa d)\rho_t+((1+\rho)\tilde{u})_s-((1+\rho)(1+\kappa d)\tilde{v})_d=0,\\ &(1+\kappa d)\tilde{u}_t+(q+\tfrac{1}{2}\tilde{u}^2)_s-\tilde{v}((1+\kappa d)\tilde{u})_d=0,\\ &(1+\kappa d)\tilde{v}_t+\tilde{u}\tilde{v}_s-(1+\kappa d)(q+\tfrac{1}{2}\tilde{v}^2)_d=-\kappa\tilde{u}^2, \end{aligned} \tag{49}$$

where $\kappa=\kappa(s)>0$ is the caustic's curvature.

We introduce again the small parameter $0<\epsilon\ll 1$, the variables ψ and η from (9)–(10), and expand in one-third powers of ϵ. However, since the equations are now nonlinear, *the overall size of* ρ, u, *and* v *is important*. We cannot start, as in (14), with $O(1)$ leading order terms and only later, when it becomes necessary for matching, multiply the whole expansion by some convenient parameter, such as the $\epsilon^{-1/6}$ needed in subsection II C. Now the prefactor is significant from the very beginning. Thus we propose the ansatz,

$$\begin{aligned}
&\rho\sim\gamma(\rho_0(\psi,\eta,s)+\epsilon^{1/3}\rho_1(\psi,\eta,s)+\cdots),\\
&\tilde{u}\sim\gamma(\tilde{u}_0(\psi,\eta,s)+\epsilon^{1/3}\tilde{u}_1(\psi,\eta,s)+\cdots),\\
&\tilde{v}\sim\gamma(\tilde{v}_0(\psi,\eta,s)+\epsilon^{1/3}\tilde{v}_1(\psi,\eta,s)+\cdots),
\end{aligned}\tag{50}$$

where $\gamma>0$ is the prefactor.

The size of the parameter γ will determine whether the dominant behavior is linear and the analysis in section II, applies or not. In particular, we want to determine the critical size or **threshold value** γ_c of γ (as a function of ϵ) for which nonlinearity first appears at leading order. For $\gamma\ll\gamma_c$ the behavior will be dominated by the linear terms; for $\gamma\gg\gamma_c$, on the other hand, nonlinearity will dominate and the scalings we used at the caustic may not even make sense (since they correspond to the folding of wavefronts in a linear system). A straightforward calculation shows that the threshold value is

$$\gamma_c=\epsilon^{2/3}.$$

When we take $\gamma=\gamma_c$ in (50), the resulting asymptotic equations are the following extension of (15)–(16) (a nonlinear Tricomi equation):

$$2\kappa(s)\,\eta\rho_{0\psi}+\tilde{v}_{1\eta}=((\beta+1)\rho_0^2)_\psi,\tag{51}$$

$$\tilde{v}_{1\psi}+\rho_{0\eta}=0,\tag{52}$$

where we have taken $\tilde{u}_0=\rho_0$ and $\tilde{v}_0=0$ [as we eventually had to do in (17)–(19), for the linear case].

B. Weakly Nonlinear Geometrical Optics (expansion away from the caustic)

An analysis similar to that in subsection II C applies away from the caustic (see for instance Refs. 13–15), with appropriate corrections for the presence of nonlinearity. Here the critical size for the variables ρ, u, and v (required to have nonlinear terms appearing at leading order in the asymptotic equations) is simply ϵ, with corresponding expansions

$$\rho\sim\epsilon\rho_0+\epsilon^2\rho_1+\cdots\,,\quad u\sim\epsilon u_0+\epsilon^2u_1+\cdots\,,$$

and

$$v\sim\epsilon v_0+\epsilon^2v_1+\cdots\,,\tag{53}$$

where the variables are functions of θ, x, and y, with θ as in (1). Then Φ satisfies the same Eikonal equation (2) that applies in the linear case and it is still true that $u_0=\Phi_x\rho_0$ and $v_0=\Phi_y\rho_0$, very much as in (21). But the transport equation (22) is modified by nonlinear terms into

$$(\nabla\Phi)\cdot\nabla\rho_0+\frac{1}{2}(\Delta\Phi)\rho_0+\left(\frac{\beta+1}{2}\,\rho_0^2\right)_\theta=0.\tag{54}$$

Notice that, in the linearized case, we can separate the θ dependence as $e^{ik\theta}$ [as in (21)]. Then this expansion reduces to the one in subsection II C.

It can be shown that equation (54) has the appropriate conservation form to deal with shocks, even though (47) is not in conservation form [The simplest way is to check that (54) does give the correct jump conditions.] The same result applies to the caustic equations (51)–(52).

We could include an explicit dependence on the time variable t in the expansion (53)—to consider ''nonsteady'' wavetrains. The only change this introduces in the equations above is an additional term ρ_{0t} in the transport equation (54). To be consistent, we should then also include a t dependence in the caustic expansion (50). This, however, introduces no changes in equations (51)–(52); that is to say, the dependence on t at the level of the caustic equations is merely parametric. We will not do any of this in this paper, since our main concern is the study of the behavior of shocks at a caustic, so that nonsteadiness is of secondary interest.

As before, we introduce the ray (characteristic) coordinates for the Eikonal equation [see (23)–(25)]. Then (54) reads, along each ray, as

$$\frac{d}{d\tau}\,(\sqrt{|\tau|}\rho_0)+\left(\frac{\beta+1}{2}\,\sqrt{|\tau|}\rho_0^2\right)_\theta=0.\tag{55}$$

Use now the ray variables (s,τ) to replace (x,y), where (as in subsection II C), s indicates the value of the arclength along the caustic where the ray is tangent to it and τ vanishes there [in particular (25) applies]. Then the transformation,

$$\rho_0(s,\theta,\tau)=\frac{1}{\sqrt{|\tau|}}\,f(s,\theta,z)$$

[where $z=\mathrm{sign}(\tau)\sqrt{|\tau|}$] reduces the equation above to the well known constant coefficient Hopf equation,

$$f_z+((\beta+1)f^2)_\theta=0,\tag{56}$$

where $z<0$ corresponds to the incident wave and $z>0$ to the reflected wave. Note that this equation is not to be applied across $z=0$, where matching with the caustic layer must be used instead.

It is clear that the limit near the caustic of this outer Weakly Nonlinear Geometrical Optics (WNGO) expansion, has exactly the same form as that provided by the linear theory (LGO): Namely, equation (34) applies, where

$$I(s,\theta)=f(s,\theta,0^-)\quad\text{and}\quad R(s,\theta)=f(s,\theta,0^+),\tag{57}$$

for the incident and reflected waves, respectively. Here 0^- and 0^+ stand for the limits $z\to 0$ with $z<0$ and $z>0$, respectively. One important difference with LGO is that, in the linear case, the wave shape does not evolve as propagation along the rays occurs, while in WNGO the wave shape does change, following equation (56). In fact, even if a wave starts with a smooth profile, shocks may form by the time the wave reaches the caustic, giving a discontinuous incident wave.

A second important point is that now we must be careful when constructing the full solution outside the caustic layer as the sum of the incident and reflected waves, as was done in the linear case. Because of the nonlinear interactions between these two waves, this cannot be done without some care. However, if the following circumstances apply, then (to leading order at least) adding the two waves is correct.

(1) If the incident wave is a single wavefront (as in the case of a single weak shock focusing), then outside the caustic layer the incident and reflected fronts do not occupy the same region of space and thus do not interact.
(2) In the case of Gas Dynamics, in the absence of leading order entropy or vorticity variations [as in the example provided by (47) and the expansion (53)], acoustic waves do not interact (resonate) at leading order.

Generally, however, for oscillatory wavetrains, one needs to worry about the incident and reflected waves being in resonance in the sense of Ref. 16. (There is also a coherence condition one has to consider.) These conditions are rather complicated and not easily satisfied by the waves associated with a caustic. Thus it seems fair to assume that, in most cases, simply adding the incident and reflected waves is the correct approach.

C. Problems with matching: Linear rates of growth and thresholds of nonlinearity

The next step is matching the outer and inner expansions. As in the linear case, this matching should provide enough information to determine the shape and amplitude of the reflected wave from knowledge of the incoming wave. In addition, it should provide the amplification factor of the wave amplitude in a neighborhood of the caustic, a most important prediction for practical purposes.

The first difficulty that arises is that we do not know a priori which are the expansions that we need to match. In the linear case, we could carry out both expansions beforehand, and later multiply the inner expansion by the amplification factor necessary for matching. For the full equations, however, this amplification factor will determine whether the inner equations are linear or nonlinear; it is thus not something that can be left as a small detail to be taken care of at the end.

To resolve this difficulty, we can attempt the following thought experiment: starting with a linear regime, where the results are known, gradually increase the amplitude of the incoming wave, until nonlinearity shows up, either in the outer, in the inner expansion, or in both. At this point, at least one of the expansions has to be replaced by its weakly nonlinear analog, as developed in the previous two subsections III A and III B.

Hunter and Keller carried out this procedure in Ref. 3 for smooth waves. They found that the threshold of nonlinearity is reached first in the outer domain, when the waves in a neighborhood of the caustic are still behaving linearly. They concluded that the caustics of smooth Weakly Nonlinear Geometrical Optics waves can be described by linear theory, as the following argument shows.

The Hunter and Keller's argument for smooth WNGO waves. If a typical wavelength for the incoming wave is ϵ, and the amplitude is small enough that linear theory applies, the amplification factor at the caustic is $\epsilon^{-1/6}$, as shown in subsection II C. The threshold amplitude of the incoming wave for nonlinearity to show up in the outer expansion is $O(\epsilon)$, as described in subsection III B. Linear theory near the caustic will then apply, since this predicts an amplitude near the caustic of order $\epsilon^{5/6}=\epsilon*\epsilon^{-1/6}$ (which is smaller than the threshold value $\epsilon^{2/3}$ for nonlinear caustic behavior described in subsection III A). Thus the first onset of nonlinearity occurs away from the caustic, when the amplitude of the incoming wave has an amplitude comparable to its wavelength, with the caustic layer still behaving linearly.

Hunter and Keller extend this argument formally, to determine for which amplitude of the incoming wave will the caustic layer behave nonlinearly. The validity of this formal extension, however, is doubtful. Once the outer domain is strongly nonlinear, there is no reason to think that the geometry of the caustic layer will scale similarly to that of a linear wave. Thus the very ansatz of the inner expansion may not make sense. For that matter, even the notion of caustic may have to be revised.

At first sight the result above, with the waves behaving linearly near the caustic (even though they behave nonlinearly outside) may seem puzzling. It is, however, as Hunter and Heller point out, not so surprising. Nonlinear behavior in small amplitude waves arises not just from amplitude considerations, but also from how long the nonlinearity acts. For a quadratic nonlinearity, it is the product of these two factors that counts. The caustic layer is fairly thin—$O(\epsilon^{2/3})$, as shown earlier. The waves spend only a short time in it: $O(\epsilon^{1/3})$, instead of the $O(1)$ time they spend outside. Thus an amplitude larger by this exact amount is needed for the nonlinearity to affect the caustic layer.

In this work, we study caustics of shock waves. The arguments above, if naively applied to shock waves, yield the result that the caustics of weak shock waves should behave linearly. There are two reasons why we cannot accept this answer.

The first reason is experimental: Sturtevant and Kulkarny studied the focusing of weak shock waves in Ref. 4. They found that the behavior at a caustic of a weak shock wave is always nonlinear, no matter how small the amplitude of the incoming wave. The caustic always has an internal wave structure including a triple shock intersection, a very nonlinear and puzzling configuration (see section V) Numerical calculations of focusing weak shocks yield the same results.[9]

The second reason is purely analytical: If the caustic of a weak shock wave were to be properly described by linear theory, then the connection between the incoming and reflected wave profiles should be given by the Hilbert Transform (38). As described in subsection II D, this implies that the reflected wave has a logarithmic singularity, with unbounded amplitude. But how can we justify making a linear, or even a weakly nonlinear approximation, based on small amplitudes, when the amplitude takes infinite values? Clearly, such asymptotic approximation would be questionable. An unbounded reflected wave, surprising yet consistent for linear waves, becomes rather unacceptable for the full nonlinear equations.

Thus the hypothesis of linear behavior at the caustics of weak shock waves does not agree with experiments and is internally inconsistent. It has to be discarded and replaced by a globally nonlinear theory, developed in the following section.

IV. CAUSTICS OF WEAK SHOCK WAVES: A NEW NATURAL SCALE

In this section we introduce a new scaling, induced by the nonlinear effects near the caustics of weak shock waves. Caustic layers for weak shock waves are thinner than those of smooth weakly nonlinear waves, and the amplification factor for the amplitude is correspondingly larger.

A. The width of the caustic layer

First we shall rewrite the arguments concluding the previous section more formally. Consider a weak shock wave which (away from focusing) has $O(1)$ curvature and $O(\epsilon)$ amplitude. The propagation of this wave can be described using Weakly Nonlinear Geometrical Optics **(WNGO)**, as in subsection III B. If a caustic forms, WNGO is valid for the description of the propagating front away from the caustic, for $d \gg \epsilon^{2/3}$, and near the front, for $|\Phi - t| \ll 1$, where Φ satisfies the Eikonal equation (2) and $\Phi = t$ is the approximate location of the front.

As the front approaches the caustic, WNGO predicts amplitudes of order,

$$a_c = \epsilon d^{-1/4}, \tag{58}$$

where d is the NCCS coordinate. The behavior of the Eikonal fronts near the caustic, on the other hand, suggests the inner variables

$$\Psi = \frac{s-t}{\epsilon} \quad \text{and} \quad \eta = \frac{d}{\epsilon^{2/3}}, \tag{59}$$

where $s = \Phi$ is the arclength along caustics, the other NCCS coordinate. As discussed in subsection III A, an expansion in terms of these variables yields the result that nonlinearity only plays a role if the amplitude at the caustic is at least $O(\epsilon^{2/3})$. But, from (58),

$$\epsilon^{7/8} \ll a_c \ll \epsilon^{5/6},$$

in the domain where both WNGO and the caustic expansion are valid, i.e., for $1 \ll \eta \ll \epsilon^{-1/6}$. Since this amplitude does not reach the critical value $\epsilon^{2/3}$, we conclude that the behavior at the caustic should be well described by linear theory. This is essentially the argument that Hunter and Keller applied to smooth waves.

Yet a problem arises when the incident front is a shock wave. We saw in subsection II D that, when the caustics are linear, the connection between the incident and reflected waves is given by a Hilbert Transform which, for a discontinuous incident front (such as a shock) yields a reflected wave with unbounded amplitude. Such waves clearly cannot be handled by either linear or weakly nonlinear theories, so we reach an impasse. Since after assuming the inner scales in (59), everything else follows by deduction, we conclude that the scales in (59) must fail for shocks.

Let us reexamine these scalings for the example of a single incident shock. Clearly, an inviscid shock *does not* provide any transversal short wave scale, since the shock has no thickness. For pure shock solutions (such as the one we are now considering) the ϵ^{-1} scale introduced by WNGO through the variable $\theta = (\Phi - t)/\epsilon$, serves only as an ordering parameter. The expansion is really an expansion in powers of the distance to the shock, and valid only near the shock front. Now, in equation (58), ϵ clearly has an interpretation as an amplitude, but in (59) the inner variables ψ and η follow from an interpretation of ϵ as a wavelength. Thus, since the incoming wave has no length scale, assuming the scales in (59) is really arbitrary: We pull out a length scale out of nowhere, and then are puzzled when the nonlinearity does not conform with this length scale leading us to paradoxical results!

The proper approach is to accept that we do not have *a priori* a transversal length scale at the caustic, i.e. the proper ''ϵ'' for length scaling is an unknown to be determined, that we shall denote ϵ_c. We shall still assume that the relation between transversal and longitudinal scalings at the caustic is the same as in the linear case, so—as in (59)—we introduce the variables

$$\psi = \frac{s-t}{\epsilon_c} \quad \text{and} \quad \eta = \frac{d}{\epsilon_c^{2/3}}. \tag{60}$$

Then we will let the nonlinearity ''choose'' ϵ_c. That is, we will let ϵ_c have exactly the right size so that nonlinear effects occur in the caustic layer (these should eliminate the infinities that linearized theory predicts). The ultimate verification of this assumption will be provided by the consistency of the resulting matched expansions.

Under this scaling, using (58), it follows that the amplification factor at the caustic is given by

$$d^{-1/4} = \epsilon_c^{-1/6} \eta^{-1/4}.$$

On the other hand, we know that the behavior at the caustic layer has to involve nonlinear effects, since otherwise a Hilbert Transform appears (giving rise to a logarithmic singularity). So ϵ_c is determined by the condition that the amplitude at the caustic reach the nonlinear threshold discussed in subsection III A. That is,

$$\epsilon_c^{2/3} = \epsilon \epsilon_c^{-1/6},$$

where the term on the right follows from the *outer amplitude* times the amplification factor at the caustic and the term on the left is the critical threshold amplitude for a nonlinear caustic. Thus we have

$$\epsilon_c = \epsilon^{6/5}. \tag{61}$$

With this new scale, the resulting caustic expansion will be valid for

$$|d| \ll \sqrt{\epsilon_c} = \epsilon^{3/5},$$

while the outer WNGO expansion is valid for

$$d \gg \epsilon^{2/3}.$$

Since $2/3 > 3/5$, the two regions overlap in the range

TABLE I. Scales near caustics of smooth and discontinuous waves.

	Smooth waves with $O(\epsilon)$ amplitude and frequency	Weak shock waves with $O(\epsilon)$ strength
Caustic basic length scales:		
Longitudinal	ϵ	$\epsilon_c=\epsilon^{6/5}$
Transversal	$\epsilon^{2/3}$	$\epsilon_c^{2/3}=\epsilon^{4/5}$
Amplitude at caustic	$\epsilon^{5/6}$	$\epsilon_c^{2/3}=\epsilon^{4/5}$
Amplification	$\epsilon^{-1/6}$	$\epsilon_c^{-1/6}=\epsilon^{-1/5}$

$$\epsilon^{2/3}\ll d\ll\epsilon^{3/5},$$

and matching of the two expansions is possible. Except for the fact that we use ϵ_c instead of ϵ, the caustic layer expansion and equations are the same as those in subsection III A. Thus equations (51) and (52) apply:

$$2\kappa(s)\,\eta\rho_{0\psi}+\tilde{v}_{1\eta}=((\beta+1)\rho_0^2)_\psi,\quad \tilde{v}_{1\psi}+\rho_{0\eta}=0. \quad (62)$$

We consider solutions of these equations that decay for η large. Thus their asymptotic behavior has the same form as that of the linearized equations, which is exactly right to match the outer WNGO expansion solutions as they approach the caustic.

Introduce now the variables $u=u(x,y)$ and $v(x,y)$, as follows:

$$\rho_0=\frac{(2\kappa\nu)^{2/5}}{2(\beta+1)}\,u(x,y)\quad\text{and}\quad \tilde{v}_1=\frac{(2\kappa\nu)^{3/5}}{2(\beta+1)}\,v(x,y),$$

where ν is an arbitrary constant,

$$x=(2\kappa)^{2/5}\nu^{-3/5}\psi\quad\text{and}\quad y=(2\kappa)^{3/5}\nu^{-2/5}\eta.$$

Then the caustic equations above take the normal form of a Nonlinear Tricomi Equation,

$$(yu-\tfrac{1}{2}u^2)_x+v_y=0,\quad u_y+v_x=0, \quad (63)$$

written in the appropriate conservation form to handle shocks.

In conclusion: for weak shock waves, the width of the caustic layer is given by

$$\text{width}=\epsilon_c^{2/3}=\epsilon^{4/5} \quad (64)$$

(thinner than for smooth waves) and the amplification of the amplitude by

$$\text{amplification}=\epsilon_c^{-1/6}=\epsilon^{-1/5} \quad (65)$$

(larger than for smooth waves). Table I compares the scales near a caustic for a smooth weakly nonlinear wave of amplitude and frequency of order ϵ, with those for a shock wave with strength ϵ.

These new scales allow us to resolve the contradictions that the use of the linear scalings lead to, with infinities in the reflected wave produced by the Hilbert Transform connection formula (38). It seems clear, though, that this is a matter that must be investigated carefully, that, in the context of (62), such infinities cannot occur. If a narrow logarithmic spike, such as the one shown in Fig. 5, starts to develop, the nonlinear terms should immediately generate a shock that will, literally, clip it to finite size.

As a counterpart, now we have to solve a more complex (nonlinear) problem in the caustic layer—one that cannot be solved exactly using separation of variables, as it is the case with the Linear Caustic Equations. Furthermore, as shown below in section V, this nonlinear problem leads to some interesting ''paradoxes'' of its own.

B. Shocks within smooth waves

In order to better understand the way in which the new nonlinear scales resolve the questions posed in subsection III C, consider a situation in which both the old and the new scales are present: a periodic incident wave, with period and amplitude of $O(\epsilon)$, including at least one shock. A simple example is provided by the saw-tooth wave,

$$I(\theta)=\tfrac{1}{2}(\theta-\pi),\quad \text{for } 0<\theta<2\pi,\quad \text{extended periodically,}$$

displayed in Fig. 5, but the argument below, sketched in Fig. 6, is general.

As the wave approaches the caustic, it is amplified by $O(\epsilon^{-1/6})$. Then, in a layer of transversal width $O(\epsilon^{2/3})$, Linear Caustic Theory applies, with an amplitude of $O(\epsilon^{5/6})$. Except that the amplitude near the shocks continues to grow, and would end up producing a singularity if not checked: The logarithmic singularity of the Hilbert Transform in subsection II D. After an additional amplification of $O(\epsilon^{-1/30})$, the amplitude near the shocks reaches the threshold of nonlinearity at $O(\epsilon^{4/5})$. Within a distance $O(\epsilon^{4/5})$ of the caustic, following the scalings in Table I, a new layer arises. In this thinner layer, the Nonlinear Caustic Equations of subsection III A apply, implemented as explained in subsection IV A (using $\epsilon_c=\epsilon^{6/5}$ and $\gamma=\gamma_c$). It should be clear from the arguments in subsection IV A that these Nonlinear Caustic Equations actually apply only in a small neighborhood of the shock front, of longitudinal width $O(\epsilon_c)$. Note that this width (in the direction of propagation) is only a small fraction of the wavelength, which is $O(\epsilon)$.

The infinities of linearized theory clearly arise because, from the point of view of an expansion that assumes leading orders of amplitude $O(\epsilon^{5/6})$, the spikes near the shocks—of amplitude $O(\epsilon^{4/5})$—are effectively infinite. This creates an error that then propagates outside the caustic, in the reflected wave.

The picture above of the process near a caustic for a high frequency weakly nonlinear wave combining smooth parts with shocks appears rather frightfully complicated, with a layer within a layer ''tracking'' each shock wave as it goes though the ''main'' (linear) caustic layer—which implies that an extra matching is also needed. There is, however, a simple way to model all this with just one set of caustic equations and a single caustic layer. Even with this simplification, the approach is still more complicated than what linear theory for smooth waves allows; but no more so than the theory for a simple shock wave we presented earlier in subsection IV A.

In order to see this, we first summarize the basic ansatz of Linear and Nonlinear Caustic Theory, using the subscripts l and nl to indicate quantities associated with the linear and nonlinear cases, respectively. Linear Caustic Theory uses the ansatz

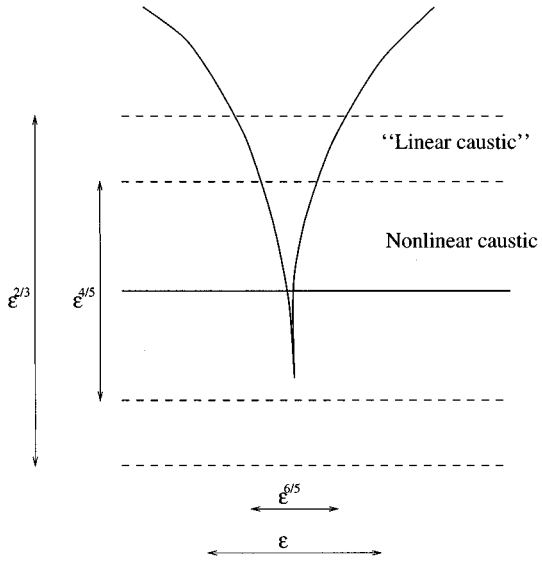

FIG. 6. Nonlinear Caustic Layer of transversal width $O(\epsilon^{4/5})$, placed inside the Linear Caustic Layer of transversal width $O(\epsilon^{2/3})$. This nonlinear layer occurs near shock fronts and is restricted to a neighborhood of the shock position of longitudinal width $O(\epsilon^{6/5})$. The picture shows a typical linear wavefront (both incident and reflected branches) placed at the location of the shock. The actual shock front will not look quite this way, particularly inside the $O(\epsilon^{4/5}\times\epsilon^{6/5})$ rectangle of the Nonlinear Caustic Layer (exactly enclosing the cusp of the linear front): see section V.

$$\rho=\epsilon^{5/6}\rho_{0l}+O(\epsilon^{7/6}) \quad \text{and} \quad \tilde{v}=\epsilon^{7/6}\tilde{v}_{1l}O(\epsilon^{9/6}), \tag{66}$$

with independent variables

$$\psi_l=\epsilon^{-1}(s-t) \quad \text{and} \quad \eta_l=\epsilon^{-2/3}d.$$

On the other hand, Nonlinear Caustic Theory uses the ansatz

$$\rho=\epsilon^{4/5}\rho_{0nl}+O(\epsilon^{6/5}) \quad \text{and} \quad \tilde{v}=\epsilon^{6/5}\tilde{v}_{1nl}+O(\epsilon^{8/5}), \tag{67}$$

with independent variables

$$\psi_{nl}=\epsilon^{-6/5}(s-t) \quad \text{and} \quad \eta_{nl}=\epsilon^{-4/5}d.$$

Thus, we have the relationships between independent variables,

$$\psi_{nl}=\epsilon^{-1/5}\psi_l \quad \text{and} \quad \eta_{nl}=\epsilon^{-2/15}\eta_l. \tag{68}$$

For the dependent variables it is tempting to write

$$\rho_{0nl}=\epsilon^{1/30}\rho_{0l} \quad \text{and} \quad \tilde{v}_{1nl}=\epsilon^{-1/30}\tilde{v}_{1l}. \tag{69}$$

But this last equation is not quite true: there are higher order terms involved in the dependent variables expressions and (even more important) the way the expansions are carried presumes that ρ_0 and $\tilde{v}_1$ are $O(1)$ quantities, independent of ϵ.

Yet, if we substitute (68) and (69) into the Nonlinear Caustic Equations (62) (which apply for the quantities with the subscript nl), the equations take the form

$$\begin{aligned} &2\kappa(s)\,\eta\rho_{0\psi}+\tilde{v}_{1\eta}=\epsilon^{1/6}((\beta+1)\rho_0^2)_\psi,\\ &\tilde{v}_{1\psi}+\rho_{0\eta}=0, \end{aligned} \tag{70}$$

where we have not written the subscript l to simplify the notation. We note that, as long as $\rho_0=\rho_{0l}$ here remains bounded, the nonlinearity is not important and the equations reduce to the linear ones. On the other hand, when the nonlinear scalings apply (by construction) these equations reduce to the appropriate ones, namely (62).

Thus, we can simply do this to take care of both pure smooth waves and mixed waves with shocks and smooth components

Set up an ansatz as in the case of smooth waves (normally giving a Linear Caustic Theory), namely, equation (50) with $\gamma=\epsilon^{5/6}$ and no ϵ_c introduced, unlike subsection IV A.

Do not, however, assume that the leading order terms are necessarily bounded, or even fully independent of ϵ. Thus, when writing the leading order equations, collect the appropriate terms from higher orders—to take care of nonlinear effects near shocks. Thus, we use (70) as the governing asymptotic equations, instead of the linear equations that would result if we proceeded as usual.

This ''recipe'' is somewhat *ad hoc*, but it does give the correct answer, as the arguments above show. A formal justification is quite likely possible, but we will not attempt it here.

A final formal small point we wish to make is the following: In the far field of the caustic layer ($\epsilon^{2/3}\ll d\ll\epsilon^{3/5}$), both in the case of smooth waves and when shocks are present, the caustic equations behave linearly. Thus the solution can be written by separation of variables (as in subsection II B). However, when shocks are present, we cannot conclude [as in equations (18)–(19)] that only Airy functions of the first kind, Ai, are relevant. This follows from the requirement that the solution vanish on the shadow side of the caustic ($\eta\to-\infty$) and can only be concluded in the case when the linear equations are uniformly valid throughout the caustic layer. But the linear equations are not uniformly valid when shocks are present. Thus Airy functions of the second kind, Bi, will also appear; with contributions tied up to the presence of shocks in the solution. Since the Hilbert Transform connection formula arises precisely from the exclusive presence of Ai in the solution, it will no longer apply when shocks occur. This should correct the problem of the infinities in the reflected wave, though what the connection formulas will be in the case with shocks is an open question.

C. Viscous effects

We have concentrated so far on inviscid flows. Real gases are viscous, of course, but their viscosity is small enough that many flows can be successfully modeled as inviscid. For weak high frequency waves and shocks, there is a threshold size of the viscosity, related to the frequency and amplitude of the waves, for which viscous effects start to play a role at the level of the Geometrical Optics approximations. This phenomenon is somewhat analogous to the thresholds of nonlinearity described in section III. Thus a question similar to the thought experiment in subsection III C may be posed: start with an inviscid weak shock wave which forms a caustic, and slowly decrease the Reynolds number (R_e) from $R_e=\infty$. Where and when will viscous

effects appear first, near the caustic or in the far field? Once these effects are present, at the critical threshold can we formulate a Geometrical Optics approximation to the flow which incorporates dissipation?

These questions arise naturally from our earlier discussions, since viscosity provides a longitudinal length scale for shocks, whose absence gives rise to the puzzling behavior of weak shock waves at caustics that we pointed out in subsection III C. Thus one may think that a sufficient amount of viscosity may result in a different solution to the puzzle from the one we developed here for inviscid shocks. We shall see, however, that this is not the case: when the viscosity is large enough to account for shocks of significant width, yet not so large that the concept of a shock stops making sense, the scales it yields at a caustic are exactly the ones displayed in Table I of subsection IV A earlier. In this subsection we develop the arguments leading to this conclusion, and display the equations they yield both near the caustic and in the far field.

It is convenient to begin by reviewing the nondimensionalization behind equations (47). Using a prime to denote the dimensional variables, we have

$$\rho' = \rho^*(1+\rho), \quad p' = \rho^*(c^*)^2 p, \quad u' = c^* u, \\ v' = c^* v, \quad x' = Lx, \quad t' = (L/c^*)t, \tag{71}$$

where ρ^* is the density of the equilibrium state, c^* is the corresponding sound speed, and L is a typical length scale for the wavefronts and caustics (e.g., a typical radius of curvature). In the simplest model to account for viscous effects, the zeros on the right hand sides of the second and third equations in (47) should be replaced by

$$\epsilon R_e^{-1} \Delta u \quad \text{and} \quad \epsilon R_e^{-1} \Delta v, \tag{72}$$

respectively, where Δ is the Laplacian in two dimensions (2-D) and R_e is the Reynolds number outside the caustic. Generally we have $R_e = LU/\nu_*$, where U is a typical flow velocity and ν_* is the kinematic viscosity. In our case, $U = \epsilon c^*$, as follows from the nondimensionalization above and equation (53). Thus we have

$$R_e = \epsilon L c^* \nu_*^{-1}, \tag{73}$$

where ϵ is the same small parameter as in our expansions earlier.

Let us now consider a viscous weak shock wave away from the caustic, of strength $O(\epsilon)$. The width l_s of such shock arises from a balance between the quadratic nonlinear compression and the dissipative spreading. We must balance terms like uu_x or $\rho\rho_x$ on the left in equations (47)—of size $O(\epsilon^2/l_s)$—with the dissipative terms above in equation (72)—of size $O(\epsilon^2/(R_e l_s^2))$. Thus

$$l_s = O(R_e^{-1}), \tag{74}$$

in the nondimensional units of equation (47).

On the other hand, as a weak shock approaches the caustic, it amplifies at a rate $1/\sqrt{|\tau|}$. We can no longer consider its amplitude as being of the form ''constant$\times\epsilon$.'' Then, the same argument leading to this last equation (74) yields that l_s decreases at a rate $\sqrt{|\tau|}$. Thus, a weak incident shock provides no length scale as it approaches a caustic, even in the viscous case! This is the crux of our argument, which we make more precise below.

From equation (74) we conclude that, for the waves of wavelength $O(\epsilon)$ in subsection III B.
If $R_e \gg \epsilon^{-1}$, the shock can be modeled as a discontinuity (for then $l_s \ll \epsilon$).

If $R_e \ll \epsilon^{-1}$ (so that $l_s \gg \epsilon$) the wave is so spread out that calling it a shock is meaningless. In fact, dissipation is so large compared with the other effects, that in the context of a Geometrical Optics approximation, the resulting equations are linear and dissipative.

Clearly then, the interesting situation arises at the critical threshold when l_s and ϵ are of the same order, since then the viscous shock thickness scale is precisely equal to the one required for a Weakly Nonlinear Geometrical Optics (WNGO) expansion. For this to occur, the Reynolds number has to be $O(\epsilon^{-1})$. Thus we write

$$R_e = (\nu\epsilon)^{-1}, \tag{75}$$

where ν is an $O(1)$ positive constant.

When this scaling applies, we can carry through a WNGO analysis, exactly as in subsection III B. The viscous terms contribute for the first time at the level of the transport equation (54), and modify it into

$$(\nabla\Phi)\cdot\nabla\rho_0 + \tfrac{1}{2}(\Delta\Phi)\rho_0 + \left(\frac{\beta+1}{2}\rho_0^2\right)_\theta = \tfrac{1}{2}\nu\rho_{0\theta\theta}. \tag{76}$$

As in subsection III B, we can introduce ray based coordinates in this equation. The same set of transformations leading there to equation (56), yield now

$$f_z + ((\beta+1)f^2)_\theta = |z|\nu f_{\theta\theta}. \tag{77}$$

Notice that, close to the caustic, as $z \to 0$, the dissipation vanishes. The physical dissipation is of course constant; it is the intensity amplification, rescaled in equation (77), which makes the dissipative term comparatively smaller. Thus, as a shock approaches the caustic, it becomes sharper and sharper. Thus, in this critical threshold case, we may still have to deal with the discontinuous input to the caustic layer from the outer expansion which was the source of trouble in linearized Caustic Theory (see subsection III C). Thus we expect linearized Caustic Theory to be inadequate for weak shocks even in the viscous case, which is the main result of this subsection.

Not all shocks, however, need to sharpen up into a profile that, from the viewpoint of the outer expansion, is essentially a discontinuity, as they approach the caustic obeying (77) above. For this to happen the shock has to have had enough time to ''relax'' to a state where nonlinearity and viscosity balance just right. If the shock strength is varying too rapidly, this balance will not be maintained and a discontinuity will not form. Of course, this situation is not that different from the one in the inviscid case: there too, some initially smooth waves develop shocks and reach the caustic with a discontinuous profile, while others do not.

Figure 7 shows a numerical solution to equation (77), starting at $z = -1$ with a smooth profile, which soon devel-

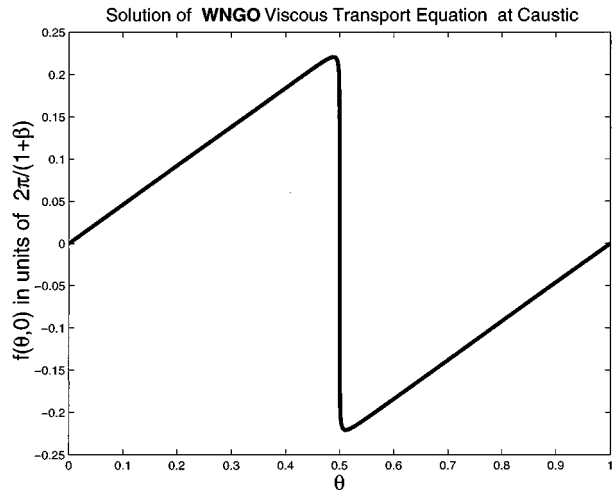

FIG. 7. Numerical solution to equation (77), to illustrate the development of nearly a discontinuity at $z=0$. We have adopted $\nu=1$; the value of β can be absorbed into the definition of f. The initial data, at $z=-1$, is $f(\theta,-1)=(2\pi/(1+\beta))\sin(\theta)$. The figure displays the solution when this wave reaches the caustic, at $z=0$; it has by then become essentially discontinuous. Notice that f is a rescaled variable; the amplitude of the physical wave is in fact growing proportionally to $1/|z|$.

ops into a viscous shock. The width of this shock gradually decreases with $|z|$, all the way to nearly zero by the time it reaches the caustic at $z=0$. The pseudo-spectral algorithm utilized for this computation will be described in a separate article, where more numerical examples will be provided. The purpose of the example here is to illustrate the way in which equation (77) may give rise to essentially a discontinuity at $z=0$.

We consider next a neighborhood of the caustic, and conduct a ''threshold'' analysis for the asymptotic behavior, similar to the one above. Specifically, consider the expansions in subsection III A, when the terms in (72) are present in the equations of motion. Then ask the following questions: (i) For which size of R_e do these viscous terms have an effect on the behavior, comparable with geometrical focusing? (ii) What are the resulting asymptotic equations in this critical case? We note that (i) is a question at the linear level. Nonlinear effects will appear at the same level automatically, by setting $\gamma=\epsilon^{3/2}$ in (50)—as this is the right choice to get nonlinear effects comparable in strength with geometrical focusing.

Foreseeing the need to eventually have a scaling in the caustic layer based on a different ''ϵ'' from the one in the WNGO outer expansion (just as we had to do in subsection IV A), let us here differentiate explicitly the small expansion parameter to be used in the caustic layer, by calling it ϵ_i. We keep the symbol ϵ to mean the small parameter in the outer WNGO expansion. Notice that this ϵ appears now explicitly in the governing equations, through the viscous terms (72), because we defined the Reynolds number R_e using the flow parameters outside the caustic.

The situation is very much as in subsection III A, except that now to the right hand sides of the second and third equation in (49) we must add the viscous terms arising from (72). There is not much point in computing these exactly, it is enough to notice that their leading order contributions are

$$\frac{\epsilon}{\epsilon_i^2 R_e}\,\tilde{u}_{\psi\psi} \quad \text{and} \quad \frac{\epsilon}{\epsilon_i^2 R_e}\,\tilde{v}_{\psi\psi},$$

respectively. Then a rather straightforward calculation shows that the critical threshold condition on the Reynolds number is $R_e=O(\epsilon\epsilon_i^{-5/3})$. Thus assume

$$R_e=\nu^{-1}\epsilon\epsilon_i^{-5/3}, \tag{78}$$

and take $\gamma=\epsilon_i^{2/3}$ in the caustic expansion (50) (replacing ϵ by ϵ_i in these formulas). Then the resulting asymptotic governing equations are

$$2\kappa(s)\,\eta\rho_{0\psi}+\tilde{v}_{1\eta}+\nu\rho_{0\psi\psi}=((\beta+1)\rho_0^2)_\psi,$$

$$\tilde{v}_{1\psi}+\rho_{0\eta}=0. \tag{79}$$

We point out that we can apply to these equations the same transformation that lead earlier to (63); though now ν is not arbitrary and is determined by the viscous dissipation. Then we obtain the following normal form of a Nonlinear Viscous Tricomi Equation:

$$(y-u)u_x+v_y+u_{xx}=0, \quad u_y+v_x=0. \tag{80}$$

The following conclusions are immediately obvious from the analysis we just carried out.

- If in the caustic layer we use $\epsilon_i=\epsilon$ and the Reynolds number is critical for the flow outside [so that equation (75) applies], then equation (78) shows that the Reynolds number will be too large [by $O(\epsilon^{-1/3})$] to have viscous effects in the caustic layer. The governing asymptotic equations in the caustic will be inviscid.
- If the Reynolds number is critical for the flow outside [so that equation (75) applies], then equation (78) shows that we need $\epsilon_i=\epsilon^{6/5}$ to get viscous effects in the caustic layer. But this is exactly the same as the value for ϵ_c we arrived at by the inviscid shock arguments in subsection IV A.

Thus, we arrive at the following result: Viscosity [in the critical threshold case (75)] does not affect the basic scaling results that we obtained for the inviscid case earlier. Furthermore, we see the following.

(a) For smooth incident waves, the caustic equations are linear and inviscid (the scalings $\epsilon_i=\epsilon$ and $\gamma=\epsilon^{5/6}$ must be used). The amplification factor at the caustic is $\epsilon^{-1/6}$ and the connection formula (38) involving the Hilbert Transform applies.

(b) For weak shocks, the caustic equations are nonlinear, viscous, and given by equation (79) above. The amplification factor is $\epsilon^{-1/5}$ and the scalings $\epsilon_i=\epsilon^{6/5}$ with $\gamma=\epsilon_i^{2/3}$ must be used in the caustic.

(c) For waves having both shocks and smooth components, we use the same ideas in subsection IV B that lead to the unified description in equation (70): The smooth wave scalings above in (a) are used, but we keep in the equations the higher order terms that become important near shocks (due to the larger amplitudes and gradients there). The resulting equations are

$$2\kappa(s)\eta\rho_{0\psi}+\tilde{v}_{1\eta}+\epsilon^{1/3}\nu\rho_{0\psi\psi}=\epsilon^{1/6}((\beta+1)\rho_0^2)_\psi,$$
$$\tilde{v}_{1\psi}+\rho_{0\eta}=0. \tag{81}$$

The fact that viscosity yields the same scalings as the nonlinear inviscid analysis is a strong argument in favor of the correctness of the theory for caustics of weak shock waves presented in this paper. Much more can be said about the behavior of viscous shocks near a caustic. The behavior of the solutions to (77), in particular, can be shown to display an interesting bifurcation, with viscous shocks away from the caustic behaving either linearly or nonlinearly at the caustic depending on the value of a critical parameter, which involves not just the orders of magnitude, but the actual sizes of the viscosity, the front's amplitude, and its curvature. A detailed discussion, however, goes beyond the scope of this article, and will be postponed to a future publication.

V. THE PROBLEM OF TRIPLE SHOCKS

One of the most striking manifestation of nonlinear behavior in the caustics of weak shock waves is the occurrence of triple shock intersections. In the experimental work of Sturtevant and Kulkarny,[4] they observe the focusing shock reflecting off the caustic as another shock. Incident and reflected waves meet at a point and then continue through as a single shock for a finite distance before decaying. This is particularly clear in Fig. 3f of Ref. 4; see also the numerical calculation in Fig. 31 of Ref. 9. Such nonlinear behavior, which a naive extrapolation of existing theory would not predict, constitutes the main motivation for the present work. We saw in the previous section how we can solve this apparent paradox, by adjusting the scale of the caustic layer of shock waves. The resulting equations (63) for the inner layer are nonlinear, and one would expect that they yield the observed triple shock structure when given boundary conditions corresponding to matching with the incident and reflected waves. This matching procedure should not only determine the structure of the caustic layer, but also the parameters of the reflected wave.

Yet a new problem arises: equations (63) do not admit triple shock intersections, at least not without an additional singularity. This fact, that we prove in subsection V A, is analogous to the one at the core of the von Neumann Paradox of oblique shock reflection. Thus we have solved a paradox just to discover a deeper one! In subsection V B, we comment briefly on the new paradox. A more complete discussion of this intriguing open problem, however, goes beyond the scope of this work.

A. A theorem on triple shocks

In this subsection, we prove that equations (63) do not admit triple-shock intersections separating three states where the solution is continuous. This proof is entirely similar to the one presented in Ref. 9 for the small disturbance unsteady transonic flow equations, which describe the oblique reflection of weak shock waves in the regime where the von Neumann paradox arises.

Consider a triple-shock intersection, such as the one displayed in Fig. 8. We assume that the variables u and v are

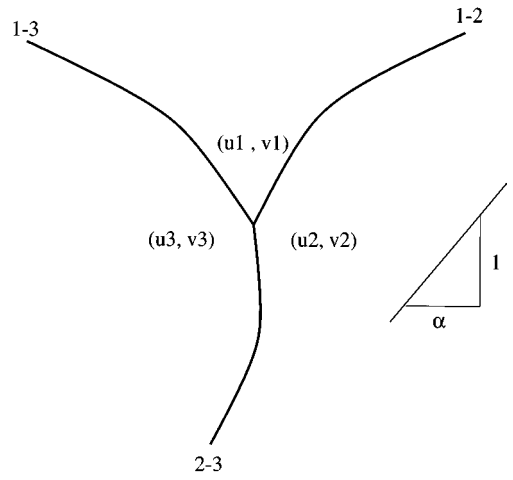

FIG. 8. Three shocks meeting at a point. We prove that such a configuration cannot hold if the three states separated by the shocks are continuous, their limits at the triple point are disjoint, and the shock slopes have limiting values, measured by their inverse α.

continuous in the three subdomains 1, 2, and 3 into which the shocks divide the plane; i.e., we assume that limiting values (u_1,v_1), (u_2,v_2), (u_3,v_3) exist as we approach the point from within each of the three subdomains. We also assume that these three limiting states are disjoint, so none of the shocks ends up with zero strength. Finally, we assume that the slopes of the three shocks also approach limiting values at the triple-shock intersection. We shall show that a triple-shock configuration satisfying these hypotheses cannot be a local solution to equations (63).

Proof: Since by hypothesis the shocks are the most singular part of a solution which is otherwise continuous, the three limiting states and slopes have to satisfy the jump conditions for (63):

$$[v]=\alpha[u], \quad \alpha^2=y-\overline{u}, \tag{82}$$

where the brackets [:] stand for the jump in the enclosed variable across a shock, $\overline{u}$ for the arithmetic mean of the two values of u across the shock, and $\alpha=dx/dy$ for the inverse of the shock's slope. The second of these conditions yields, for the three shocks,

$$\begin{aligned}
\alpha_{12}^2&=y-\tfrac{1}{2}(u_1+u_2),\\
\alpha_{13}^2&=y-\tfrac{1}{2}(u_1+u_3),\\
\alpha_{23}^2&=y-\tfrac{1}{2}(u_2+u_3),
\end{aligned} \tag{83}$$

from which we obtain

$$\begin{aligned}
u_1-u_2&=2(\alpha_{23}^2-\alpha_{13}^2),\\
u_2-u_3&=2(\alpha_{13}^2-\alpha_{12}^2),\\
u_3-u_1&=2(\alpha_{12}^2-\alpha_{23}^2).
\end{aligned} \tag{84}$$

Now we apply the first jump condition in (82) to (84), and obtain

$$v_1 - v_2 = 2\alpha_{12}(\alpha_{23}^2 - \alpha_{13}^2),$$
$$v_2 - v_3 = 2\alpha_{23}(\alpha_{13}^2 - \alpha_{12}^2), \tag{85}$$
$$v_3 - v_1 = 2\alpha_{13}(\alpha_{12}^2 - \alpha_{23}^2).$$

Adding up equations (85), we obtain the following relation between the shock slopes:

$$\alpha_{12}(\alpha_{23}^2 - \alpha_{13}^2) + \alpha_{23}(\alpha_{13}^2 - \alpha_{12}^2) + \alpha_{13}(\alpha_{12}^2 - \alpha_{23}^2) = 0. \tag{86}$$

Now think of (86) as an equation for α_{12}, with roots $\alpha_{12} = \alpha_{13}$ and $\alpha_{12} = \alpha_{23}$. Since the equation is quadratic, these are the only two roots. However, if $\alpha_{12} = \alpha_{23}$, it follows from the jump conditions (82) that $u_1 = u_3$ and $v_1 = v_3$, so in fact we have only one shock. The same applies, of course, if $\alpha_{12} = \alpha_{13}$, and the proof of the theorem is complete.

B. The triple-shock paradox and its relation to the von Neumann paradox of oblique shock reflection

The theorem of the previous subsection closes the circle of this paper in a puzzling way. Let us summarize this puzzle here. We started with the caustics of smooth weakly nonlinear waves. Hunter and Keller showed that the behavior of such caustics is governed by linear theory. For weak shocks, however, this result is inconsistent. Moreover, the experimental results of Sturtevant and Kulkarny display strongly nonlinear behavior near the caustics, with triple-shock intersections as the most prominent nonlinear feature. In section IV, we solved this apparent paradox, showing that the behavior of weak shock waves near the caustic is in fact governed by nonlinear equations. However, the theorem we have just proved shows that these nonlinear equations do not admit triple shocks!

This later development may look a bit disappointing; it could suggest a failure of the asymptotic theory. Yet there is another situation, where triple shocks show up in equations that in principle do not admit them. This is the von Neumann paradox of oblique shock reflection, taking place when a weak shock wave hits a wedge at nearly glancing incidence. For a quite broad range of parameters, triple shocks are observed that the equations of gas dynamics do not appear to support. Moreover, the asymptotic equations describing this phenomenon are very similar to (63); and a proof similar to the one above[9] shows that they cannot hold triple shocks either! Yet numerical experiments with these asymptotic equations match very closely the real scenario; in particular, they do display a local structure with three shocks apparently meeting at a point.

There is more to this analogy. The asymptotic equations describing oblique shock reflection also model the behavior of weak shock waves near arêtes.[9] In this context, they have been solved numerically. These solutions include not only the arête, but also a piece of the caustic nearby. And these caustics have triple shocks at their core, even though the equations being solved do not appear to allow them!

These paradoxical triple shocks are the subject of present study by various groups of researchers, including Brio, Canic, Colella, Gamba, Henderson, Hunter, Keyfitz, Morawetz, Rosales, and Tabak (see for instance Refs. 5, 6, 7, 8, and 9); often with conflicting views. The two leading proposed solutions to the paradox involve, in one case, a shock with zero strength at the triple point, and in the other, an extra singularity which makes one of the states between shocks not continuous at the triple point. The latter scenario is the one favored by the authors; however, none of the proposed solutions is by any means complete, and much work remains to be done on this elusive yet fascinating problem. We believe that, whichever the solution to the problem in the case of oblique reflection of shocks, it will apply to the caustics as well. We hope to find such solution one day; however, we will not attempt to find it here.

VI. CONCLUSIONS

The caustics of weak shock waves have been studied through matched asymptotic expansions. These caustics have been found to be thinner and correspondingly more intense than those of smooth waves with comparable amplitudes. This difference in scales gives rise to a nonlinear set of equations in the caustic layer, thus solving a contradiction that would have the caustics of weak shocks behave linearly, even though linear theory predicts an infinite amplification for these shocks.

The new nonlinear scales have been shown to survive the addition of a small amount of viscosity to the equations. In addition, a unified framework has been introduced, which simplifies the study of the caustics of shock waves immersed within smooth weakly nonlinear waves.

Finally, an outstanding open problem has been reported, which connects the nonlinear behavior near the caustics of weak shock waves with the von Neumann paradox of oblique shock reflection. Both problems have three weak shocks apparently meeting at a point, even though the equations describing this local structure do not admit triple-shock intersections without an additional singularity.

ACKNOWLEDGMENTS

The work of Rosales was partially supported by National Science Foundation (NSF) Grant No. DMS-9311438. The work of Tabak was partially supported by the NSF Grant No. DMS-9501073.

[1] R. N. Buchal and J. B. Keller, ''Boundary layer problems in diffraction theory,'' Commun. Pure Appl. Math. **XIII**, 85 (1960).
[2] D. Ludwig, ''Uniform asymptotic expansions at a caustic,'' Commun. Pure Appl. Math. **XIX**, 215 (1966).
[3] J. K. Hunter and J. B. Keller, ''Caustics of Nonlinear Waves,'' Wave Motion **9**, 429 (1987).
[4] B. Sturtevant and V. A. Kulkarny, ''The focusing of weak shock waves,'' J. Fluid Mech. **73**, 641 (1976).
[5] M. Brio and J. K. Hunter, ''Mach reflection for the two-dimensional Burgers equation,'' Physica D **60**, 194 (1992).
[6] P. Colella and L. F. Henderson, ''The von Neumann paradox for the diffraction of weak shock waves,'' J. Fluid Mech. **213**, 71 (1990).
[7] S. Čanić and B. L. Keyfitz, ''Riemann Problems for the two-dimensional unsteady transonic small disturbance equation,'' SIAM J. Appl. Math (in press).
[8] C. S. Morawetz, ''Potential theory for regular and Mach reflection of a shock at a wedge,'' Commun. Pure Appl. Math. **45**, 1 (1993).
[9] E. G. Tabak and R. R. Rosales, ''Focusing of weak shock waves and the

von Neumann paradox of oblique shock reflection,'' Phys. Fluids **6**, 1874 (1994).
[10]J. B. Keller, ''Rays, waves and asymptotics,'' Bull. Am. Math. Soc. **84**, 727 (1978).
[11]M. J. Lighthill, ''Reflection at a laminar boundary layer of a weak steady disturbance to a supersonic stream, neglecting viscosity and heat conduction,'' Q. J. Mech. Appl. Math. **III**, 303–325 (1950).
[12]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[13]Y. Choquet-Bruhat, ''Ondes asymptotiques et approachées pour des systèmes d'équations aux dérivées partielles non linéaires,'' J. Math. Pures Appl. **48**, 117 (1969).
[14]J. K. Hunter, ''Nonlinear geometrical optics,'' Multidimensional Hyperbolic Problems and Computations, The IMA Volumes in Mathematics and its Applications, edited by J. Glimm and A. Majda (Springer Verlag, New York, 1991), Vol. 29, pp. 179–197.
[15]R. R. Rosales, ''An Introduction to weakly nonlinear geometrical optics,'' in Ref. 14, pp. 281–310.
[16]J. K. Hunter, A. Majda, and R. R. Rosales, ''Resonantly interacting weakly nonlinear hyperbolic waves II: Several space variables,'' Stud. Appl. Math. **75**, 187 (1986).