

On the method of caustics: An exact analysis based on geometrical optics

ARES J. ROSAKIS and ALAN T. ZEHNDER

Graduate Aeronautical Laboratories, California Institute of Technology, Pasadena, CA 91125, USA

(Received April 9, 1984)

Abstract

The method of caustics, as developed and applied to fracture mechanics over the last twenty years, contains several approximations which limit its applicability. In this paper the development of caustics is reviewed and the implicit assumptions made in the past are clarified and discussed. The exact equations are derived for caustics formed by the reflection of light from a general surface. In addition, the conditions for the formation of a caustic curve are derived and explained in detail. Numerically generated shadow spots are given for the case of light reflected from a surface deformed due to the presence of a plane stress, mode-I, elastic crack. Attention is focused on the near tip region where severe deformation gradients violate the assumptions made by previous approximate analyses. The results demonstrate significant deviation from the approximate analyses resulting in errors as large as 15% in the determination of the stress intensity factor from shadow spot measurements.

1. Introduction

In the past few decades, a number of optical methods based on the principles of light wave interference have been introduced for the study of stress and strain fields in elastic transparent materials. By their application a number of important advances in the study of the deformation of solids have been achieved [1,2,3]. However these methods are restricted generally to transparent materials or to opaque materials with transparent coatings [4]. More recently the optical method of caustics, a technique based on geometrical optics, was the first one to be successfully applied to the study of deformation fields in opaque solids. The advantage of caustics relative to other optical experimental techniques is that the method can be used in either a reflection or a transmission arrangement and thus can be applied to the investigation of both transparent or opaque materials.

In transmission the technique can be used to record changes in the refractive index caused by the presence of stress or density fields. Transmitted caustics can therefore yield information regarding the stress or density fields, provided that these fields can be related to the changes in the refractive index. Similarly, caustics in reflection can be used to record the changes in elevation which are introduced to an initially flat and reflective surface by the establishment of a stress field. Reflected caustics can therefore

yield useful information regarding the stress state below the reflecting surface, provided that the surface deformation can be related to the existing stresses.

In this paper, our attention is mainly focused on the study of caustics obtained by reflection of parallel light rays from the mirrored surface of a solid. However, a special effort has been made to keep the analysis as general as possible. As a result, the main concepts presented here, with minor adjustments, hold for both the cases of reflection (opaque solids) and transmission (transparent solids or fluids).

In the fluid mechanics literature, transmission caustics are usually referred to as "shadowgraphs" and have been applied to a number of interesting applications, e.g. qualitative study of compressible flow [5]. In solid mechanics, the method is customarily referred to as either the shadow-spot technique, or the method of caustics. In solids, caustics have been initially introduced by Schardin [6] and Manogg [7] and in recent years have been applied to a number of interesting fracture mechanics problems by a variety of investigators [8,9,10,11,12].

Manogg, who was the first to apply the technique quantitatively, used caustics in a transmission arrangement. He was able to record changes in the optical path of rays traveling through transparent material at the vicinity of a crack tip, where the elastic stress field introduces changes in the refractive index as well as changes of thickness. The resulting difference in optical path produces a caustic pattern on a screen placed behind the specimen. He showed that the geometrical characteristics of the caustic depend on the nature and intensity of the crack tip singularity and was able to measure the intensity of the near-tip stress field.

Manogg's result was later applied to the study of reflected caustics. His analysis, obtained on the basis of a transmission arrangement, was assumed to describe both transmission and reflection shadowgraphs. The adaptation of Manogg's equation for the reflection problem involved a number of simplifying assumptions [8,9] (to be discussed in detail later). These assumptions prove to be useful in cases involving small angles of incidence but limit the generality of the result and impede the complete understanding of the optical process. This seems to be particularly true at the vicinity of steep gradients in the reflector shape, areas that are usually of special interest to most investigators. Unfortunately, in spite of such limitations, no effort has been made to understand completely the mechanism of the optical mapping and in most of the applications of the method the approximate adaptations of Manogg's results were indiscriminately applied without examination of the validity of the assumptions under which they were initially derived.

In the first part of this paper a three dimensional analysis is performed and the complete equations of the optical mapping are established for the first time. No assumptions regarding the nature of reflector or the optical set-up are made. For the cases of light reflected from regions of abrupt changes in optical path created by singularities such as point force, stress waves, or singular crack tip displacement fields, the notion of an interruption curve is introduced. The interruption curve is defined as the locus of points surrounding the singularity and including an area which will not contribute to the formation of an optical image. The curve represents an upper limit to how close one can "approach" the singularity using the optical method of caustics, since the information obtained by reflection from points within this curve is lost.

In the second part of this paper, the condition for the formation of a caustic curve is re-examined. The accepted condition, which seems to appear again and again in the

literature is never adequately explained [8,9]. An explanation in terms of the intensity of light reaching the image plane (screen) and producing the optical pattern is provided here. It is shown that the intensity of light reaching the "screen" at a point P , is inversely proportional to the Jacobian determinant of the mapping equations derived in the first part of the section. The caustic curve thus corresponds to reflections from points on the specimen for which the Jacobian vanishes. The vanishing of the Jacobian produces infinite intensity in the screen plane and the existence of bright curves (the caustics) is thus predicted. More generally, it is noted that once the equation of the reflecting surface is given, the value of the Jacobian is completely determined and it can be used to provide the intensity of light at any point on the screen.

In the third part of this paper, shadow patterns are produced numerically using the complete equations of the optical mapping. The shadow patterns are then compared to the patterns obtained using the approximate mappings. Through this comparison, the errors involved in using the approximate mappings for fracture mechanics applications are investigated. Of particular interest is the region very near to a plane-stress crack tip where the steep out of plane displacement gradients result in large deviations from the customarily used simplified caustic equations.

2. Discussion of previous work on caustics

In the original work on transmitted caustics [7] there are implicit assumptions which limit the applicability of the results. These assumptions have been carried through in most of the subsequent applications of caustics [8,9,10] with very little or no explanation. In this section the approximations that appear in the caustics literature are discussed in detail.

Consider a planar medium lying in the x_1, x_2 plane at $x_3 = 0$. The medium is such that it causes a non-uniform change in the optical path of light transmitted through it, or reflected from its surface. For a transparent material the change in optical path is due to non-uniform changes in thickness of the medium and also due to gradients in the index of refraction of the material. For an opaque material the change in optical path is due to a non-uniform surface elevation of the medium.

Consider further light travelling in the $-x_3$ direction normally incident on the medium at $x_3 = 0$, as illustrated in Fig. 1. This is equivalent to a family of plane waves incident on the medium. A light ray is parallel to the vector normal to a surface representing wave fronts of light. If $S(x_1, x_2, x_3)$ represents the optical path of the light ray, then the wave front is given by $S(x_1, x_2, x_3) = \text{Const}$. The vector $\nabla S(x_1, x_2, x_3)$, where ∇ is the gradient operator, is then normal to S and thus parallel to the light ray passing through S at point (x_1, x_2, x_3) . For a transparent material, the plane waves travel through the material and are distorted due to the introduced variations in optical path. This causes light rays passing through the medium to be deflected. (Equivalently, waves reflected from the surface of an opaque solid are distorted due to nonuniform surface elevation.) If a screen is placed at a distance z_0 behind the medium, then the light ray intersecting it at the point $x = (x_1, x_2)$ will be mapped to a point $X = (X_1, X_2)$ on the screen. The (X_1, X_2) coordinate system is identical to the (x_1, x_2) system, except that the origin of the former has been translated by $x_3 = -z_0$. Assuming that the medium is of infinitesimal thickness and using the geometry of Fig.

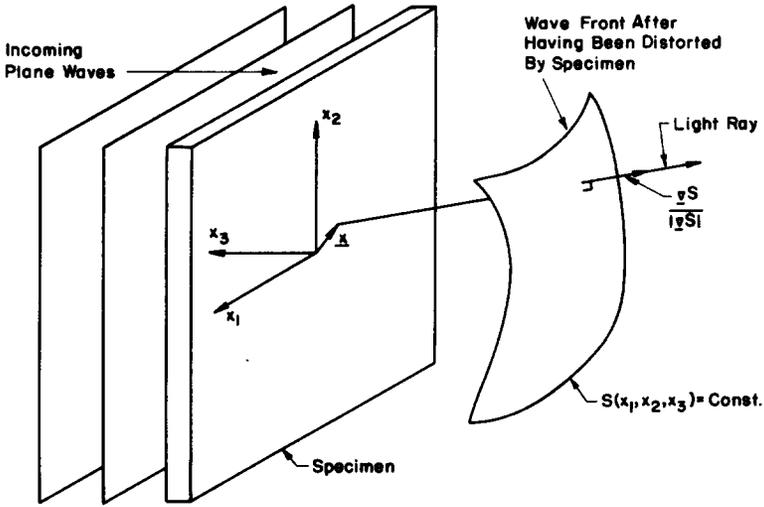


Figure 1. Deflection of light rays by specimen.

2, we see that the mapping is given by

$$X = x - z_0 \left(\frac{\nabla S(x_1, x_2, 0)}{(\nabla S(x_1, x_2, 0), e_3)} - e_3 \right) \tag{2.1}$$

where e_3 is the unit vector along the x_3 axis.

In the original work by Manogg it was assumed that S is described by a family of self-similar surfaces translated in the $-x_3$ direction as in Fig. 3. Thus, according to his

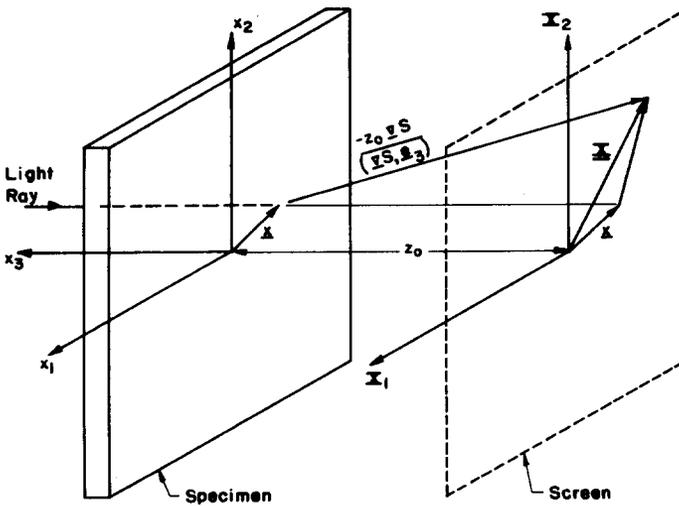


Figure 2. General optical mapping.

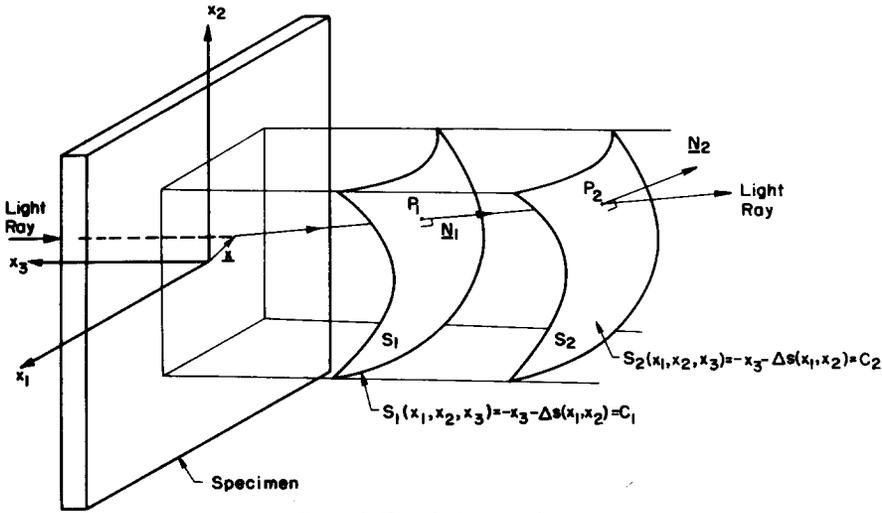


Figure 3. Translating wavefronts.

assumption, S is given by

$$S(x_1, x_2, x_3) = -x_3 - \Delta s(x_1, x_2) = \text{Const.} \tag{2.2}$$

where $\Delta s(x_1, x_2)$ represents the change in optical path introduced by the medium. The limitations of this assumption become evident if a light ray is followed from a surface S_1 to another, S_2 , as in Fig. 3. The light ray parallel to the vector normal to S_1 at point P_1 intersects S_2 at P_2 . However the normal at P_2 is not parallel to the normal at P_1 and thus does not coincide with the direction of the light ray. Only for very small angles of deflection, will the normal at P_2 be parallel to the normal at P_1 . Thus Manogg's assumption implicitly limits the applicability of his results to small ray deflections. As is evident from Fig. 4, the wave fronts are actually a family of expanding waves, resulting in light rays which are always normal to the surface. For very large radii of curvature, i.e. very small deflections, expanding and translating waves are coincident and S may be approximated by equation (2.2).

With Manogg's assumptions and the further assumption of a medium of infinitesimal thickness, the mapping of light rays is found by substituting equation (2.2) into (2.1) and is given by

$$X = x - z_0 \nabla(\Delta s(x_1, x_2)), \tag{2.3}$$

which is Manogg's original result and is the equation used in the normal evaluation of transmitted caustics for fracture mechanics applications.

The application of caustics to the study of opaque materials requires the reflection of light from the highly polished surface of a planar solid. Theocarlis [9] applied Manogg's equation to caustics in reflection by assuming that the change in optical path is given by

$$\Delta s(x_1, x_2) = 2f(x_1, x_2), \tag{2.4}$$

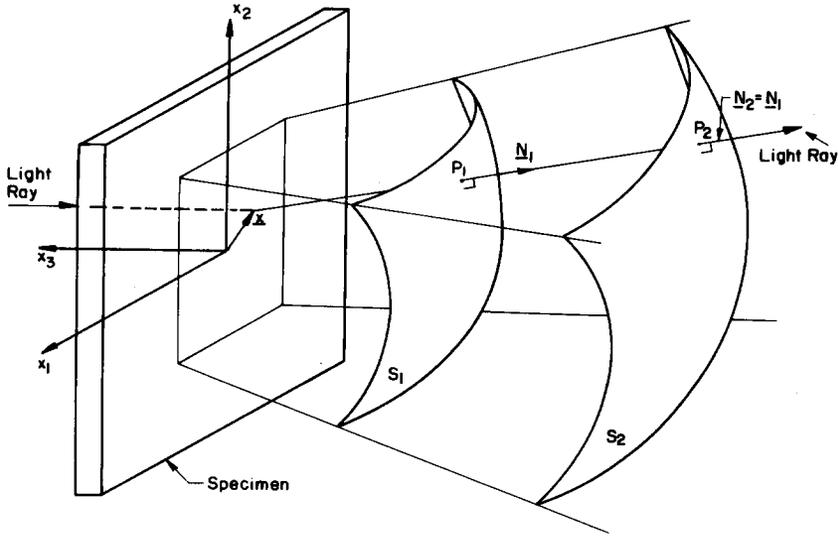


Figure 4. Expanding wavefronts.

where $x_3 = -f(x_1, x_2)$ is the equation of the specimen surface. Substituting equation (2.4) into (2.3) yields

$$X = x - 2z_0 \nabla f(x_1, x_2). \tag{2.5}$$

Equation (2.4) introduces an additional approximation since the actual change in optical path is greater than $2f$, due to the angular deflection undergone.

In cases of severe gradient in optical path, the light rays reflected from or transmitted through the medium deviate strongly from parallelism. In such cases Manogg's and Theocaris' assumptions lose validity and a more accurate analysis is required. The exact analysis for reflected caustics is given in the following section.

3. Reflected caustics: The mapping equations

In this section, the particular example of caustics obtained by reflection will be considered. Shadowgraphs obtained by reflection are associated with changes in the optical path of the light rays, introduced due to the nonuniform elevation of the reflector surface. The equations describing the optical mapping of points of the reflector on to points of the image plane (shadowgraph plane), with minor modifications, also describe the transmission shadowgraphs. The basic difference between reflection and transmission is in the nature of the optical path change, which in the latter case is caused by changes in the refractive index of the transparent medium.

Consider a family of light rays, parallel to the x_3 axis, incident on the reflective surface $x_3 = -f(x_1, x_2)$ of an opaque material, see Fig. 5. The reflected light field is recorded by a camera positioned in the front of the reflecting surface. The focal plane of this camera, which shall be called the "screen" for convenience, is located behind the reflector and intersects the virtual extensions of the reflected rays at the plane $x_3 = -z_0$ where z_0 is positive. In the most general case, z_0 is positive for a virtual image (focal

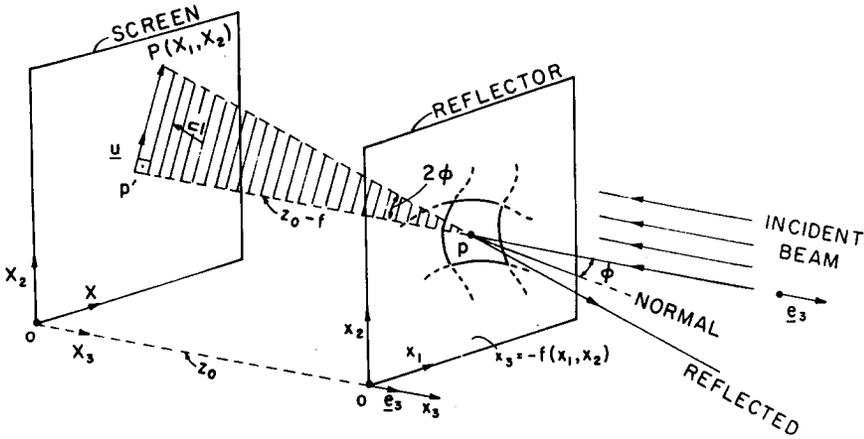


Figure 5. Optical mapping for reflected caustic.

plane is located behind the reflector) and negative for a real image (focal plane is located in front of the reflector.).

Consider further an incident light ray which intersects the reflecting surface at $x_3 = -f$. After reflection from point $p(x_1, x_2)$ on the polished surface, the ray will deviate from parallelism. The virtual extension of the resulting reflected ray will then intersect the screen at the virtual image point $P(X_1, X_2)$ whose position will depend on the slopes of the reflecting surface at $p(x_1, x_2)$ and on the normal distance z_0 . The X_1, X_2 coordinate system is identical to the x_1, x_2 system, except that the origin of the former has been translated the distance z_0 to the screen.

If φ is the angle between the incident light ray and the normal to the reflector at $p(x_1, x_2)$ and p' is the normal projection of p on the "screen", then the position vector of the image point P will be given by

$$X = x + [(z_0 - f) \tan 2\varphi] u \tag{3.1}$$

where

$$X = X_i e_i, \quad x = x_i e_i$$

and $u = \overline{p'P} / |\overline{p'P}|$ is the unit vector along $\overline{p'P}$.

Also, if N is the unit normal to the surface $x_3 = -f(x_1, x_2)$ at the point $p(x_1, x_2)$, then N is equal to $\nabla F / |\nabla F|$ where $F(x_1, x_2, x_3) = x_3 + f(x_1, x_2)$. Further, if e_3 denotes a unit vector in the x_3 direction (direction of light incidence), the plane which contains the incident light ray and the corresponding reflected ray will be defined by the normal unit vector $n = N \times e_3 / |N \times e_3|$. The intersection of this plane with the plane of the "screen" is then the line which has a direction specified by the unit vector u given by

$$u = n \times e_3 = - \frac{\frac{\partial f}{\partial x_1} e_1 + \frac{\partial f}{\partial x_2} e_2}{\left[\left(\frac{\partial f}{\partial x_1} \right)^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \right]^{1/2}} = - \frac{\nabla f}{|\nabla f|} \tag{3.2}$$

Also, the tangent of the angle φ between the incident ray and the normal to the reflecting surface can be expressed as

$$\tan \varphi = \frac{\sin \varphi}{\cos \varphi} = \frac{|\mathbf{e}_3 \times \mathbf{N}|}{|\mathbf{e}_3 \cdot \mathbf{N}|} = \left[\left(\frac{\partial f}{\partial x_1} \right)^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \right]^{1/2} = |\nabla f|. \quad (3.3)$$

By substituting equations (3.2) and (3.3) into (3.1), the position vector of the image point P of the "screen" can be written as¹

$$\mathbf{X} = \mathbf{x} - 2(z_0 - f) \frac{\nabla f}{1 - |\nabla f|^2}. \quad (3.4)$$

The above expression describes the optical mapping of points $p(x_1, x_2)$ of the reflector on to points $P(x_1, x_2)$ of the "screen". The choice of the sign of z_0 depends on whether the image is real or virtual. The use of a virtual image (positive z_0) and a reflection arrangement, has certain advantages in experimental solid mechanics as discussed in [14,15].

Equation (3.4) can be simplified by introducing a number of assumptions based on the nature of the changes in optical path introduced by the medium studied. For z_0 sufficiently large, it can be accurate and useful to assume that $(\partial f / \partial x_i)^2 \ll 1$ and that $f(x_1, x_2)$, the magnitude of the optical retardation introduced by the medium, is small compared to z_0 . These assumptions reduce the equation of the mapping to the following form as indicated in equation (2.5)

$$\mathbf{X} = \mathbf{x} - 2z_0 \nabla f,$$

which is the approximate expression given by Theocaris.

Although equations (2.5) seem to give satisfactory results in many cases of practical interest, the complete understanding of the reflection process requires the introduction of the general form (3.4) of the mapping equations. This becomes obvious by considering the second term in (3.4) which becomes singular for $|\nabla f|^2 = 1$. This singularity is artificially eliminated in the approximate form of the equations used so far.

The condition

$$|\nabla f|^2 = 1 \quad (3.5)$$

defines a curve on the plane of the medium considered whose points will map at ∞ on the reference plane screen. Points inside this curve will not be reflected back and will not contribute to the formation of the optical image on the "screen".

The curve defined by the equation $|\nabla f|^2 = 1$ is introduced here for the first time, and will be referred to in this paper as the *interruption curve*. For cases in which singularities in the optical path difference introduced by a medium are present, the interruption curve places an upper limit to how close one can approach the singularity using the optical method of caustics, since information from points within the curve is lost or interrupted.

The physical meaning of the condition can be understood by referring to equation (3.3). Setting $|\nabla f|^2 = 1$ in (3.3) is equivalent to setting $\tan \varphi = \pm 1$ or $\varphi = \pm \pi/4$. An angle of incidence of $\pm \pi/4$ will result in a reflected ray parallel to the x_1, x_2 plane

¹ An early attempt to present the accurate optical mapping equation was given in [13]. Unfortunately this analysis was two-dimensional resulting in a term missing in the denominator of expressions (3.4).

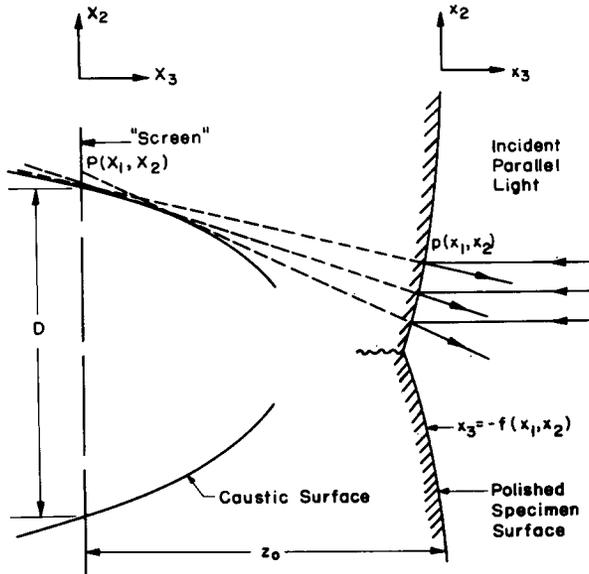


Figure 6. Formation of caustic envelope upon reflection.

whose virtual extension will not intersect the “screen”. Similarly, slopes larger than one will produce reflections that will not contribute to the formation of an optical image.

4. The formation of caustic curves and shadow spots

Changes in optical path caused by reflection or refraction of parallel light will deflect the light from its original course. Depending on the nature of the optical retardation, regions of high intensity (caustic) or areas of low intensity (shadow spots) will be created on a “screen” placed perpendicular to the initial ray direction, see Fig. 6.

In the following section, the conditions necessary for the formation of caustic curves are established. The analysis is carried out in reference to a reflection arrangement but the general results hold for both reflection and transmission as long as the appropriate mapping equations are used. (Equations (3.4) must be used for reflection and equations (2.3) for transmission.)

Consider light of intensity $\xi(x_1, x_2)$ incident on the reflective surface of an opaque solid. After reflection, the rays will deviate from parallelism and the reflection process will be described by the mapping equations (3.4) derived in the previous section. For perfect reflectivity, the light energy incident on an area D of the highly polished specimen is an invariant of the transformation. Let D' be the image area on the transformed plane “screen” corresponding to reflections from the area D . Conservation of light energy requires that the total energy E incident onto the area D must be equal to the total light energy E' , reaching the area D' of the transformed plane. Thus,

$$E = \iint_D \xi(x) dx_1 dx_2 = E' = \iint_{D'} \Xi(X) dX_1 dX_2 \tag{4.1}$$

where $\Xi(X)$ represents the light intensity field on the "screen".

Using the equations of the mapping (3.4) and changing variables in the right hand side of (4.1), we obtain

$$E = \int \int_D \xi(x) dx_1 dx_2 = E' = \int \int_D \Xi(X(x)) |J(x)| dx_1 dx_2 \quad (4.2)$$

where $J(x) = \det(\nabla X)$ is the Jacobian determinant of the transformation equation (3.4).² Since expression (4.2) is true for an arbitrary area D , the integrands involved can be equated, giving

$$\Xi(X(x, f)) = \xi(x)/J(x) = \xi(x)[\det \nabla X]^{-1}. \quad (4.3)$$

With the equations of the mapping, the Jacobian determinant can be written as

$$J(x) = \det(\nabla X) = \det \left\{ I - 2\nabla \left[(z_0 - f) \frac{\nabla f}{1 - |\nabla f|^2} \right] \right\}, \quad (4.4)$$

where I is the identity tensor.

Once the equation of the reflecting surface $x_3 = -f(x_1, x_2)$ is given, the value of the Jacobian is completely determined and equation (4.3) can be used to provide the intensity of light at any point on the "screen". The entire image on the screen can thus be predicted. Since the intensity of light reaching the "screen" is inversely proportional to the Jacobian, the vanishing of J in (4.3) corresponds to infinite intensity of light on the "screen" plane and the existence of highly luminous curves (caustics) is thus predicted.

The caustic curves therefore correspond to reflections from points on the reflecting surface for which the Jacobian determinant vanishes. The loci of points on the reflector on which $J = 0$ are called the initial curves and are characterized by the property that rays reflected from their immediate neighborhood are responsible for the generation of the caustic curves. The resulting caustics obtained on the screen are regions of multiple reflection, that is for those points on the caustic curve the mapping (3.4) is not invertible and the determinant of the Jacobian matrix is expected to vanish.

Note that J in equation (4.4) depends parametrically on z_0 and thus the initial curve depends on z_0 . In the discussion on the numerical work that follows, z_0 is varied thus varying the initial curve from a curve far from the crack tip to one very near the crack tip.

As indicated earlier the method has potential for use in the study of abrupt changes in optical path introduced by singularities in stress, density, etc. It is of interest therefore to carry the investigation one step further and consider some general features of the reflection process in the vicinity of a singularity on the reflecting surface or more generally a singularity in the additional optical path introduced in a medium either by refraction or reflection.

At points far away from such a singularity (as an example consider a crack tip) the reflector can be considered almost flat and $\nabla f \rightarrow 0$. The second term in (4.4) is very small, so $J \rightarrow \det\{I\} = 1$, and $\Xi \rightarrow \xi$. Thus, if the intensity of the incident light is uniform, $\xi(x) = \xi_0$, then uniform intensity of light is predicted on the screen, see Fig. 7.

² In transmission expression (3.4) must be replaced by equations (2.5).

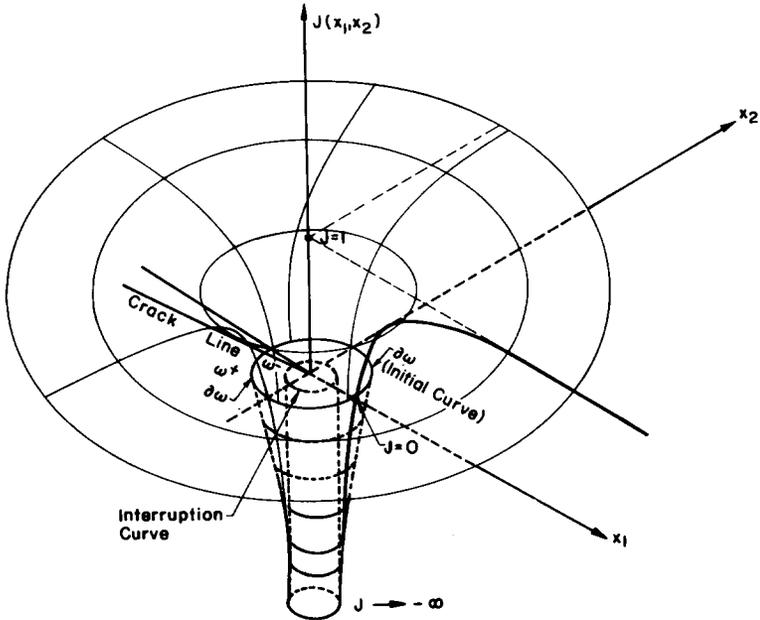


Figure 7a. Jacobian of the mapping.

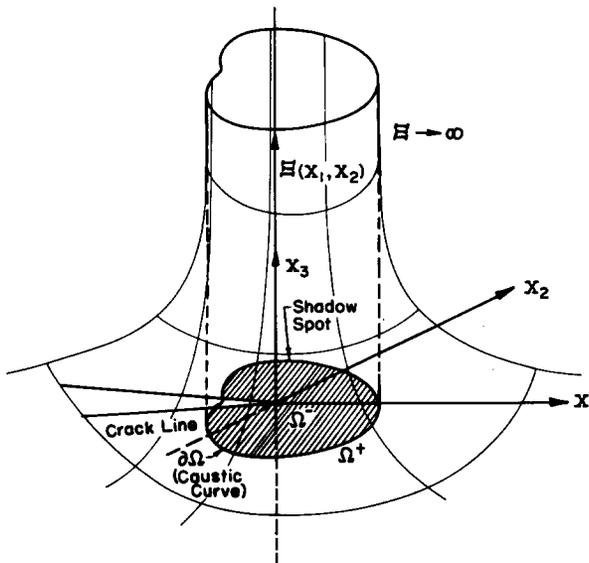


Figure 7b. Intensity of light on screen.

As the singularity (crack tip) is approached, the value of J decreases and the light intensity on the “screen” increases. At some point J vanishes. The place where that happens is referred to as the initial curve. The intensity corresponding to reflections from the initial curve is infinite (4.3) and a caustic is the formed, see Figs. 7a, b.

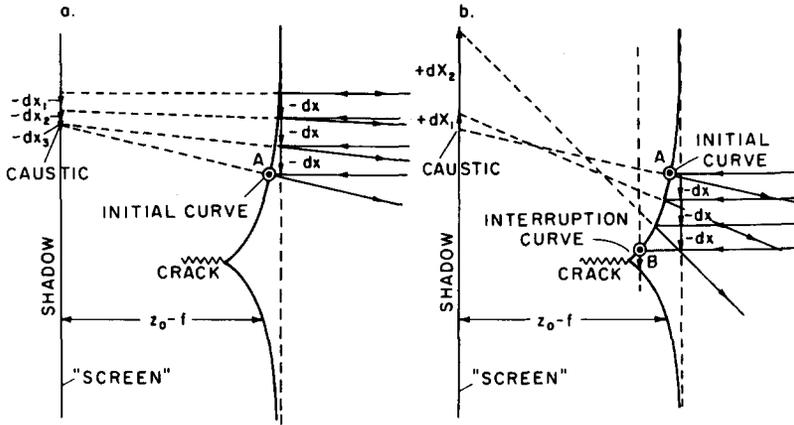


Figure 8. Folding of the optical mapping.

At points just inside the initial curve, the determinant in (4.4) takes small negative values. As we move closer to the center of the singularity, J decreases further and in the vicinity of the interruption curve ($|\nabla f|^2 = 1$) it tends to $-\infty$. (One can see that by referring to (4.4) and observing that J is singular in the vicinity of the interruption curve.)

It is worth noting at this point that at the initial curve there is a reversal of the sign of the Jacobian. This reversal implies "folding" of the mapping. The rays reflected from points outside the initial curve map outside the caustic curve. The rays reflected from points on the initial curve, map directly onto the caustic and the rays reflected from points inside the initial curve again map outside the caustic. No ray will reach the screen at points inside the caustic and a dark region, the shadow spot, will be created.

In order to investigate the mechanism of reflection more completely, let us consider a two-dimensional reflector $z = -f(x)$ as shown in Fig. 8. Assume that the reflector has a unit depth and that the incident light has a constant intensity represented here by the constant spacing dx of the incoming rays.

For this specific case,

$$J = \frac{dX}{dx}$$

where dX is the spacing of the virtual extensions of the reflected rays reaching the "screen". As shown in Fig. 8a, at points away from the singularity $dX \rightarrow dx$ and $J \rightarrow 1$. As the initial curve is approached, dX decreases and eventually is reduced to zero. Decreased spacing of rays represents high light intensity and a caustic curve is obtained by reflections coming from the vicinity of the initial curve. So far, for points outside the initial curve the mapping is one to one and all points are mapped outside the caustic curve. Right on the initial curve, all points are mapped directly on to the caustic. As we move inside the initial curve, the Jacobian changes sign, and a "folding" of the mapping is observed. That means that, as shown in Fig. 8b, points inside the initial curve map again outside the caustic. At points just inside the initial curve, the value of J is small and negative (small spacing of rays on "screen"). As we move further in, J becomes larger in magnitude (increased spacing dX) until it reaches the

value of $-\infty$ at the interruption curve (point where the slope of the reflector is $\pm 45^\circ$). The main features of the mapping can therefore be summarized as follows:

- i. Points on the reflecting surface lying outside the initial curve will map onto points of the “screen” lying outside the caustic.
- ii. Points on the initial curve will map directly on the caustic.
- iii. Points lying in the region between the initial and the interruption curve will map again outside the caustic.
- iv. Points inside or on the interruption curve will not be mapped on the screen.

In other words, there are no points on the reflector that will map inside the caustic curve, and thus a dark region (shadow spot) surrounded by a highly luminous curve (caustic) is formed on the “screen”.

Letting $\partial\omega$ denote the initial curve defined in the reflector plane (x_1, x_2) and $\partial\Omega$ denote the caustic curve defined in the screen plane X_1, X_2 , we can write

$$\begin{aligned} x \in \omega^+ \cup \partial\omega \cup \omega^- &\rightarrow X(x) \in \Omega^+ \cup \partial\Omega \\ x \in \partial\omega &\rightarrow X(x) \in \partial\Omega \end{aligned}$$

which implies

$$X^{-1}(\Omega^-) = \phi$$

where ω^+ represents the region outside the initial curve, ω^- represents the annular region between the initial and the interruption curves and Ω^+, Ω^- are the regions outside and inside the caustic respectively see Fig. 7.

From the above it is clear that the total intensity at a point $P(X_1, X_2)$ on the “screen” is given by

$$\Xi_{\text{tot}} = \xi \left(\frac{1}{|J|_{x \in \omega^+}} + \frac{1}{|J|_{x \in \omega^-}} \right). \tag{4.5}$$

The two terms represent the contribution to the intensity at $P(X_1, X_2)$ from rays reflected from regions both outside and inside the initial curve. For example, the intensity at a point far away from the optical singularity on the screen will come from rays reflected at points very far from the deformation singularity where $J_{x \in \omega^+} \rightarrow 1$ and from points very close to the interruption curve where $J_{x \in \omega^-} \rightarrow -\infty$. Thus from (4.5) $\Xi_{\text{tot}}(X)$ far away $\rightarrow \xi(x)$ (no optical distortion).

At points closer to the optical singularity both terms increase and they both become infinite on the caustic where contributions coming by reflection from both just inside and just outside the initial curve are added.

Mappings of the nature discussed above have been experimentally observed in a number of different situations. For the case of a crack in a planar metallic specimen, the change in optical path is produced by a nonuniform contraction in the thickness direction due to the existence of singular strains at the crack tip [8,9,10]. The type of optical pattern obtained is shown in Fig. 9. Far from the singularity we see a matrix of uniform intensity where the Jacobian is very close to one. As the tip is approached, the value of the Jacobian decreases, the light intensity increases, and a highly luminous curve (the caustic) surrounding a dark area (shadow spot) is observed as predicted by the analysis.

A closer observation of Fig. 9 will reveal a set of “fringes” surrounding the caustic curve. The existence of the fringes implies phase interference and consequently multiple

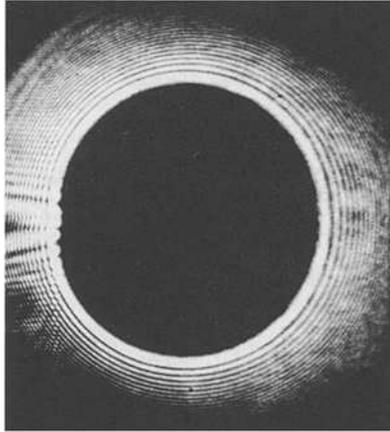


Figure 9. Caustic formed by reflection of coherent light from the vicinity of a mode-I elastic crack. Interference fringes demonstrate folding of the mapping.

mapping. The phenomenon of the “folded” mapping discussed above can be used to explain the existence of the fringes as follows.

The light source used to produce the photograph shown in Fig. 9, was a laser which emits monochromatic, single phase light. As explained above, the illumination outside the caustic curve results from a double reflection or mapping. That is, light waves reflected from both inside and outside the initial curve on the reflector strike the “screen” outside the caustic. Because of the deformation of the specimen surface, however, the light rays reflected from inside the initial curve travel a distance different from that traveled by the rays reflected from outside the initial curve. This difference in path length leads to a difference in phase at the screen which results in the observed interference pattern.

A similar phenomenon has been observed in cases of optical patterns obtained by transmission of laser light through a compressible fluid (shadow graph). Abrupt changes in pressure and consequently density, created for example by shock waves, change the refractive index and create singularities in the additional optical path imposed on the light by the fluid. As in the case of reflection, a bright region is observed along the shocks bordering a dark line, the caustic. Interference fringes are observed outside the caustic line as predicted by the phenomenon of the “folding” of the mapping discussed above.

5. The simplified application of caustics to fracture mechanics

Consider a highly polished planar specimen of uniform thickness d in the undeformed state, occupying a region of the x_1, x_2 plane. The specimen contains an edge crack and is subjected to traction and/or displacement boundary conditions, see Fig. 10. When the loading is applied, the resulting change in thickness is nonuniform at the vicinity of the crack tip and the equation of the deformed specimen surface is given by

$$x_3 = u_3(x_1, x_2) = -f(x_1, x_2) \quad (5.1)$$

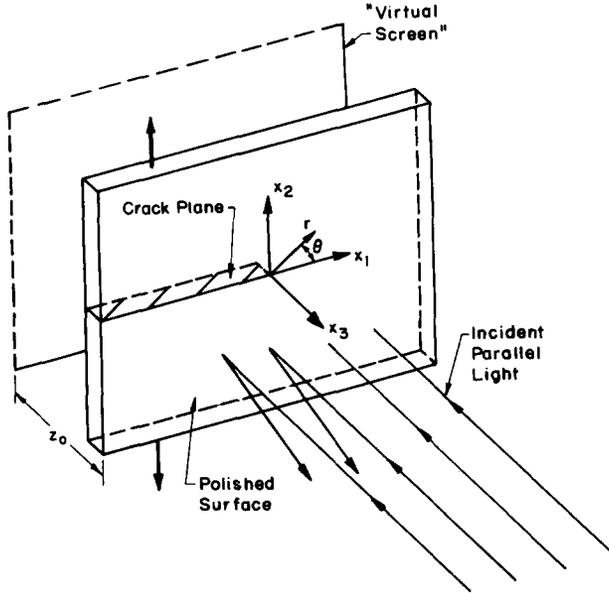


Figure 10. Specimen and optical set-up.

where u_3 is the displacement in the thickness direction. For a plane-stress, elastic, Mode-I crack, the displacement u_3 of the initially flat specimen is given by

$$u_3(x_1, x_2) = -\frac{\nu K d}{E\sqrt{2\pi r}} \cos \frac{\vartheta}{2}, \quad (5.2)$$

where ν = Poisson's ratio, K = stress intensity factor, E = modulus of elasticity, d = thickness of specimen, $r = \sqrt{x_1^2 + x_2^2}$ and $\vartheta = \tan^{-1} x_2/x_1$.

Substitution of the above expression for $f(x_1, x_2)$ in the approximate equation (2.5) provides the equations customarily used for the interpretation of reflected caustics. These equations, expressed in polar form are given by [7]:

$$\begin{aligned} X_1 &= \frac{r_0}{3} \left[3 \cos \vartheta \pm 2 \cos \frac{3\vartheta}{2} \right], \\ X_2 &= \frac{r_0}{3} \left[3 \sin \vartheta \pm 2 \sin \frac{3\vartheta}{2} \right], \end{aligned} \quad -\pi < \vartheta \leq \pi, \quad (5.3)$$

where r_0 is the radius of the circular initial curve and is given by

$$r_0 = \left(\frac{3d\nu K |z_0|}{2\sqrt{2\pi} E} \right)^{2/5}. \quad (5.4)$$

Equations (5.3) are the parametric equations of an epicycloid, see Fig. 11, where a positive sign indicates a real image (screen behind specimen, i.e. $z_0 > 0$) and the negative sign indicates a virtual image (screen in front of specimen, i.e. $z_0 < 0$). The details of the derivation of equations (5.3) and (5.4) can be found in [7,8], where it is also shown that the stress intensity factor K can be expressed as a function of the

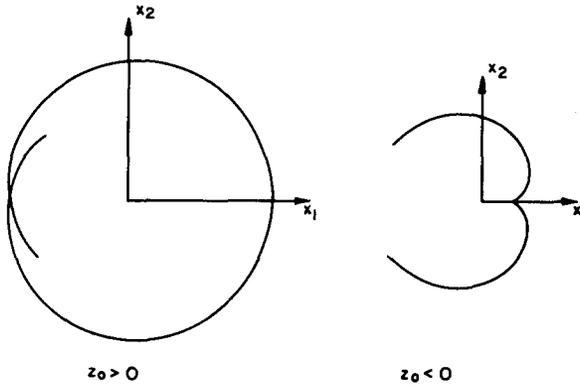


Figure 11. Positive and negative branches of caustic curve for mode-I elastic crack.

maximum transverse diameter, D , of the caustic curve as follows

$$K_{\text{caus}} = \frac{D^{5/2}E}{10.71(z_0 \nu d)}. \quad (5.5)$$

6. Numerical work

As previously discussed, the approximations made in the usual analysis of caustics are invalid in regions of severe optical path gradient. For example very close to the crack tip severe surface deformation gradients exist, resulting in severe optical path gradients of light reflected from the surface. It is often desired to investigate such areas [15]. Thus it is important to know what errors occur in evaluating the caustic using the approximate mapping equations.

For convenience a plane-stress, elastic, Mode-I crack is considered. In the presence of non-linear behavior the singularity in surface displacement at the near tip region is expected to be less severe than in the purely elastic case. Thus this work provides a conservative estimate of the error as the crack tip is approached.

For a given surface deformation ($\nu Kd/E = \text{const.}$ in equation (5.2)) the position of the screen ($x_3 = z_0$) was varied and the corresponding shadow patterns were obtained numerically. A square array of points representing light rays on the specimen surface was mapped point by point through the exact equations (3.4) to the screen. The range covered was from $z_0 \rightarrow +\infty$ to $z_0 \rightarrow -\infty$ corresponding to virtual and real images of the shadow patterns. Special attention was given to the transition region between the virtual and real images. This region corresponds to small initial curves and is of particular importance in the study of the near tip area.

In order to obtain more information on the shape of the caustic the initial curve was computed numerically at each position of the screen (see equation (4.4)). Its points were mapped through equations (3.4) producing the caustic curve. The initial and caustic curves are displayed alongside the discrete full field shadow patterns in Fig. 12.

It is worth observing at this point that for large values of z_0 (either positive or negative) the complete equations of the mapping described in (3.4) reduce to the

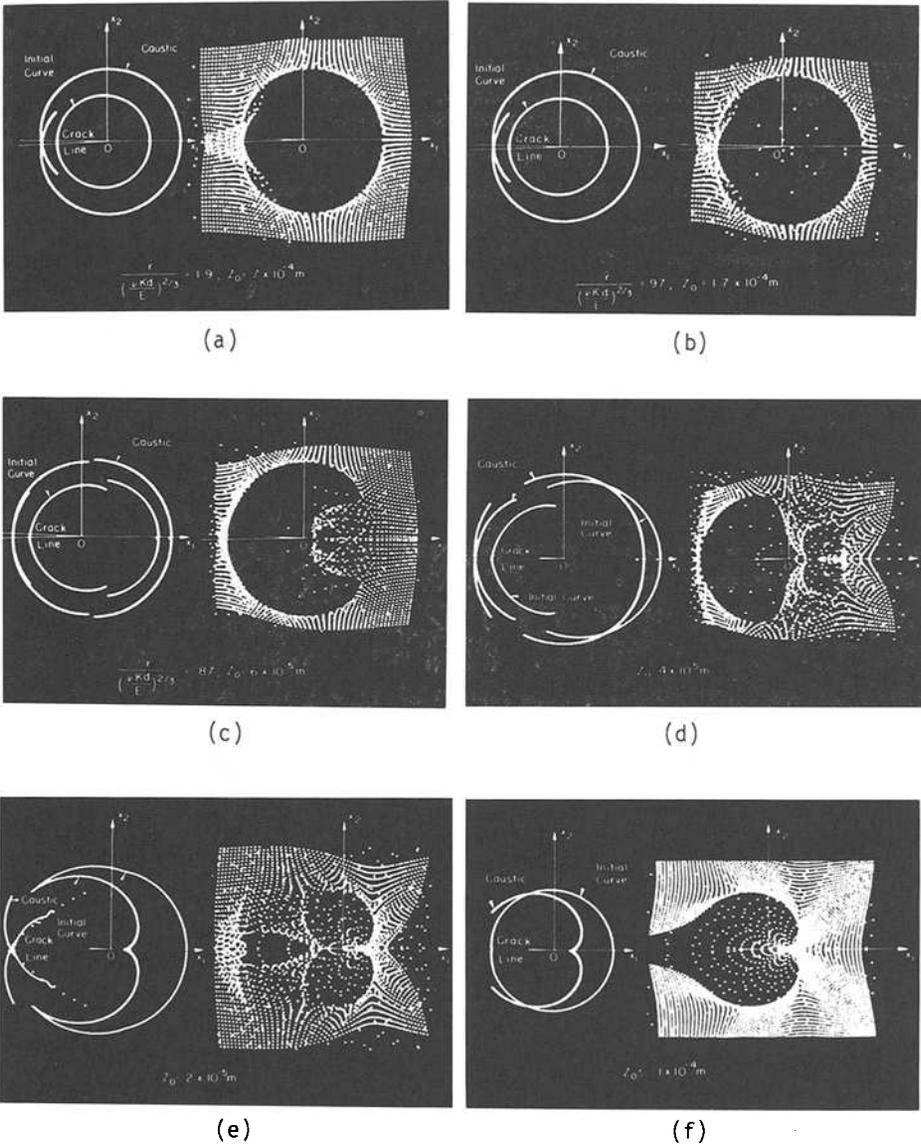


Figure 12. Computer generated shadow spots and caustic curves. Transition from a real to a virtual image is shown.

approximate one considered by Manogg (2.3). This is true since for large enough z_0 , the initial curve can be made large enough so that the caustic curve is formed from regions of less severe displacement gradients than those at the immediate vicinity of crack tip, thus $(\partial f / \partial x_i)^2 \ll 1$ and $z_0 \gg f$. The simplified equations (2.3) predict self similar initial and caustic curves which are circular and epicycloidal respectively.

Comparison of the computed initial and caustic curve shapes to the expected shapes, Figs. 12a (large z_0), 11, demonstrate that for large z_0 (large initial curves) the

approximate equations (2.3) are indeed a good description of the mapping. However our results show that as z_0 becomes smaller (the initial curve approaches the crack tip) the initial curve is no longer circular and the caustic shape deviates from that of an epicycloid. As a direct consequence the customary evaluation of the stress intensity factor through caustic (5.5) becomes increasingly inaccurate.

The transition from virtual to real image is shown in Figs. (12b–d). As z_0 decreases some light begins to map inside the caustic curve producing a blurring of the right edge of the shadow pattern. As z_0 gets even smaller, of the order of f , a discontinuity in the initial and caustic curves occurs as shown in Fig. 12d. For small enough initial curve, there are points on the initial curve where $f(x_1, x_2)$ becomes equal to and then larger than z_0 as ϑ goes from π to zero ($f \sim (1/\sqrt{r}) \cos \vartheta/2$). Where this occurs, the sign of the second term in (3.4) changes and part of the caustic shifts from being a virtual image to a real one. As a result, part of the virtual and part of the real image are present on the same shadow pattern. Comparison of Figs. 12a and 12f illustrates the distinctly different caustic shapes for the virtual and real images. In Fig. 12e part of the real caustic is seen to the right and part of the virtual caustic is seen to the left. In addition to the discontinuity in the caustic curve there is a deterioration of the shape and definition of the shadow pattern. This deterioration limits the values of z_0 that will give meaningful information according to the customary simplified analysis.

By measuring the maximum transverse diameter of the caustic curve, the stress intensity factor K_{caus} that would be calculated in an experiment on the basis of the observed caustic curve was evaluated through the approximate equation (5.5). Its value was then normalized with respect to the value of K input to the numerical analysis and

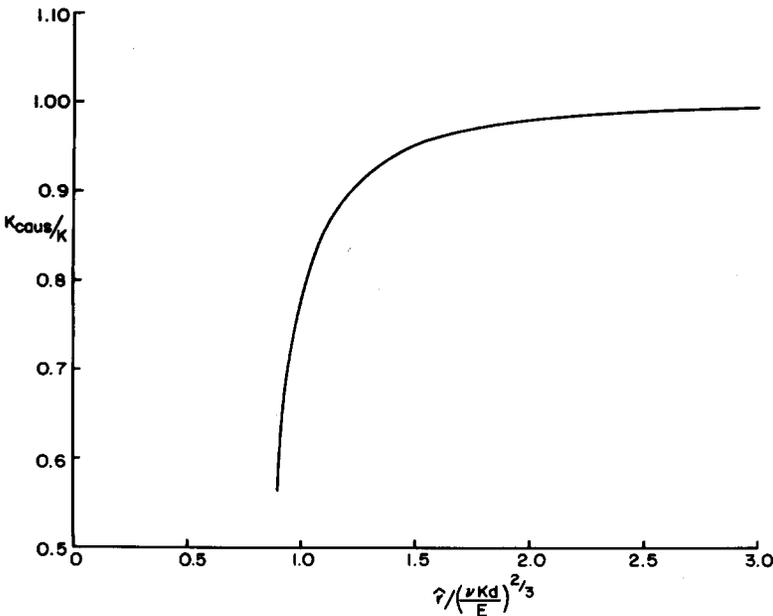


Figure 13. K_{caus}/K vs. $\hat{r}/(\nu K d/E)^{2/3}$, deviation from unity indicates error incurred in calculation of K using simplified mapping.

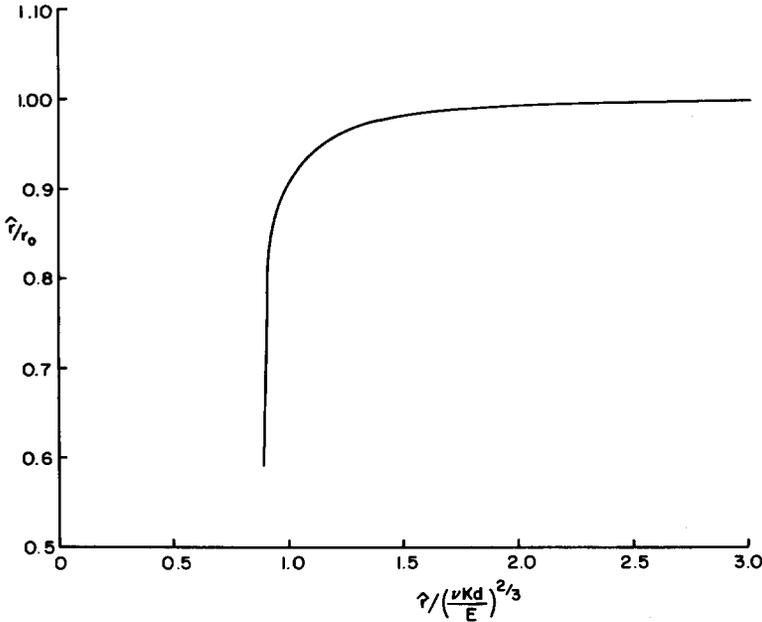


Figure 14. \hat{r}/r_0 vs. $\hat{r}/(\nu Kd/E)^{2/3}$, deviation from unity indicates error in calculation of initial curve using simplified mapping.

plotted as a function of the dimensionless parameter $\hat{r}/(\nu Kd/E)^{2/3}$ in Fig. 13. The initial curve radius, \hat{r} , is defined as the distance from the crack tip to the point on the initial curve that maps to the maximum value of X_2 . ($D = 2X_{2\max}$). The deviation of K_{caus}/K from unity indicates the error incurred due to the approximations used in the caustics analysis. The parameter $\hat{r}/(\nu Kd/E)^{2/3}$ was chosen because $\nu Kd/E$ represents the magnitude of the surface deformation f and is the only relevant length scale in the problem. In Fig. 14, \hat{r}/r_0 is plotted as a function of $\hat{r}/(\nu Kd/E)^{2/3}$. This curve indicates the deviation of the initial curve size from that calculated from the simplified mapping, using equation (5.4). Comparison of Figs. 12b and 13 shows that even before the visible deterioration of the caustic shape, the error in using the simplified analysis for the evaluation of K reaches 15% in K . To give a quantitative indication of the range of initial curve radii for which the errors become significant, the case of a 6 mm thick steel plate (containing a mode-I crack) loaded to a stress intensity factor $K_1 = 100\text{MPa}\sqrt{\text{m}}$ was considered; see Table 1.

Although the calculated values of \hat{r} for which the errors become significant seem to be small, there are cases where it is desired to investigate deformation fields very close to the tip where the above effects could become significant. For such cases it is obvious from the above that the complete mapping equations are necessary.

7. Summary and conclusions

By developing and using the exact mapping equations which describe caustics by reflection, it is demonstrated that some details of the formation of the shadow spot are

Table 1.

\hat{r}_0 (mm)	Error (%)
0.11	10
0.13	5
0.20	2

obscured when the usual small angle reflection approximations are made. This becomes important near regions of severe elevation gradients in the reflective surface. The exact analysis presented here enables us to extend the fracture mechanics applications of the method to regions such as near the crack tip, where such severe conditions prevail. It is demonstrated that in the evaluation of stress intensity factor by shadow spot measurements, based on the approximate analysis, the errors can be as large as 15% before any visible change in the caustic shape occurs. As the crack tip is further approached, the shadow spot shape is shown to change drastically and the transition from a real to a virtual image is demonstrated.

Acknowledgements

The authors would like to acknowledge the advice and encouragement of Professor J.K. Knowles of the Division of Engineering and Applied Science, Caltech. The senior author would also like to express his gratitude to professor L.B. Freund of the Division of Engineering, Brown University, for many useful discussions and suggestions. The research support of the National Science Foundation, contract number MEA-83-7785, is gratefully acknowledged.

References

- [1] J.W. Dally, Developments in Photoelastic Analysis of Dynamic Fracture, Proc. I.U.T.A.M. Symposium on Optical Methods in Mechanics of Solids, Poitiers, France, September 1979.
- [2] A.S. Kobayashi, An Investigation of Propagating Cracks by Dynamic Photoelasticity, *Experimental Mechanics*, Vol. 10 (1970) p. 106.
- [3] A.S. Kobayashi, Current and future experimental methods. In: N. Perrone et al. (eds.), *Fracture Mechanics*. University Press Virginia, Charlottesville (1978) pp. 481–496.
- [4] T. Kobayashi and J.W. Dally, Dynamic photoelastic determination of the a - K relation for 4340 Alloy Steel. In: G.T. Hahn and M.F. Kanninen (eds.), *Crack Arrest Methodology and Applications*, ASTM STP 711. American Society for Testing and Materials (1980) pp. 189–210.
- [5] H.W. Liepmann and A. Roshko, *Elements of Gas Dynamics*, John Wiley and Sons (1957) pp. 157–163.
- [6] H. Schardin, In: Averbach et al. (eds.), *Fracture*. John Wiley and Sons (1959).
- [7] P. Manogg, *Anwendungen der Schattenoptik zur Untersuchung des Zerreißvorgangs von Platten*, Dissertationsschrift an der Universität Freiburg (1964).
- [8] J. Beinert and J.F. Kalthoff, Experimental determination of dynamic stress intensity factors by the method of shadow patterns. In: G. Sih (ed.), *Mechanics of Fracture*, Vol. VII. Sijthoff and Noordhoff (1981).
- [9] P.S. Theocaris, Elastic stress intensity factors evaluated by caustics. In: G. Sih (ed.), *Mechanics of Fracture*, Vol VII. Sijthoff and Noordhoff (1981).
- [10] W. Goldsmith and F. Katsamanis, Fracture of notched polymeric beams due to central impact, *Experimental Mechanics*, Vol. 19 (1979) pp. 235–244.

- [11] W.L. Fourney and H.P. Rossmannith, Crack tip position and speed as determined from Rayleigh wave patterns, *Mechanics Research Communications*, Vol. 7(5) (1980) pp. 277–281.
- [12] K. Ravi-Chander and W.G. Knauss, Dynamic crack-tip stresses under stress wave loading — A comparison of theory and experiment, *International Journal of Fracture*, Vol. 20 (1982) pp. 209–222.
- [13] P.S. Theocaris and E.E. Gdoutos, Surface topography by caustics, *Applied Optics*, Vol. 15, No. 6 (1976) pp. 1629–1638.
- [14] A.J. Rosakis, Experimental determination of fracture initiation and dynamic crack propagation resistance of structural steels by optical method of caustics, *Ph.D. Thesis*, Brown University, June 1982.
- [15] A.J. Rosakis, C.C. Ma and L.B. Freund, Analysis of the optical shadow spot method for a tensile crack in a power-law hardening material, *Journal of Applied Mechanics*, Vol. 50 (1983) pp. 777–782.