Three-Dimensional Effects in Nonlinear Fracture Explored With Interferometry

Thesis by

Richard D. Pfaff

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1991

(Submitted February 25, 1991)

Acknowledgements

I wish to thank:

The faculty for all of their kind assistance, and particularly, Professor Knauss, my research advisor, for his standards, for his dedication, and for giving as much as there was to give; Professor Rosakis for being free with the loan of his equipment, and Professor Ravichandran for remembering best what it is like to be a graduate student and understanding; the staff for having a superb sense of humor and an equanimity which keeps us all nearly sane, and particularly, George Lundgren who never growls when growled at and has found ways to transform many an unworkable set of mechanical drawings into a workable mechanical device; the graduate students, for our shared successes and failures, and especially three: Carl Schultheisz for a long research collaboration where his intelligence and good nature kept us from more than one bad end; Peter Washabaugh for all of our shared endeavors and all of the things that he gave just for friendship and the fun of the research, like spending 23 hours straight in developing and checking photographic negatives so that Carl and I might carry out a particularly difficult and involved experiment in one long continuous 26 hour period rather than one long continuous 38 hour period; Guillermo Pulos for the office that we shared and the friendship that that brought with all of the shared observations and long discussions wherein we tried to make sense of the world in matters large and small; my family for understanding the importance of research without necessarily having to understand the research itself and for steadfastly loving me right through all of my blackest and most despairing periods.

ABSTRACT

The prospects for understanding fracture mechanics in terms of a general material constitutive description are explored. The effort consists of three distinct components.

First, optical interferometry, in its various forms (Twyman–Green, diffraction moire, etc.), can potentially be used under a wide range of conditions to very accurately measure the displacement and strain fields associated with the deformation surrounding a cracktip. To broaden the range of fracutre problems to which interferometry may be applied, certain of the necessary experimental improvements have been developed:

1. High speed camera designs capable of extremely high ($> 10^9$ frames/second) framing rates with large array sizes, ($> 4000 \times 4000$ pixels per frame) so that the application of optical techniques to solid mechanics may be considered without limitation on the rate of deformation.
2. An accurate and adaptable device for dynamic loading of fracture specimens to high load levels utilizing electromagnetic (Lorentz force) loading with ultrahigh ($> 2,000,000$ Amp/cm^2) current flux densities.
3. Implementation of high sensitivity (2 nm), large range (2 nm $\times$ 3,200,000) interferometry achieved with wide field array sizes of $50,000 \times 50,000$ and 8 bit gray scale (error restricted to 1 bit) for surface deformation measurements on fracture specimens.

Second, functional descriptions for certain aspects of the displacement fields associated with fracture specimens are developed. It is found that the fully three-dimensional crack tip field surrounding a through-thickness crack in a plate of elastic-plastic material shows a hierarchical structure of organization and that the primary aspects of the deformation field would seem to have a relatively simple form of expression if the deformation is viewed in a properly normalized form.

Third, a comparison is made between interferometrically measured surface displacements for a notched 3-point-bend speciemn of a ductile heat treatment of 4340 steel and a numerical simulation of the specimen based on a material constitutive description determined from uniaxial tests performed on the same material. The small but finite notch tip radius (0.15 mm) fabricated by a wire-cutting electrical discharge machine allows one to explore the limits of applicability of standard continuum plasticity theories without involving a process zone model for the very near tip region extent in a cracked specimen geometry.

Table of Contents

Preface

An understanding of fracture, especially when nonlinear material behavior is involved, requires both the capability for reliable stress analyses as well as for experimental observations which are commensurate in accuracy. This statement is true with respect to both quasistatic and dynamic fracture conditions.

In earlier studies of dynamic fracture at GALCIT, questions had arisen with respect to the more detailed mechanism(s) of fracture under dynamic conditions that involved high rate processes in a small region surrounding the crack tip. The interest centered on the development of small microfractures which lead through coalescence to the growth of the macroscopic crack. A better understanding of this basic phenomenon and its role under the static or dynamic conditions requires that one be able to observe relevant phenomena with sufficient temporal and spatial resolution on the one hand, but that one be also able to provide the requisite (possibly nonlinear) stress analysis in order to make associations of the observed phenomena with continuum mechanical stress or deformation fields at the tip of propagating cracks; it is preferable that one possess this capability for dynamically moving cracks.

This thesis provides some building blocks to the larger framework of accomplishing that goal. One of these topics addresses improved designs of high speed optical recording (high speed camera design) with high spatial resolution; another is concerned with the capability of the mechanics community to model the stress and displacement fields at the tip of moving cracks with sufficient accuracy so that improved dynamic recording methods and the ensuing results can be buttressed by commensurate analyses.

The latter objective requires the development of experimental methods that allow high spatial resolution so that deformations at the tip of cracks can be determined with a high degree of accuracy. Optical interferometric techniques are candidates for this purpose. In order to assess whether present day (finite element) analysis methods and material characterizations are able to model crack tip defor-

mations reliably it is necessary to compare experimentally determined deformations with those computed for the same test geometry.

This comparison is one of the building blocks for the overall intent of this thesis, except that the experimental/analytical work is restricted here to deal with the out-of-plane deformations in the crack tip region under static conditions. The additional complication of dynamic motion does not affect the optical part of this comparison because of the high speed of light relative to mechanical deformation rates. On the other hand, the treatment of the dynamical crack propagation problem with nonlinear (elasto-visco-plastic) material description is considerably more open to question than the static problem, in significant part because of the lack of experience with the appropriate constitutive behavior: Consequently, the comparison is limited to the static problem. Additional motivation for this study is delineated in Chapter 2 of this thesis.

The improvement of high speed optical recording facilities possesses two phases: One consists of the improvements in the operation of the existing rotating mirror camera, specifically with respect to increasing the angular aperture to reduce earlier diffraction limitations and increasing the versatility of operation to allow for easier set-up conditions when a wide variety of test conditions need to be accommodated. This particular aspect of equipment design is not delineated here.

Although funds were not adequate to complete the design of a novel high speed camera with significantly improved recording ability, both with respect to space and time, plans are outlined for the configuration of such a camera of the raster type. This paper study is developed in Chapter 3 of this thesis.

Finally it should be remarked that a part of the dynamic fracture studies alluded to above depend on one's ability to control the dynamic loading applied to a test geometry as well as supplying sufficiently high load levels to induce fracture in the materials under study. With this need in mind an existing electro- mechanical loading device was redesigned and constructed to allow fracture of Aluminium specimens, where formerly only polymer specimens could be accommodated. This

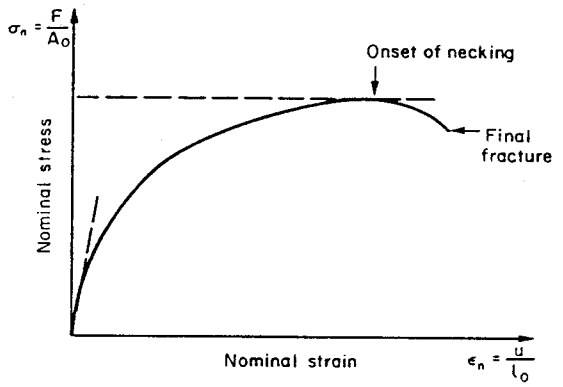
development is not recorded in this thesis, but is mentioned here to round out the presentation of the overall accomplishment of providing the experimental and analytical building blocks to study static and dynamic fracture.

Chapter 1

Introduction

The irreversible energy absorption represented by a particular test specimen's "fracture toughness" arises from the volume-specific energy absorption integrated over the volume of disturbed material that extends away from the crack surfaces. Tabulations of fracture toughness extracted from data on standardized specimen configurations have been a useful engineering guide during the time period when it was too difficult to perform calculations based on actual, detailed constitutive behavior of the material. However, in many endeavors (an example being the design of fracture resistant composites) it is the full constitutive description of the component materials which is of importance and not the measured bulk fracture toughness of the individual material components as taken from homogeneous specimens. Ideally, one would expect to eventually reduce the present concept of fracture toughness characterization to merely a particular and calculable large scale manifestation of a constitutive model applicable down through the small scale dimensions appropriate to the crack tip region.

To view the problem in an illustrative albeit simplified light, let us consider a schematic plot of the data from a uniaxial tension test as it is typically displayed in introductory engineering material science texts. The example here is taken from Ashby & Jones [1].



One cannot very well develop a complete constitutive description of a material

from a curve which ends in midair at this point labeled "final fracture." Conversely, the energy absorbed in fracture in a standard fracture toughness specimen does not consist only of energy absorbed by material undergoing strains beyond that at which final fracture would be said to be occurring in the uniaxial tension specimen.

In effect, one should like to extend the full, three dimensional constitutive description of a material, such that a plot in true stress vs. true strain would pass smoothly to zero at some sufficiently large strain. For those crystalline materials which (because of dislocation immobility) do not exchange next nearest atomic neighbors, such a constitutive model generally already exists in the form of a known interatomic potential. For most other materials, no such model exists. One could reasonably expect that the development of such a description and an understanding of its microstructural origins would be of considerable utility in designing fracture resistant composites.

The "engineering fracture toughness" of bulk material has been a very poor guide in this regard, since fracture toughness is properly thought of as the property of a "body"* of material rather than the property of a material: In the continuum description of materials, one idealizes or homogenizes materials such that any infinitesimal cube is considered to have a set of associated properties (density, melting temperature, heat capacity, tensor coefficients of a modulus of elasticity, etc.). Traditional fracture toughness is not a "material" attribute ascribable to an infinitesimal element in this same way. However, the constitutive description of many

---

* Consider, for example, the usual necessity of defining a plane strain fracture toughness vs. a plane stress fracture toughness [2]; fracture under small scale yielding conditions vs. the fracture of a wire; a stress intensity factor related to crack propagation velocity (for dynamically propagating cracks) even for materials which exhibit no equivalent rate dependence in their constitutive description because the influence of inertial terms, alone, is to alter the stress and deformation fields surrounding a dynamically propagating crack from those for a quasistatically propagating crack.

materials might be fruitfully extended such that it would encompass all of the behavior inherent in the fracture of macroscopic bodies made up of such material.

Before this understanding can be formulated, a number of difficulties need to be addressed.

1.1 Basic Issues in Need of Resolution

1. The deformation field surrounding a crack front is, generally, fully three-dimensional. This fact limits the simplifications which can be introduced into the mathematics of the problem through a reduction of dimension (limitation on describing a 3-D problem in terms of a 2-D problem, for example).
2. The more interesting (useful) materials are, more or less by definition,the tougher materials. Consequently, a large strain formulation of fracture problems is often necessary, large deformations being one of the common mechanisms of increased toughness. Large strains pose both mathematical and experimental difficulties. An additional concern from an experimental point of view, though, is that these same tougher materials will be capable of sustaining relatively higher stress levels. This limits the way in which the testing apparatus may be constructed.
3. Fracture involves material behavior ($d\sigma/d\epsilon < 0$) which would be unstable in a homogeneous state of stress, but which does not lead to instability in nonuniform fields as in a fracture specimen, where the material in the unstable range of its response is surrounded by material in the stable portion of its response curve.
4. Associated with this material instability and the fully three dimensional nature of the fields common to fracture, then, is an inability to establish the material constitutive model in a simplified geometry and separate from a fracture test, this model then being generalized and applied to crack problems. If small scale stress measurements and deformation measurements were easily made,

then establishing the complete constitutive description for a material would be an entirely experimental endeavor. However, it has generally proved difficult to directly measure stresses on a small, local size scale (in those cases where the body is not in a homogeneous state of stress) even near the external surface of a body. Measuring deformation in the deep interior of a body is also generally difficult. Measurements of surface deformation are most readily made, but by themselves are insufficient to establish a constitutive relationship. These observations would seem to suggest the use of theoretical calculations made in conjunction with what is expected to be an incomplete set of full field experimental measurements as part of the modeling process.

5. The fracture process involves material description down to angstrom size scales where most of the classic continuum descriptions of materials become invalid. At the same time, the fracture process is embedded in a body which is usually quite large in relation to the fracture process zone. The approximately 10 orders of magnitude in size scale over which the material model must apply makes its formulation considerably more difficult. From an experimental aspect, it makes matters more tractable if the size range for experimental observations can be reduced. It is not possible to follow the movement of each atom in a fracture specimen whose external dimensions are of the order of $10cm$. While in some cases it may be necessary to follow the motions of individual atoms near the crack tip, continuum descriptions are generally quite sufficient for the larger size scales. In this sense, one's measurement techniques may average over a larger and larger volume of atoms as one becomes progressively farther and farther removed from the highest strain regions in the fracture process zone. This general idea is reflected more or less universally in finite element formulations of fracture problems: the character of the mesh is of diminishing element size as the distance from the crack tip diminishes. It is less clear how one devises a completely general experimental technique which has this same character.

6. Most nonnumerical descriptions of the strain fields in fracture specimens con-

sist, even for linearly elastic and monotonically hardening plastic (HRR) materials, of asymptotic analyses which are mathematically asymptotically correct in a limit which is precisely where the model description is completely invalid.

7. Most materials in use are also inhomogeneous (polycrystalline, etc.) at one or more size scales of description. The specific details of this inhomogeneity (interface strengths, for example) can be of dominant importance in the fracture behavior of the conglomerate substance.

8. There is a consequent necessity of developing experimental techniques which are free of *a priori* assumptions in their interpretation and which are suitable under the wide range of material strain mentioned. Methods like photoelasticity and transmission optical interferometry are inherently unsuitable in that their application is substance dependent. One desires techniques which are independent of material choice.

9. The nature of a crackfront is to be hidden (in most materials and in most senses) in that, by definition, it is interior to the original surface of the body.

10. Experimental techniques using only measurements of surface displacements and tractions to infer the internal strain field of a fracture specimen incur the limitations inherent in the formulation of a Cauchy or inverse problem [3]. An intuitive sense of this is perhaps more easily grasped by viewing such techniques in light of the familiar St. Venant principle—that in linear elastic media certain self-equilibrating distributions of tractions have an associated stress field which diminishes exponentially with distance. One may see then that severe, but localized stress fields might exist within a material (from void creation, for example) while the associated effect some distance away, at the exterior of the body, would be quite limited or nonobservable. This difficulty may be partially overcome by increasing the sensitivity of these surface measurement techniques. However, it is only completely overcome in this way under fairly limited circumstances.

11. While it may usually be convenient to hope for materials which exhibit no time

dependency in their constitutive description, many important materials show pronounced rate dependence in response to part of the load spectrum to which they may be subjected. Perversely, material toughness often diminishes with an increase in loading rate. Perversely, because the shorter the overall event, the higher the required experimental data acquisition rate for the capture of the phenomenon.

12. Another unfortunate behavior of most solids is their susceptibility to "subcritical" crack growth such as results under fatigue, stress-corrosion, etc. Since this feature is so important in determining the useful lifetime of most structures, a general theory of fracture would be quite incomplete if it did not encompass these behaviors. Consequently, one should prefer experimental measurement techniques which allow not only for a high data acquisition rate, but for a high data storage capacity as well. In this way, one might hope to effectively observe high frequency fatigue fracture. Alternately, some particularly effective way of observing the fracture process which would reduce the raw data storage and raw data storage requirements would serve as well.

It is not the intention to make here a completely inclusive list of all the technical difficulties associated with the reintegration of fracture mechanics into the larger continuum mechanical description of engineering solids. The items mentioned are taken here as one particular starting point for the subsequent developments and will be explored to greater and lesser degree in what follows.

A way of summarizing the foregoing discussion of technical difficulties may be found through consideration of the concept of Lagrangian description of the deformation in structural solids. Throughout, it has been implicitly assumed that a formulation which follows the displacement of individual material points will continue to be an effective and unifying view. The technical difficulties then revolve around the various requirements for extending this Lagrangian description to the

material behavior involving larger strains at smaller size scales exhibited in fracture mechanics.

One might then phrase this process in the still broader context of the primary steps usually taken in scientific reductionism:

1. Experimental Observation
2. Model Development
3. Model Verification
4. Development of Mathematical Tools—for the ready application of the model to the full range of its applicability.

In this context, it may be said that the description of the material at different distances from the crack in a fracture specimen is presently at different stages of development in the process of scientific reduction of the exhibited behavior.

1.2 Test specimen as extension of the testing machine and relation to Cauchy problem

The usual distinction between the testing apparatus and the test specimen is that the testing apparatus is considered to be well enough understood in its behavior that it can be utilized to perform a prescribable action upon the test specimen. The reaction of the specimen is then observed to categorize its behavior in light of the prescribed action. By extension, a standard fracture test specimen might be treated in such a way that part of this specimen (the material at sufficient distance from the crack that its exhibited constitutive behavior is already modeled and computationally tractable) could really be treated as if it were part of the testing apparatus, while the as yet not fully understood material response nearest the crack might be viewed as the test specimen proper. Admittedly, this leads to a "testing apparatus" the actions of which may be only weakly prescribable and then only in a way which is strongly coupled to the response of the "test specimen." Consider constructing an imaginary surface surrounding the "test specimen" region of the

fracture specimen. Suppose one could measure the tractions across this boundary (or the displacements associated with this boundary) as the load is progressively increased on the fracture specimen. Then, given the known constitutive description for the "testing apparatus" region of the fracture specimen and the known far-field boundary conditions, a completely specified boundary value problem is available for the "testing apparatus," and the unmeasured displacements (or tractions) associated with this imaginary surface might be calculated, so that the complete set of conditions (tractions and displacements) applied to the "test specimen" would be known. Unfortunately, it is just such displacements (and tractions) associated with points internal to the body which are so difficult to measure directly, in general. Suppose, on the other hand, that one merely assumes some traction boundary condition associated with this imaginary internal surface and performs the same calculation on the "testing apparatus." The "testing apparatus" will have a displacement response throughout (not just at the imaginary surface) which is different from what it would have been had one used the actual traction boundary condition. An ability to detect such differences at the exposed exterior of the "testing apparatus" is the basis for the inverse problem of trying to establish the conditions at the imaginary internal surface indirectly from the external response of the "testing apparatus."

Fortunately, there are techniques of optical interferometry (Twyman-Green surface interferometry and diffraction moire interferometry) which have a theoretical capacity to measure surface displacements on the exterior of an object with high sensitivity (better than $(\frac{\lambda}{4}\ |_{\lambda=500nm})\frac{1}{64} \sim 2nm$) while simultaneously retaining a large measurement range (changes of surface slope of ± 15 deg; of in plane strains $\sim 60\%$) over a substantial field of view (array sizes of $100,000 \times 100,000$ pixels) with a minimum pixel size (or observational averaging area) $\sim \frac{1}{2}\mu m^2$ for visible wavelength light.

However, before proceeding with the inverse problem, one ought to assure oneself first that the ability to calculate the deformation of the "testing apparatus"

will match what would be measured in a simplified case where one actually does know the boundary conditions at such an imaginary internal surface surrounding the "test specimen." Such a simplified case is obtainable in virtually any material through the expedient of a blunt notch machined (or otherwise fabricated) into the virgin test material of the fracture specimen. In this case, the pseudo-"test specimen" becomes the "nothingness" inside the notch. The notch tip radius should probably be something like the expected radial extent of the "test specimen" zone in an actual fracture specimen.

Further, it is important not to forget the limitations inherent in the inverse or Cauchy problem. For cases where the "test specimen" is of extreme aspect ratio (that is, where the length of the crack front extended through the body is large in comparison with the radial extent of the "test specimen"), surface measurements at the exterior of the body (the traction-free lateral faces of a plate, for example) will be insensitive to the specific details of the conditions at the imaginary surface deep in the interior of the body. To state this limitation in the form of an example, exterior surface measurements would be expected to be a poor means of studying plane strain fracture, or what sometimes approaches plane strain fracture at those points on the crack front most removed from the exterior of the body (from the plate faces). Exterior surface displacement measurements are really only expected to clearly elucidate the behavior of material surrounding crack fronts in the region nearest to where the crack intersects the free surface (exterior) of the body. The nature of fracture near a free surface has its own distinctive characteristics, which are quite different from what occurs well away from the free surface. One may recall fracture of thick metal plates, which have a cleavage type fracture in the interior and shear lip formation associated with the free surface [2].

Beyond surface interferometry using visible and ultraviolet frequency light there are a number of techniques for establishing the surface deformation of a fracture specimen at very small size scales (variations in electron, scanning tunnelling, atomic force and low energy x-ray microscopy). While there is a considerable amount of

development work still to be done on these techniques for small scale deformation measurement, there is no fundamental breakthrough required. It seems that one may expect to measure accurately the characteristics of deformation near a fracture as the fracture intersects a free surface of the fracture specimen.

Penetrating x-ray techniques can be expected to improve the resolution for observations of gross void formation in the deep interior of materials down well into the submicron range. Again development work is required.

It is not yet clear how deformation and stress measurements in the deep interior of a body may be readily made. In the interim, there is a considerable amount which can be learned by adapting highly accurate techniques for measuring surface deformations to the widest possible set of circumstances. With these measured surface displacements (and hence strains) and utilizing extensive computations it is possible to infer something of the character of the constitutive description for material within the "test specimen" even without being able to directly measure stresses on a small scale.

It would also seem possible that the constitutive behavior exhibited at crack fronts under conditions of plane strain (or in general for states with nonzero stress parallel to the crack front, i.e., away from the free surface of a planar fracture specimen) might be observable as incipient void nucleation at the surface of an ordinary planar specimen (rather than on a cracked or notched fracture specimen) where the surface is under general biaxial loading. To avoid the fracture instability which occurs in specimens under homogeneous states of stress, it would be necessary that the condition of biaxial surface stress be created by some global loading conditions, such as bending in orthogonal directions of a plate, which would lead to stable, natural fracture initiation. The strain and stress gradients associated with such a bending specimen would also match more closely than a uniform stress the gradient which would be expected near a crack tip. One might think of this loading geometry as the case of a notched fracture specimen under pure bending where the depth of the notch has been taken to the limit of zero.

1.3 Accurate Displacement Field Characterization

In the broader attempt to adapt highly accurate techniques for measuring surface deformations to a wide range of problems, certain of the necessary experimental and analytical improvements have been developed and are described:

1. *High speed camera designs capable of extremely high ($> 10^9$ frames/second) framing rates with large array sizes, ($>$ 4000×4000 pixels per frame) so that the application of optical techniques to solid mechanics may be considered without limitation on the rate of deformation.* This topic is covered in Chapter 3 of the thesis.
2. *An accurate and adaptable device for dynamic loading of fracture specimens to high load levels utilizing electromagnetic (Lorentz force) loading with ultrahigh ($> 2{,}000{,}000$ amp/cm^2) current flux densities.* It is not necessary to have room-temperature superconductive materials available in order to implement certain designs which require conductors which can sustain ultrahigh current flux densities. Ordinary copper can sustain $2 \times 10^6 amp/cm^2$; it is merely the case that copper cannot sustain such a current flux density continuously - it melts or softens from Joule heating after a time, t, determinable from the relation

$$t \approx [(\rho C)/(j^2(p\,|_{T=20^0C}))](\frac{1}{y_T})ln\left[\frac{1+y_T(T_f-293.16K)}{1+y_T(T_i-293.16K)}\right]$$

where

$\rho \equiv$ density

$C \equiv$ heat capacity

$j \equiv$ current flux density

$p\,|_{T=20^0C} \equiv$ resistivity at 20^0C

$y_T \equiv$ temperature resistivity coefficient

$T_i \equiv$ the initial temperature of the conductor

$T_f \equiv$ the acceptable final temperature of the conductor.

The annealing temperature for most metals falls between $\frac{1}{3}$ and $\frac{1}{2}$ of the melting temperature. If one evaluates the foregoing relationship for copper and chooses the

initial temperature as room temperature (20^0C); the final temperature as $\frac{1}{2}(\frac{1}{3}+\frac{1}{2})$ of the melting temperature; the current flux density as $2 \times 10^6 amp/cm^2$, then

$$(t\,|_{Cu;T_i=20^0C;T_f=300^0C;j=2\times10^{10}amp/m^2}) \approx 95\mu sec$$

Such a time duration of approximately $100\mu sec$ is appropriate for a dynamic loading device. The details of implementing a device for dynamic loading utilizing Lorentz force loading are set out in SM GALCIT Report 1982-2 [4] and SM GALCIT Report 1991-00 [5].

3. *Functional descriptions for certain aspects of the displacement fields associated with fracture specimens.*

A) A useful perspective on the development of fracture mechanics may be gained by comparison with the development of the theory of elasticity. Much of classical elasticity is built around a solid with the simplest possible constitutive description (linearly elastic, isotropic and homogeneous). Having taken such a constitutive description, various analyticians then set forth the mathematics necessary for finding solutions with the widest possible set of boundary conditions, and all of the special cases (or subsets of boundary conditions) were explored in great detail. Alternatively, fracture mechanics, when faced with what might be considered the widest possible range of constitutive behavior, very soon settled on what amounts to the simplest possible boundary condition—the stress intensity factor field as a far-field condition. Having taken such a boundary condition, fracture mechanicians have then sought to characterize fracture behavior for the widest possible range of constitutive model, and various special cases (or subsets of constitutive behavior) have been explored.

While it is convenient to characterize the boundary condition on the elastic far-field of a fracture specimen in terms of a constant(s) of proportionality for the embedded singularity which is associated with the crack tip, such a characterization *is* certainly only an approximation. The failings of this approximation have gradually become more apparent and attempts have been made to formulate alternative

criteria for characterizing the boundary conditions applied to a fracture specimen as these boundary conditions influence the fracture behavior. Thus were born a number of specialized integrals, the most notable of which has been the J integral [6].

Alternatively, one can of course return to the infinite, separable series (in r and θ) for the plane problem and the antiplane shear problem (of which the three K fields (K_I, K_{II} , and K_{III}) are merely the least singular of all of the singular terms of such a series) and characterize the boundary conditions applied to fracture specimens using the infinity of coefficients of the infinite number of terms (singular and nonsingular) in this series. Unfortunately, this infinite, "K-field" series is suitable for describing, in generality, the boundary conditions on a split annulus with the split or crack being traction free; and the coefficients for the nonsingular terms of the series may not be computed solely from the boundary condition on the larger diameter circle of the annulus any more than the coefficients of the infinity of singular terms in the series may be computed solely on the basis of the boundary conditions on the smaller diameter circle of the annulus. When the ratio of the two characterizing diameters of the split annulus (within which the material remains linearly elastic in its response) becomes extreme, such as is associated with small scale yielding, it becomes a reasonable approximation to calculate the coefficients of the nonsingular terms solely on the basis of the boundary conditions at the far-field (large diameter contour) of the annulus (by ignoring the existence of all singular terms save the least singular term or K term). The boundary conditions on the fracture specimen are then viewed in light of the stress intensity factor (coefficient for the K term) and the coefficients of higher order (nonsingular) terms in the series.

A third approach is considered in Appendix A of the thesis. The basic ideas underlying this third approach are simple and intuitive and are found as part of the body of fracture mechanics literature. Briefly, for the plane problem, it is reasonable to replace the "K-field" series with a different type of series which is no longer separable but which exhibits much more rapid convergence in expressing the

elastic stress field surrounding a crack or notch tip. The first three terms in the series are solutions for an elliptical notch in pure, far-field bending, dilatation and distortion. Any of these three "non-selfequilibrating" stress distributions has an associated stress intensity factor. However, (with the same stress intensity factor) the character of the yield zone surrounding the crack tip can be significantly different depending on the preponderance of bending, dilatation, or distortion generating that stress intensity factor. An example of this effect is explored by Shih [7] in finite-element calculations for an elastic-plastic material. In Appendix A this basic idea is made formal and mathematical.

B) While there are analytic solutions for a circular hole in a finite thickness (linearly elastic, homogeneous, isotropic) plate [8] and for a circular hole in a infinite halfspace[9], there is no comparable solution for a through-thickness crack in a finite thickness plate. As a result, various researchers have generated finite element solutions for this "3-D K-field" problem. It would be advantageous if these approximate numerical solutions could be used to establish an approximate mathematical (symbolic) solution if, in fact, a reasonably simple "curve fit" could be found. Such a mathematical formula is a useful tool in interpreting and organizing experimental surface displacement measurements made on actual fracture specimens. For this reason, the out-of-plane, free-surface displacement data from the FEM investigation of the "3-D K field" by Nakamura and Parks [10] has been fitted to an approximate mathematical formula. It is possible to fit the available data with a relatively simple formula. When more accurate and extensive numerical data becomes available it will be possible to ascertain whether a simple formula can be retained, with minor adjustment, for higher accuracy approximations.

The formula and a comparison with the numerical data from [10] are to be found in Appendix B.

C) Investigations by Benthem [11] have explored the "State of Stress at the Vertex of a Quarter-Infinite Crack in a Half-Space" and, in finite element investigations [10], it has been shown that such a solution dominates even for finite thickness

plates in two spherical regions surrounding the two points of intersection of the crack front with the traction free planar surfaces of the plate. The size of the spherical region of dominance is a small fraction of the thickness of the plate. While these investigations are for a linearly elastic material, a similar character would be expected when, at higher load levels, a fracture specimen would begin to exhibit non-elastic deformation. This in-elastic deformation is generally exhibited first at this point or corner [12] of the "quarter-infinite" crack, given that the distortional strains will be highest at this point where the traction free surface limits the triaxiality which may develop as compared to points along the crack front nearer to the mid-plane of the plate. One observes experimentally that the region of the plastic zone associated with shear lip formation in metal specimens follows this pattern of developing from the corner of the crack outward into the body of the specimen in a process which is initially insensitive to the thickness of the plate as long as the shear lip zone is a fraction of the plate thickness. If such a deformation field were to expand, with increasing load levels, in something of a self-similar fashion this "scaleability" would be an appealing mathematical nicety if, in fact, it were applicable in characterizing a significant class of experimental observations of deformation fields in fracture specimens.

With this idea in mind, the data for out-of-plane surface displacements generated from an FEM analysis of a notched (with small tip radius) fracture specimen of an elastic-plastic material [13] were manipulated to explore the degree to which this surface displacement field would exhibit some form of scaleability. It seems to be the case that the out-of-plane surface displacements associated with elastic-plastic deformation do scale with load level when these displacements are appropriately normalized. A simple but relatively effective normalization relation for the out-of-plane surface displacements is described in the following way:

$$(w_s(x,y) - w_{se}(x,y))/w_{se}(x,y)$$

Where

$w_s(x,y) \equiv$ the out-of-plane surface displacements associated with the elastic-plastic material description.

$w_{se}(x,y) \equiv$ the out-of-plane surface displacement which would have been exhibited if the fracture specimen had remained linearly elastic at higher load levels.

This particular approach to characterizing the surface displacements is illustrated in Appendix C.

4. *Implementation of high sensitivity ($2nm$), large range ($2nm \times 3,200,000$) interferometry achieved with wide field array sizes of* $50,000 \times 50,000$ *and 8 bit gray scale (error restricted to 1 bit) for surface deformation measurements on fracture specimens.*

Interferometry is widely used in the optical industry to establish deviations from figure for lens elements. In this application, the deviation of the lens surface from "perfection" is generally only a fraction of the wavelength of visible light for high quality optical components. As manufacturers have improved their manufacturing techniques, the surface error of optical components has become progressively smaller, and the sensitivity of testing interferometers has kept pace, given that the interferometer has optical elements as well, so that interferometers can now measure displacement differences on the order of $\frac{\lambda}{4}\frac{1}{64}$ or $\sim 2nm$. Unfortunately for the study of solid mechanics, the optical industry has little interest in establishing an ability to maintain this high sensitivity while measuring severely "imperfect" surfaces.

In order to apply high sensitivity interferometry to the comparably very severe distortions encountered in the study of deforming solids, it is necessary that measurement range not be sacrificed for measurement sensitivity. This necessity is doubly true when one is particularly interested in large strain deformations, as is to be expected in the study of fracture mechanics. The most essential element of an interferometric apparatus suitable to fracture mechanics studies is what might best be referred to as a very long working distance, large angular aperture, very large field of view microscope objective lens of near diffraction limited performance. Historically, microscope objectives attained diffraction limited performance by mak-

ing their geometric aberrations smaller through the expedient of making the lenses smaller (since geometric aberrations scale with the size of the lens) [14]. Obviously, this limited the working distance for the lens as well as the field of view. As the technology of optics has improved, it has become possible to design lens systems which have reduced geometric aberrations so that they may be made physically larger while still retaining near diffraction limited performance. However, the state of optical technology is not yet such that microscope objectives can be made, as it were, infinitely large so that they maintain the same small, minimum resolvable spot size while being able to image an infinite field of view. In the interim, it is possible to attain performance significantly better than might be expected from the available microscope objectives by optimizing a lens design for the specific purpose of wide field interferometric use, which is a different criterion than is ordinarily applied in microscope optimization. As an example, most lens designs are optimized for white light use, but interferometry with monochromatic laser light does not require the consideration of chromatic aberrations, a consideration which has a pronounced and restrictive influence on designs which must image white light.

Utilizing a multielement lens design more suited to wide field interferometric microscopy, it has been possible to obtain interferograms in regular usage which attain an overall pixel array size of approximately 50,000 x 50,000. It is to be expected that with more advanced designs, one may obtain array sizes larger than 100,000 x 100,000.

1.4 Example Experiment and Analysis

A quasistatic fracture test was carried out on a 3-point-bend fracture specimen of medium-ductile 4340 steel plate of 10 mm thickness with the intent of measuring simultaneously all three components of the displacement field over the traction free faces of the plate using Twyman-Green interferometry for the out-of-plane displacements and moire interferometry for the in-plane displacements. Given the

considerable ductility of the steel and the thickness of the specimen, the eventual plastic zone size was observed to be larger in extent (mean diameter) than 10 mm. As well, the associated surface slopes (norm of the in-plane gradient of the out-of-plane displacement) of what were originally the planar surfaces of the plate attained a maximum local value near the crack tip greater than 15 deg. Similarly, the in-plane true strains would be expected to attain maxima in the vicinity of the crack tip commensurate with the breaking strains ($\sim$ 70% true strain) observed in uniaxial test specimens of the same steel. It is under such conditions of observing large deformations over a large field of view that optical imaging systems capable of resolving a large field of view while retaining a large angular aperture are necessary. For a diffraction limited lens, this product of the field of view times angular aperture is directly proportional to the resolvable array size.

Example portions of the out-of-plane surface displacement interferograms from this 3-point-bend fracture test are to be found in Chapter 2 of the thesis. These measured displacements are compared with an FEM analysis [13] of this specimen geometry and material (the material attributes were based on uniaxial tests of the material). It should be pointed out that the average photocopy machine can only represent a pixel array size of approximately 2000 x 2000 without gray scale over a 15 cm by 15 cm page area. It is for this reason that only portions of interferograms which amount to pixel array sizes of 50,000 x 50,000 with 8 bit gray scale are presented here. This limited pictorial representation capability of the printed page supplies as well the motivation for finding an effective means of compressing the data found in such interferograms thru representation with appropriate mathematical functions. Appendix A, B, and C detail one possible group of such mathematical functions, hopefully appropriate in having captured some of the essential characteristics of the deformation fields to be expected in a class of fracture specimens. The ultimate goal, of course, is to find the most appropriate mathematical functions by virtue of their being elegantly compact closed form analytical solutions to these fracture problems. However, the solutions to fully three dimensional, nonlinear frac-

ture problems are by no means assured of being representable with a finite number of elementary functions. In the interim, high accuracy representations of experimentally measured deformation fields in fracture specimens will serve as a useful tool in deducing the detail with which a materials constitutive behavior must be described in order for numerical simulations (whether by finite element technique or by something more advanced) to adequately model the deformation fields of these fracture specimens.

Chapter 2

Surface Interferometry Compared to a Three–Dimensional Finite Element Method Simulation of the Deformation Field in a Notched, Three-Point-Bend Specimen of Ductile 4340 Steel

The rapid advances in the computer industry have made the Finite Element Method (FEM) a fairly powerful technique for establishing an approximate numerical solution for problems involving nonlinear material constitutive descriptions. While the FEM technique is inelegant, a large computational capacity can compensate for a lack of elegance. In this light, it has now become reasonable to explore the degree to which the deformation fields in fracture specimens are predictable on the basis of a general constitutive description for the material.

It was the initial plan to perform this study with an aluminum alloy because of its relevance to aircraft construction, particularly as the project was funded by AFOSR. However, it turned out that developments originating with computational efforts by Professor Ravichandran, and by Professors Rosakis and Narasimhan, provided the rudiments of a computer model for a steel specimen. It was therefore decided to use 4340 steel as a model material instead of aluminum. As a material choice, 4340 steel has a number of advantages: 4340 steel is one of the most studied of engineering structural materials and as such there is a considerable body of work which elucidates its constitutive behavior; 4340 steel exhibits very little creep for a fairly wide range of testing conditions and this lack of time dependence in deformation response allows for more latitude in how experimental observations may be carried out; 4340 steel is widely used, and a broad and readily accessible technological base exists by which the material may be controlled in specifying its microstructure and large scale geometric form through heat–treating, grinding, polishing, etc.

Clearly, it would be preferable if one could easily make measurements of the strain and stress fields throughout the interior of a deformed solid of any type. Such ease of measurement is not presently available. However, the surface displacements

of a wide range of deformed materials may be measured using diffraction moire interferometry and Twyman-Green interferometry.

A collaborative experiment involving two researchers has been carried out where moire interferometry and Twyman-Green interferometry have been performed simultaneously on the planar surfaces of a notched, three-point-bend specimen of 4340 steel. An FEM analysis of this specimen configuration was carried out using the classical elasto–plasticity description determined from uniaxial tension tests. A complete description of the FEM analysis and of the various tests used to establish a material constitutive description may be found in [13]. A brief summary may be found in Appendix C of this thesis. Additionally, the results of the moire interferometric measurements of the in-plane surface displacements are found in [13]. Representative results of the out-of-plane interferometric measurements are reported here along with a comparison to the out-of-plane displacements predicted by the numerical simulation.

2.1 Comparison of Interferograms with Numerical Predictions

The comparison between out-of-plane surface deformation interferograms and the associated numerical predictions may be illustrated by representing the numerical data as a pseudo-interferogram (that is by plotting lines or contours of constant out-of-plane displacement) and by then superimposing the experimental and numerical contour lines. Figures (2.3) through (2.10b) provide such a comparison. These figures are to be found at the end of this chapter. It is first necessary, however, to make a few clarifying remarks.

The original experimental interferograms were captured in one large photographic image at each load step increment applied to the specimen. The field of view on the specimen was an octagonal region of 40.96 mm width across the flats of the octagon which corresponds to $\sim$ 4 plate thicknesses. At each illustrative load point a $10X$ and $100X$ magnification (of life size) from a portion of the photographic negative are displayed. The contrast is reversed between the prints of the two magnifications. The selected load points (for the center load indenter) are 0

kN , 35 kN , 52.3 kN , and 73.5 kN , which may be compared with the load for unstable crack propagation of approximately 78 kN. It should also be noted that crack tunnelling (initial development of the cleavage portion of the ultimate fracture surface before the development of the shear lip portion of the fracture surface) begins at approximately 68 kN. The experimentally determined crack profile versus load was established by heat-tinting a group of specimens each of which was taken to a maximum load point less than the ultimate sustainable load of 78 kN, thereby each giving an indication of the geometry of the developing fracture up to that load point.†

The wavelength (λ) of the laser light source employed in the interferometer was 632.8 nm. The out-of-plane displacement increment associated with a repetition in the gray scale of the interferogram (from black fringe to black fringe or from white fringe to white fringe) is $\lambda/2$. The FEM analysis plots the predicted contour locations in contour increments of $5(\lambda/2)$. Also, the numerical contours are plotted in the deformed specimen coordinates, since there is sufficient in-plane stretching that the shape and spacing of the contours when plotted relative to the undeformed in-plane coordinates is significantly in error.

A very small rigid body rotation is superposed onto the FEM calculation of the deformed surface. These differing rigid body rotations for each load point are small, but strictly speaking necessary, since the experimental interferograms sometimes retain a small but not negligible amount of rigid body rotation.

The initial deviation of the specimen surface from perfect planarity in the unloaded state is illustrated in figure (2.3), a $10X$ magnification of fringe contours

† The approximate nature of the crack tunnelling profile as a function of the applied load was introduced into the FEM analysis by releasing nodes of the mesh in the plane of the cleavage crack tongue in simulation of the crack propagation with load.

for the 0 kN load point. The straighter and more evenly spaced the interferometric fringes the smaller the deviation of the surface from perfect planarity. The evenly spaced and parallel black lines superimposed on the interferogram give a visual indication of the inital out-of-flatness of the specimen. This level of imperfection will be inconsequential relative to the discrepancy between experiment and simulation observed at the higher load points. The initial undeformed shape of the notch tip may be observed from figure (2.4), a $100X$ magnification of the 0 kN load point.

There are differences between the FEM approximation and the measured out-of-plane displacements which are clearly discernable. The primary of these differences is that the simulation underestimates these displacements in the region surrounding the notch tip by a fraction which becomes progressively larger at points nearer to the notch tip and progressively larger with increasing load. The discrepancy between the experimentally observed out-of-plane displacement at the notch tip and the numerically predicted displacement is approximately 9 % for the 35 kN load point and approximately 20 % for the 73.5 kN load point.

These differences may be reduced by relatively obvious improvements to the material modeling and FEM analysis (or to a lesser degree by improvements to the experiment); alternatively, the discrepancy between simulation and experiment may result from something as yet unforseen.

2.2 Principle Sources of Error

2.2.1 Errors in Numerical Simulation

Some of the obvious improvements which may be made to the FEM analysis are as follows:

a) The finite element mesh could be made substantially finer. Generally, the finiteness of the number of mesh elements leads to an underestimation of the displacements.

b) Now that the approximate nature of the deformation field in this specimen is known, the specimen could be remeshed with a mesh distribution which better reflects the nature of the strain field. For example, meshing with uniform

(planar) layers of elements, as is the case with the present analysis, is simple to implement but underutilizes a fixed number of elements. Similary, using equal angular intervals between radial lines in the in-plane mesh pattern is inappropriate given the higher shear strains found in the regions next to the notch tip at approximately 45° to the symmetry line ahead of the notch.

c) The FEM code used was not a finite deformation code and, since the strains near the notch tip become substantial (> 70%) for this ductile variety of steel even for loads which are well below that ultimately sustainable, using a small deformation kinematic description would also overestimate the stiffness of the material as the material description was formulated in "true strain" while this code is, in effect, operating in "nominal strain."

d) A somewhat more realistic constitutive model might be implemented. The present analysis used a power–law–hardening description, which is probably unnecessarily unsophisticated. For example, power–law–hardening ignores the phenomena underlying Lüder's band behavior, which does exist in this material and is readily observable experimentally.

e) The notch shape implemented in the numerical analysis consisted of straight and parallel sides with a semi-circular notch tip and a straight notch tip front through the thickness of the plate. The spacing between the parallel sides of 0.3 mm was thought to be of a comparable size scale to the surface roughness associated with cleavage cracks in 4340 steel. Unfortunately, at such small notch dimensions there are relatively few machining and fabrication processes which may be employed to create the actual notch in the steel specimens. An electrical discharge (wire cutting) machine was used to form this very slender notch in the test specimens discussed here. The electrical discharge machine creates a very smooth and reproducible notch tip shape, but it is not a semi-circular curve. The actual tip shape is closely approximated by a parabola with a small faired region connecting the parabolic tip section to the straight and parallel sides of the notch. It would improve the exactness of the comparison between

numerical analysis and experiment if the numerical mesh more closely matched the electrical discharge cut notch tip, and this may be easily accomplished if the results of the analysis are, in fact, sensitive to differences in the notch tip shape.

2.2.2 Errors in Material Characterization

An obvious improvement which may be made in experimentally establishing the constitutive description which is provided to the computer code is as follows:

It has been the practice to develop a constitutive description for the wide variety of structural materials using a series of simple loading geometries, such as uniform uniaxial, biaxial, and tension–torsion loading. However, even in such simple geometries, it is not possible to keep the deformation fields simple at larger strain levels since, very often, an instability develops, such as necking in uniaxial tension. This added complexity in establishing a large strain constitutive model from the specimen may be overcome by using more sophisticated techniques of measurement once necking, etc. has occurred : The same techniques of moire interferometry and Twyman-Green interferometry may be applied to establishing a more precise constitutive description at higher strain levels from observations of the "simple" (uniaxial, biaxial, tension–torsion, etc.) loading geometries.

2.2.3 Errors Associated with the Fracture Test

As regards the estimation of experimental error in the interferometric data collected from the test of the notched three-point-bend specimen, the interferometric techniques themselves are quite accurate, and most of the uncertainty in the observed experimental deformations arises from statistical variations from specimen to specimen resulting from an inability to more closely control the various influencing parameters. However, certain of the physical parameters which cause a significant percentage of the observable variation in deformation for different specimens may be better controlled as follows: Three-point-bend loading has associated buckling modes. Imperfections in specimen geometry and loading alignment cause the specimen to warp in these buckling mode shapes to a degree which is proportional to

the load on the specimen and to the weighted deviation in the initial shape of the specimen if we may use elastic buckling analysis as a guide. It is possible to reduce the degree of warpage under load by improving the precision of the load frame and by decreasing the initial (unloaded) geometric imperfection of the specimen. Also, the initial stresses built up in the steel from the heat-treatment process (which also lead to warping under load) can be minimized somewhat by maintaining as much symmetry as possible in the quenching process.

Alternatively, there are other specimen loading configurations which have less tendency to accentuate initial imperfections. Switching from a 3-point-bend geometry to a single-edge-notched tension configuration can eliminate the buckling modes underlying the warpage.

Even with the greatest of attention to detail there is always some assymetry which develops in the plastic zone surrounding the notch tip. The numerical simulation assumes the plastic field to be doubly symmetric. Interferometric measurements can be made on both surfaces of the plate to ascertain more clearly the degree of asymmetry involved. That the occurrence of asymmetry is unavoidable may be deduced from the nature of the fracture surface associated with the shear lips of a fracture steel plate. This shear lip fracture being asymmetric by definition, it is the culminating manifestation of the accentuation of the initial assymetries of the notched plate.

2.2.4 Data Loss Due to Angular Aperature Limitation

The optics used for imaging the interferometric fringes have a fixed angular aperature (angular aperature is the angle between the centerline of the lens and a ray emanating from a point on the centerline which can just be captured and imaged by the extreme outer edge of the lens). This limitation to the angular aperature of a lens introduces restrictions to surface interferometry (both Twyman-Green and Moire) manifested by an absence of fringe data from portions of the surface which are very severely deformed. This aspect of interferometry involves not an error in the collected data, but a complete absence of collected data from within certain

regions of the surface.

Out-of-plane (Twyman-Green) interferometry offers the simplest illustration of this effect. Twyman-Green interferometry involves specular (mirror-like) reflection of light rays from a polished surface. If the incident light rays are perpendicular to the original planar surface of the plate, then the reflected rays emanate from the surface at angles to this perpendicular which are twice the angle between that normal to the undeformed planar surface and the new local normal to the deformed surface. In this situation, it is unreasonable to expect that the imaging lens used can collect and image rays from points on the deformed surface where the rotation of the normal to the surface is approaching 45 degeees. In actuality, functional imaging lenses with angular aperatures of greater than $\pm 45^o$ (corresponding to an observable surface normal rotation of $\pm 22.5^o$) have only been achieved with extremely small lenses (high-power microscope objectives) where geometric abberations have been reduced at the expense of field of view by the very small lens diameter. However, trading a reduction in the observable field of view for an increase in angular aperature is an acceptable sacrifice if the regions of the surface farthest from the crack tip are so well categorized that they no longer need to be observed.

2.3 The Twyman-Green Interferometer

There are actually several similar, but still significantly different, optical layouts which fall under the classification of a Twyman-Green interferometer. There are many subtleties which must be observed in the design of interferometers in general. Some of these subtleties will be introduced in the course of an elaboration on the underlying reasons for the choices made in the arrangement of the optical components for this particular case. First, refer to Figure (2.1) for a schematic of the out-of-plane surface interferometer used in this experiment.

If one were to look up Twyman-Green interferometer in *Principles of Optics*, by Born & Wolf [15], the configuration of optical components that one would find would be essentially the same as that in Figure (2.1). On the other hand, a similar search of *Optics*, by Hecht [16], would turn up a configuration (refer to Figure (2.2))

where the positions of the imaging and collimating (beamforming) optics had been interchanged as compared with the version in Born & Wolf. What are the respective advantages and disadvantages of these two versions? From consideration of Figure (2.1), it is readily apparent that difficulties may be encountered in attempting to sandwich a camera into the region designated for the imaging optics and associated image plane in cases where the specimen (object plane) is extensive. Of course, one may utilize an additional mirror after the imaging optics in order to turn the beam path by 90 degrees. Still, this difficulty may be more readily handled if it is the collimating optics which are located next to the specimen as per the arrangement in Hecht.

However, there is a very important advantage associated with the choice of placing the imaging optics next to the specimen, which easily outweighs the attendant difficulty, where precision performance is required. In Figure (2.1), one may observe that the imaging optics view the specimen by virtue of a reflection off of the half-mirrored surface of the beamsplitter. Alternatively, if the imaging optics were instead to be in the location occupied by the collimating optics in Figure(2.1), then the imaging optics would be viewing the specimen through the glass plate which constitutes the beamsplitter. Severe geometric aberrations (primarily astigmatism) are associated with viewing through a glass plate turned at 45 degrees. No such aberrations are associated with viewing in reflection from the half-mirrored surface of the beamsplitter so long as the mirrored surface is optically flat. So, while the configuration in Figure (2.1) may seem more cumbersome, it is by far the more precise. This little trick of configuration is possible because the one spherical harmonic wave form which is undistorted by passage through a glass plate at 45 degrees is the plane wave, and hence the truncated plane wave reflected from the reference mirror is not significantly adversely affected by its additional passage through the glass beamsplitter.

Of course, one might think that the solution would be to replace the beamsplitter plate with a beamsplitter cube and thereby avoid the geometric aberrations

associated with the glass plate turned at 45 degrees. However, a glass cube can be the source of extreme spherical aberration. This spherical aberration can be eliminated by placing the beamsplitter into as close as possible proximity to the specimen surface and by introducing refractive index matching oil into the gap between the specimen and the glass of the beamsplitter as is done when using high power microscope objectives. In this case, though, the imaging optics become highly specialized since there are relatively few common imaging systems which incorporate glass cubes located adjacent to an image plane. These difficulties not withstanding, some inexpensive interferometers use a beamsplitter cube as the optical component placed closest to the object.

A fourth form of out-of-plane surface interferometer is to be found in expensive microscopes. One may note that the form of the interferometer sketched in Figure (2.1) is limited in its f/number to no better than 1. To improve upon this angular aperture, one needs to place some of the imaging optics closest to the object rather than the beamsplitter. This arrangement puts very severe requirements on the imaging optics. To illustrate these requirements, consider the case where both the object and the reference mirror are optically flat. In this case, one is, in effect, performing interferometry on the imaging optics between the object and the beamsplitter, and any deviation from an ideal optical transform device by these imaging optics will result in interference fringes, even though the object is flat. Such "ideal" optics must have negligible geometric aberrations and, as well, require optical surfaces which are everywhere within a fraction of the wavelength of light of their intended shapes. Such "ideal" optics with f/number better (smaller) than 1 are generally only possible at present in very small lenses (microscope objectives) where the associated geometric aberrations, which scale with size, have been diminished along with the size of the optical components. Additionally, expensive and sophisticated tecchniques must be employed in the fabrication and mounting of the various lens elements.

Since the present experiment involved a field of view of approximately 50 mm,

it was not to be expected that this fourth option (placing the imaging optics between the object and the beamsplitter) would provide any improvement over the chosen configuration. Present day optics are simply not yet that "ideal."

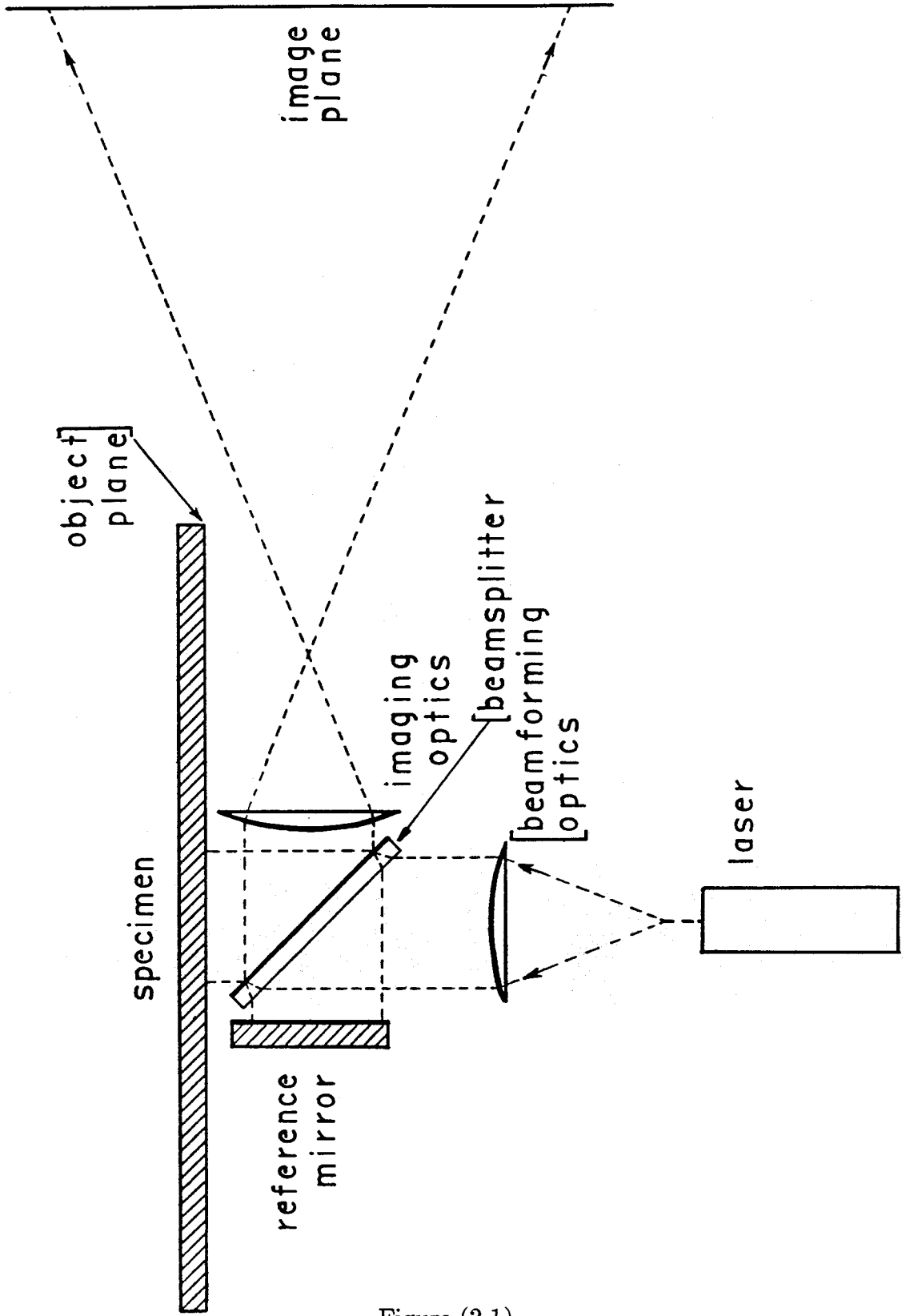


image plane
object plane
specimen
imaging optics
beamsplitter
beamforming optics
laser
reference mirror

Figure (2.1)

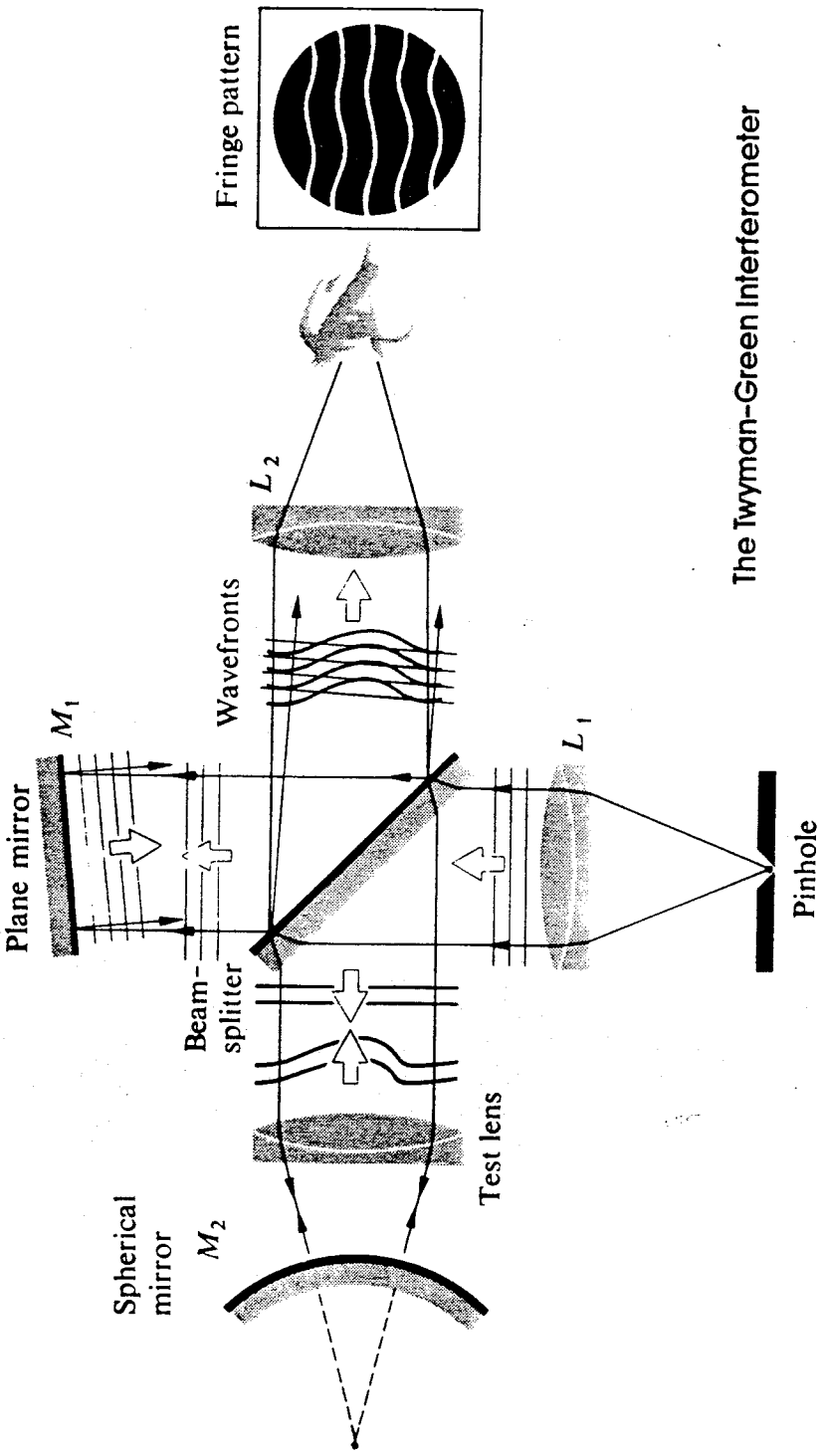

The Twyman-Green Interferometer
Fringe pattern
L_2
Wavefronts
M_1
Plane mirror
L_1
Pinhole
Beam-splitter
Test lens
Spherical mirror
M_2

Figure (2.2)

2.4 Interferograms

Figure 2.3

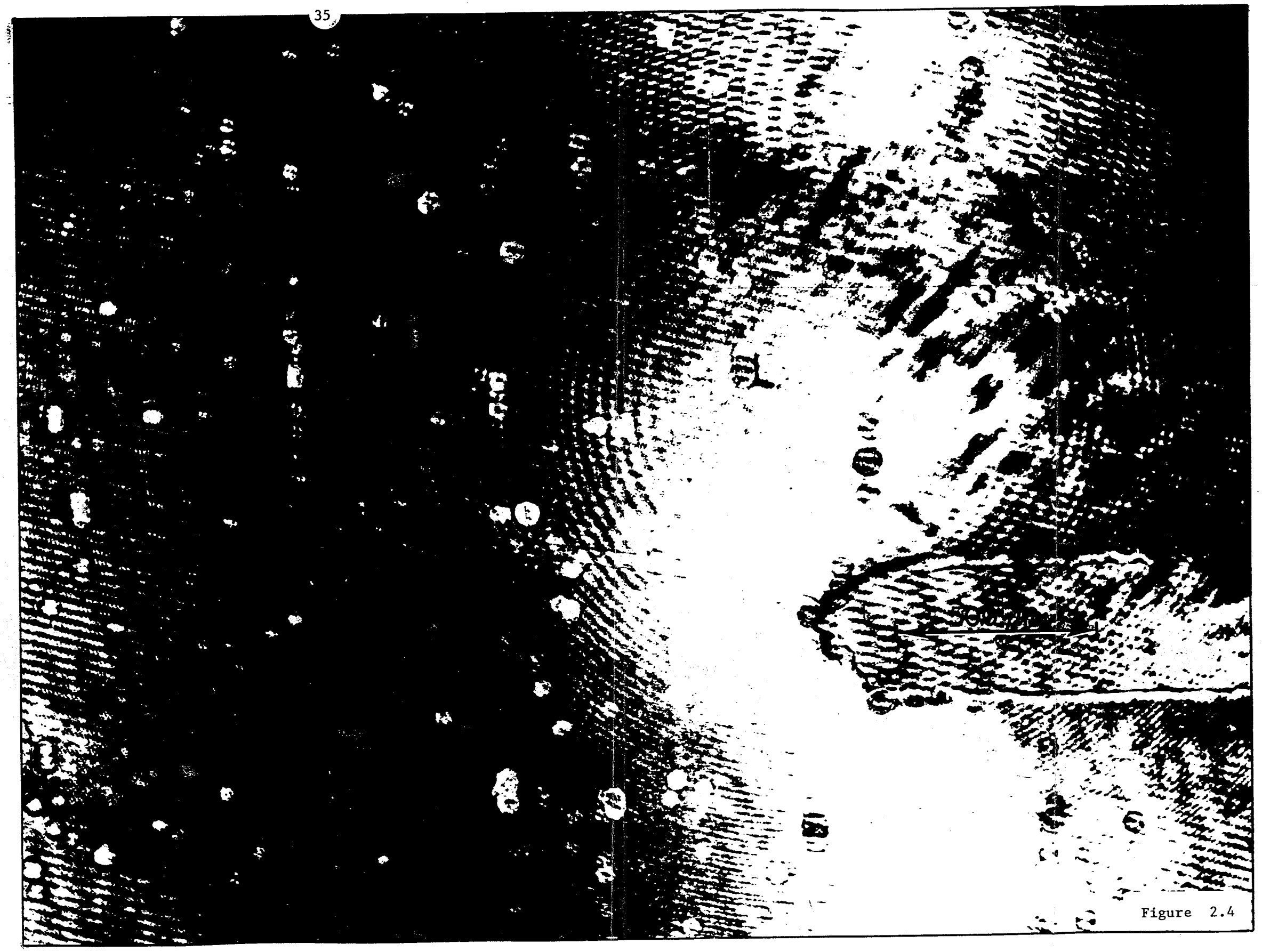


Figure 2.4

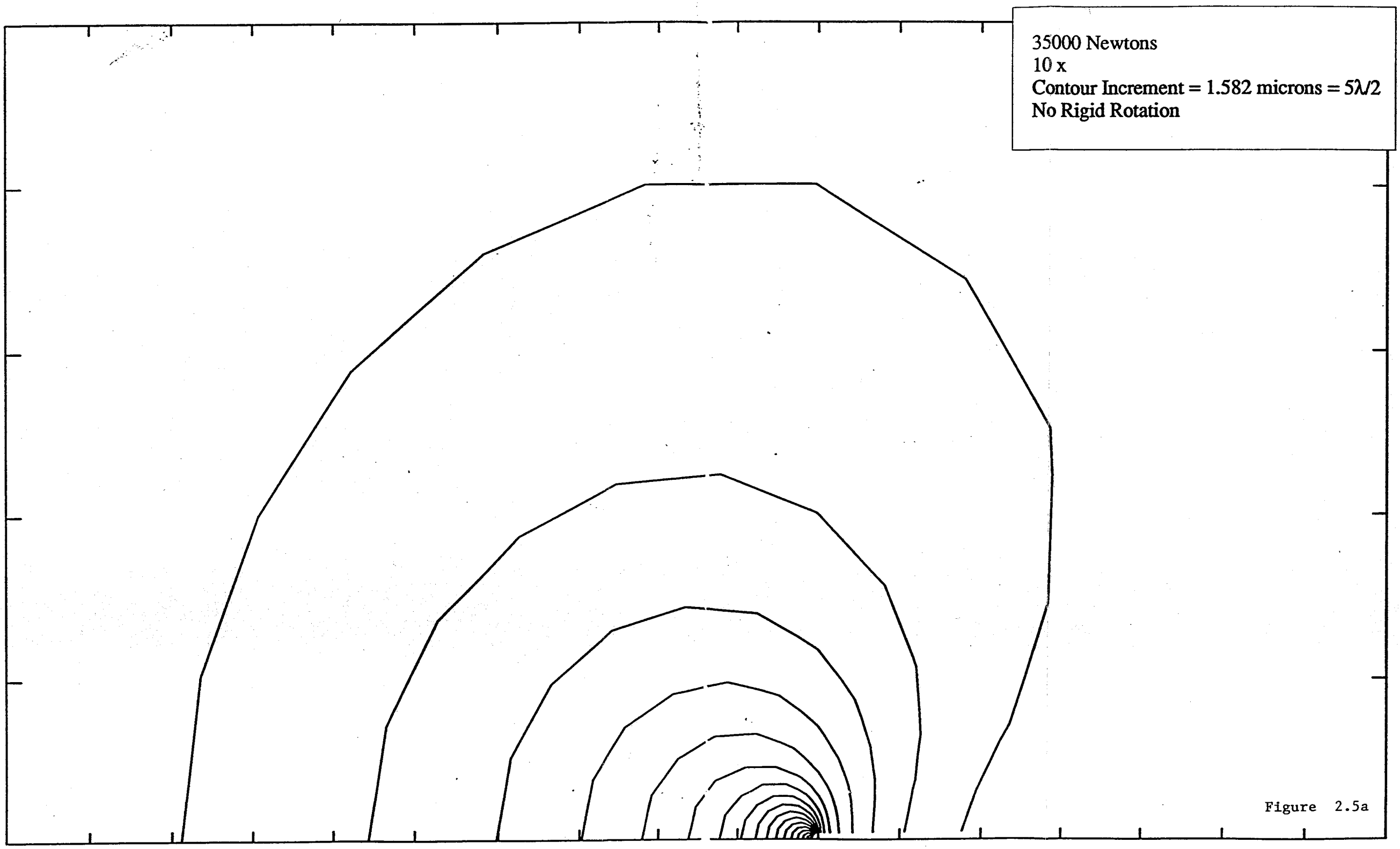

35000 Newtons
10 x
Contour Increment = 1.582 microns = 5λ/2
No Rigid Rotation

Figure 2.5a

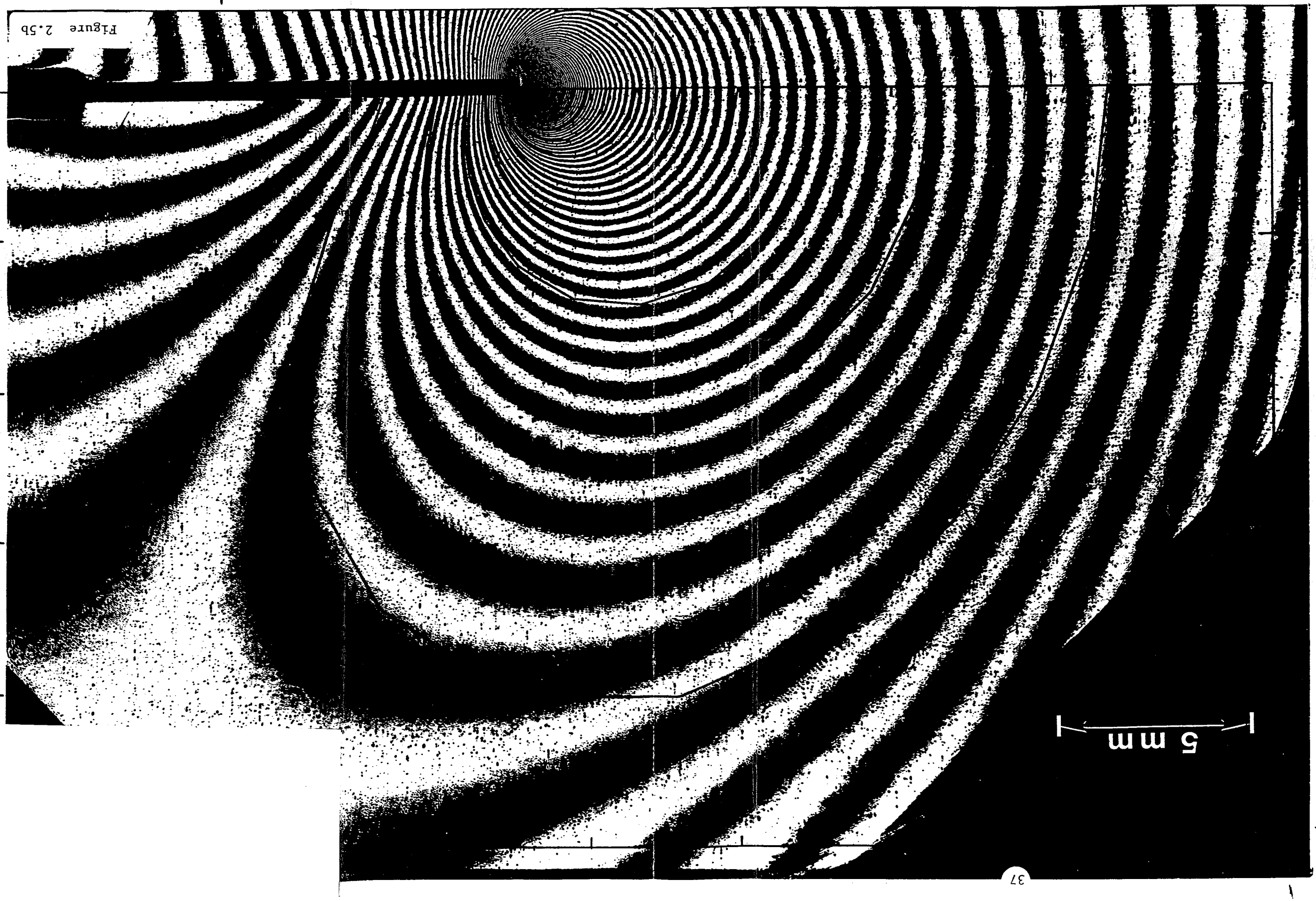


Figure 2.5b

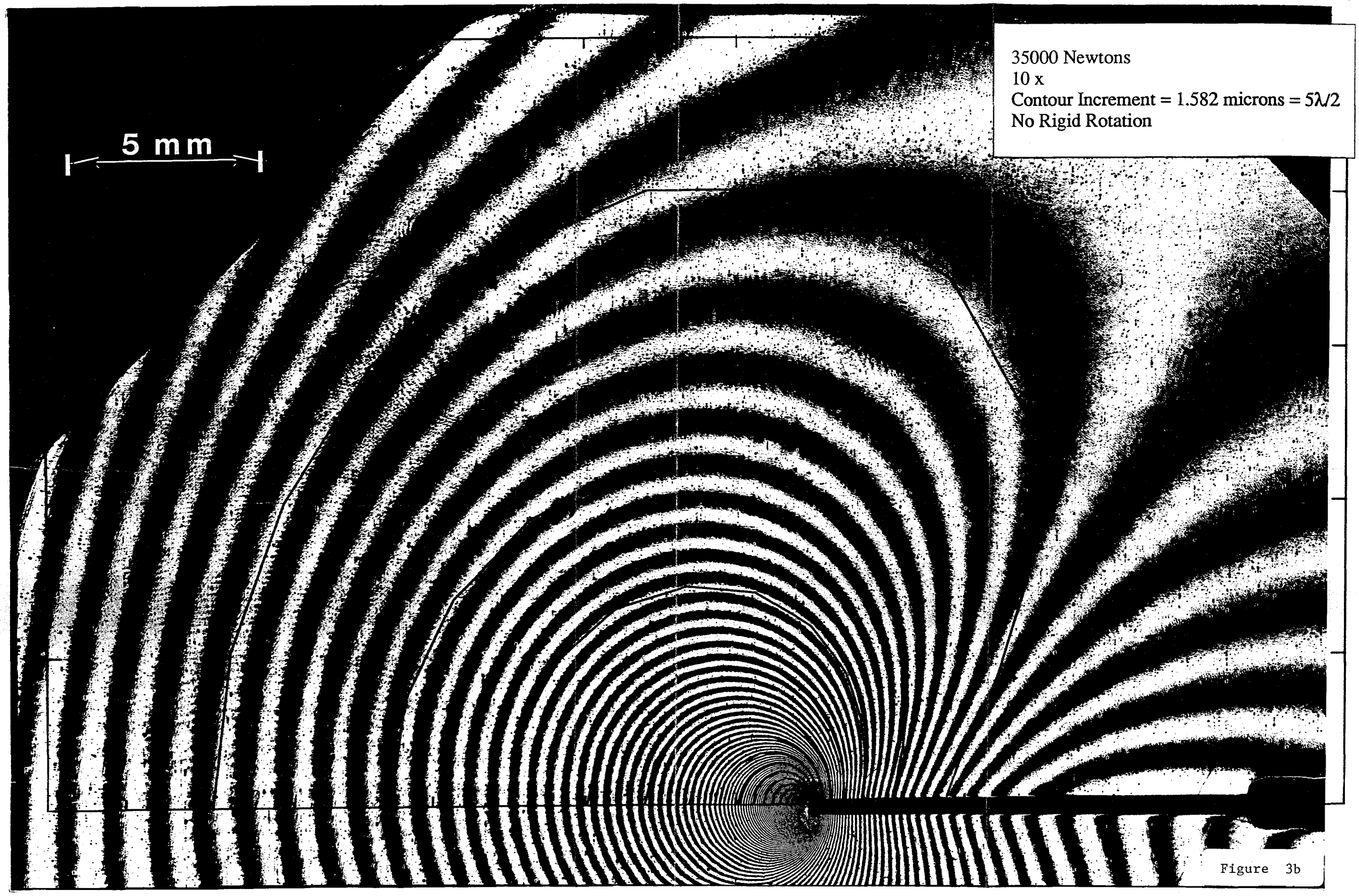

5 mm
35000 Newtons
10 x
Contour Increment = 1.582 microns = 5λ/2
No Rigid Rotation

Figure 3b

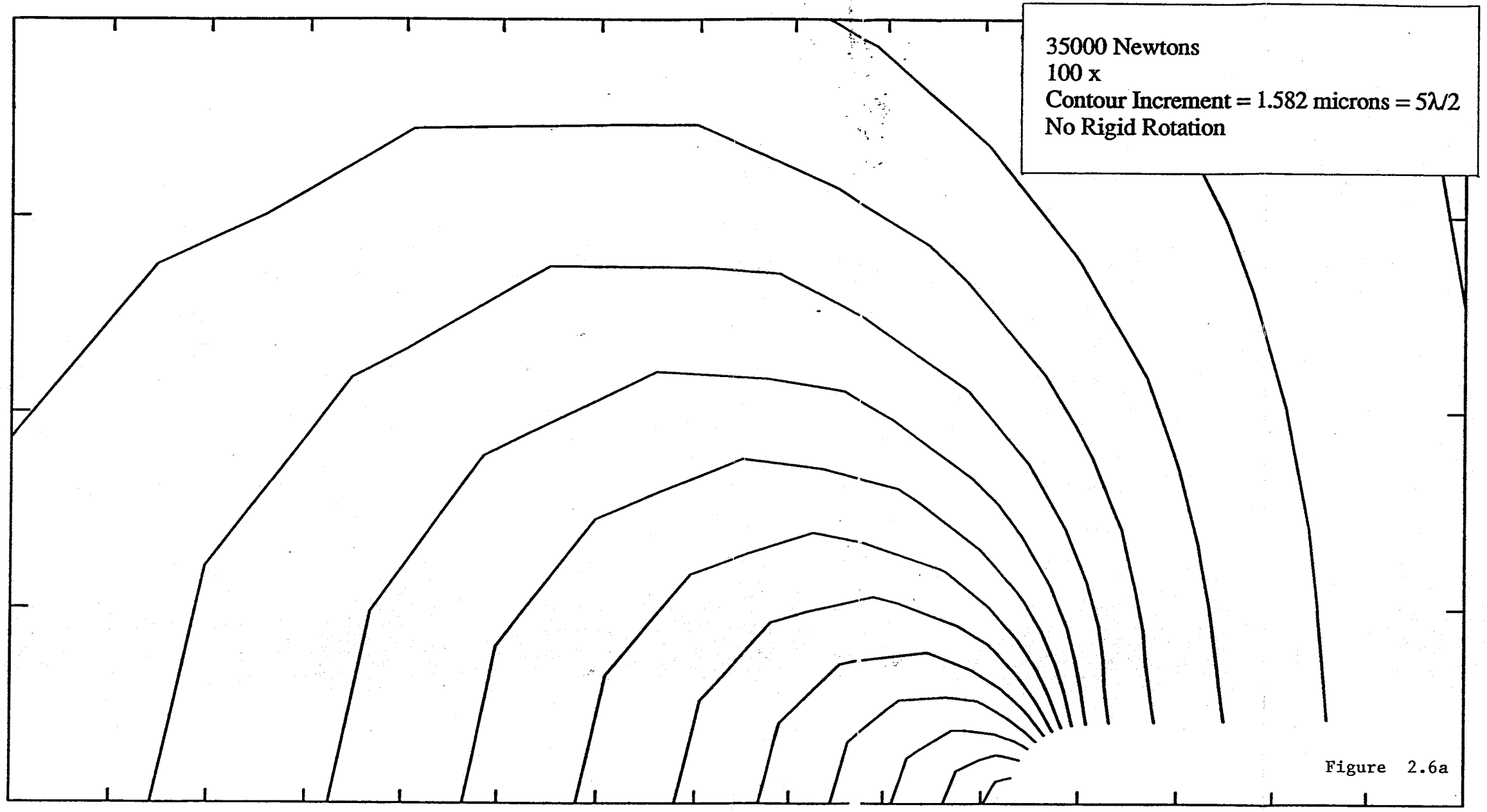

35000 Newtons
100 x
Contour Increment = 1.582 microns = 5λ/2
No Rigid Rotation

Figure 2.6a

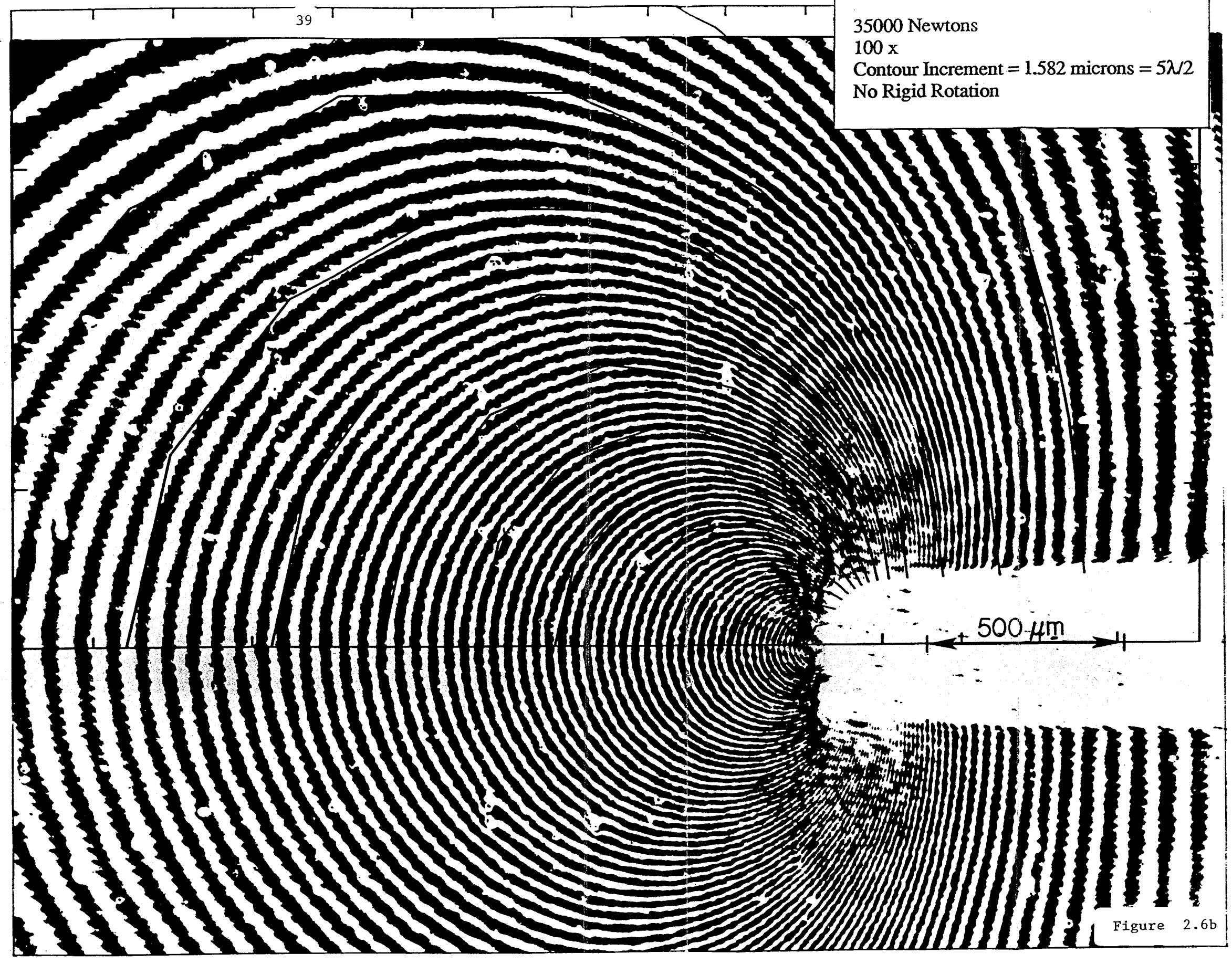

35000 Newtons
100 x
Contour Increment = 1.582 microns = 5λ/2
No Rigid Rotation
500 μm

Figure 2.6b

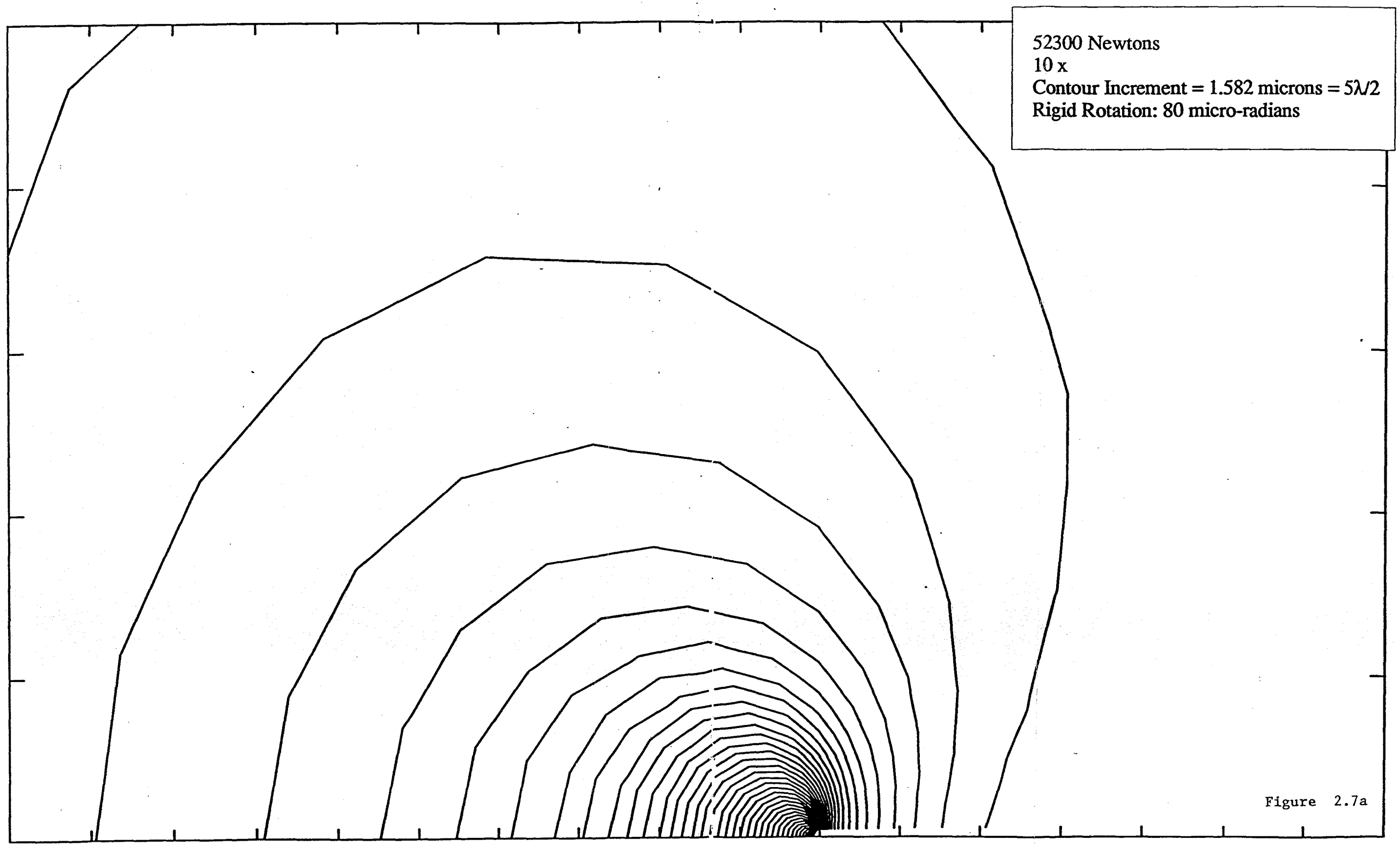
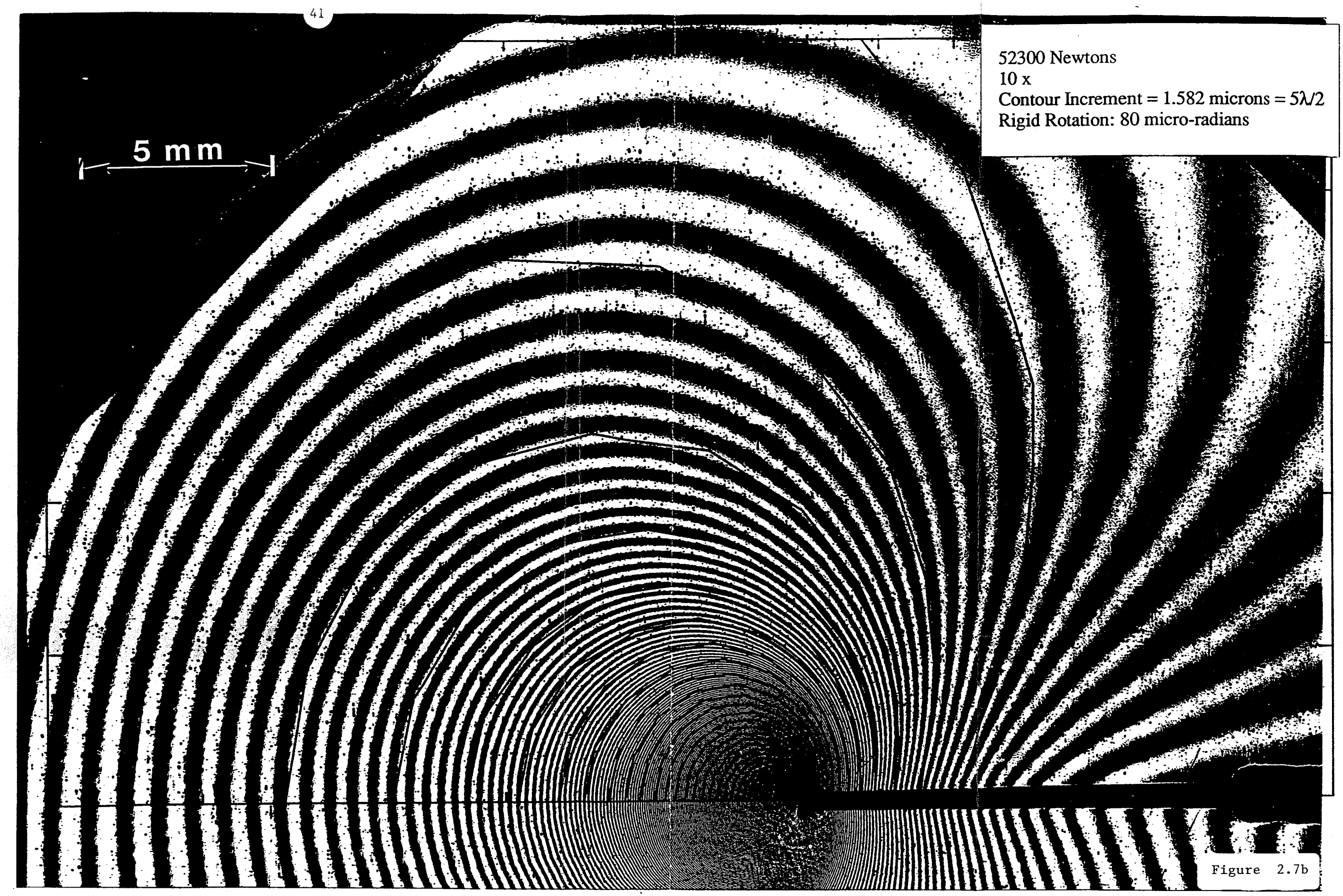

52300 Newtons
10 x
Contour Increment = 1.582 microns = 5λ/2
Rigid Rotation: 80 micro-radians

Figure 2.7a



52300 Newtons
10 x
Contour Increment = 1.582 microns = 5λ/2
Rigid Rotation: 80 micro-radians

Figure 2.7b

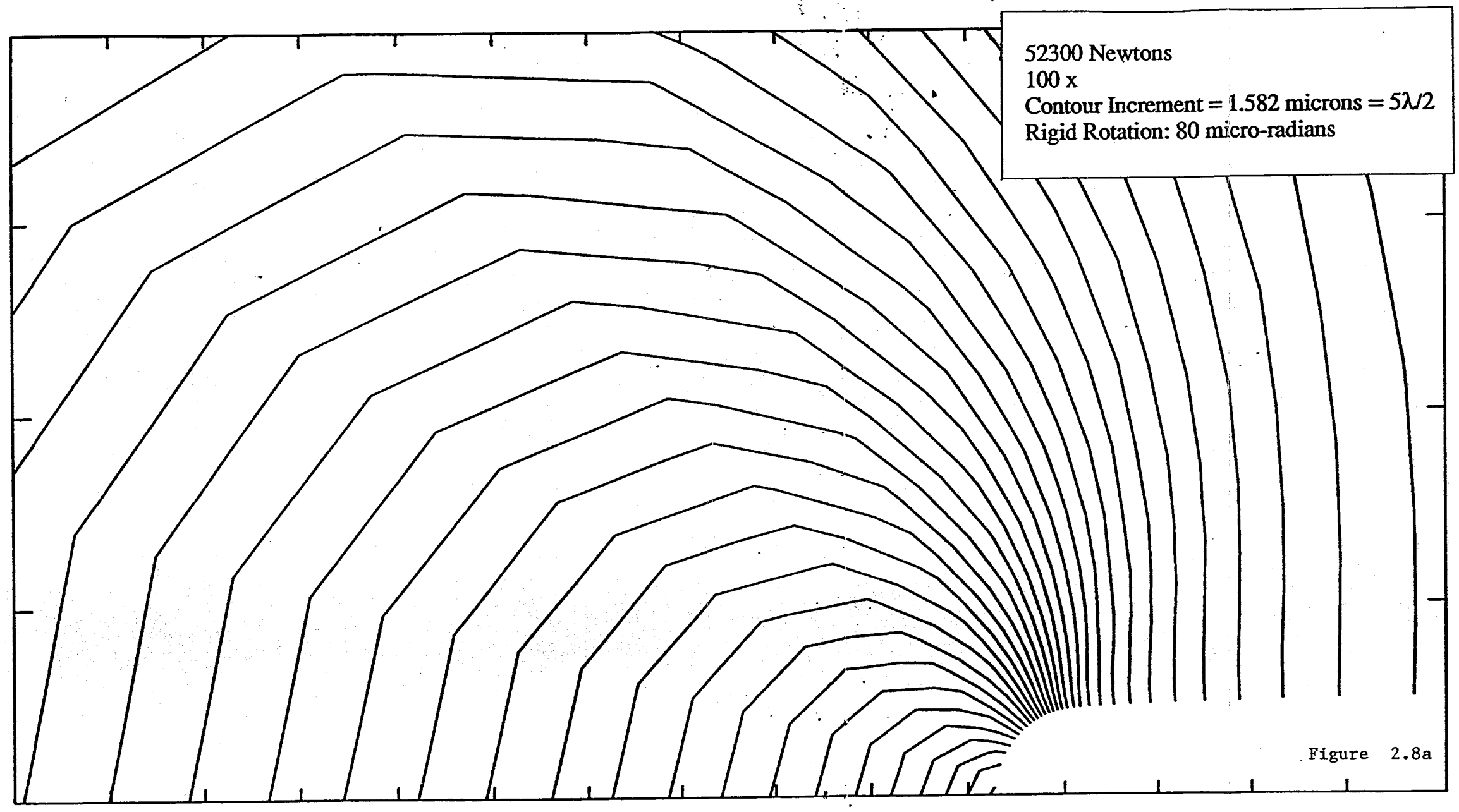

52300 Newtons
100 x
Contour Increment = 1.582 microns = 5λ/2
Rigid Rotation: 80 micro-radians

Figure 2.8a

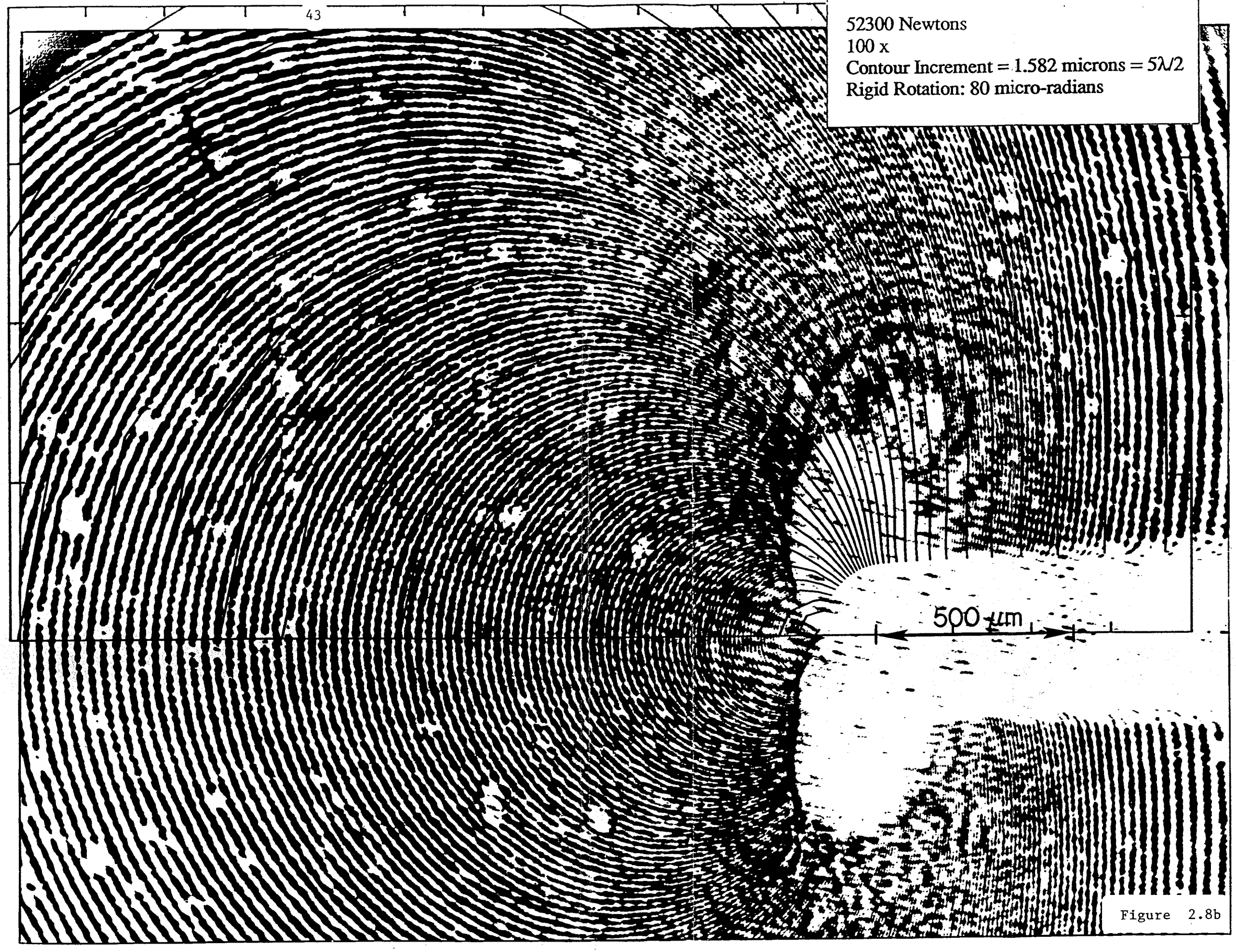

52300 Newtons
100 x
Contour Increment = 1.582 microns = 5λ/2
Rigid Rotation: 80 micro-radians
500 μm

Figure 2.8b

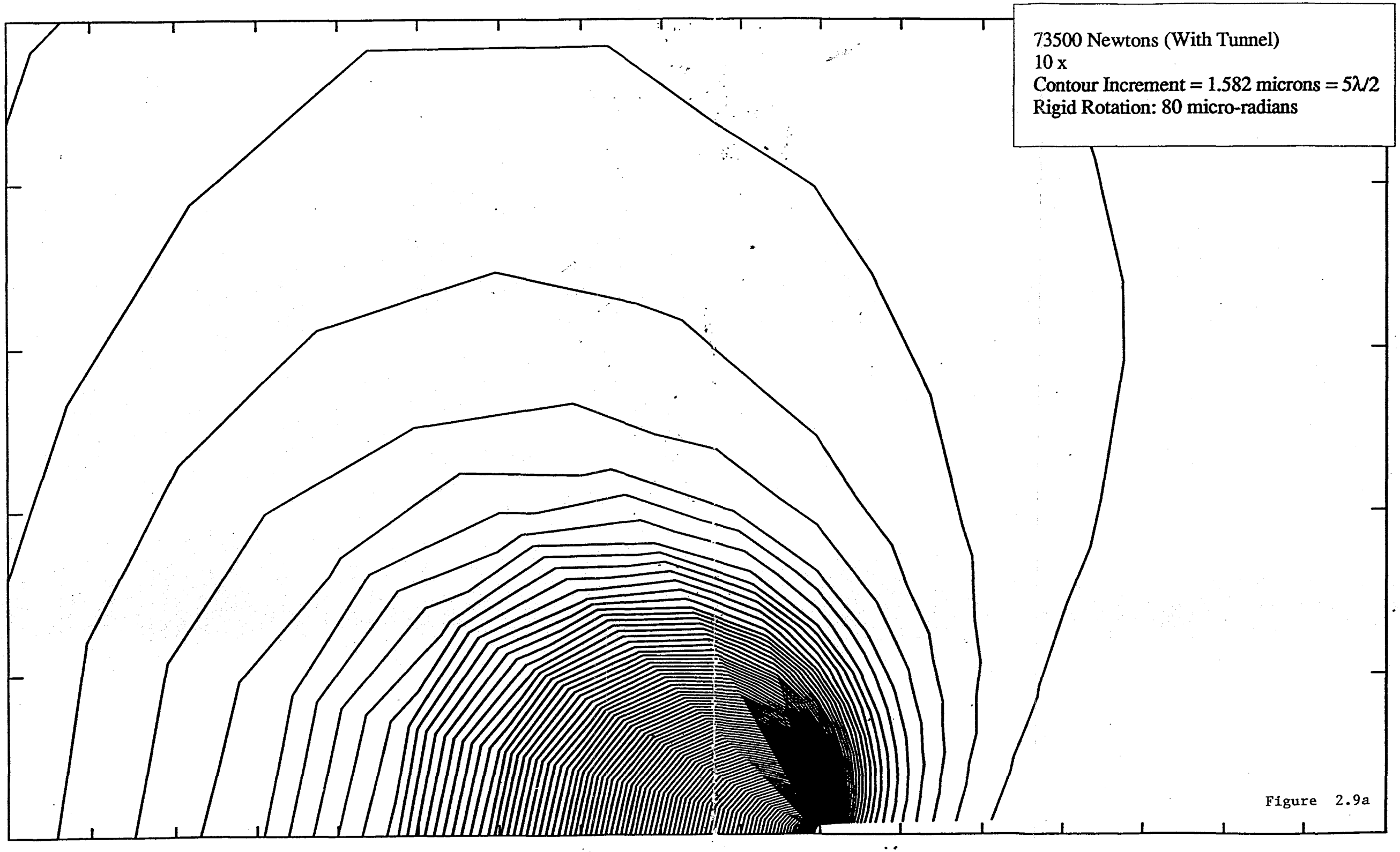

73500 Newtons (With Tunnel)
10 x
Contour Increment = 1.582 microns = 5λ/2
Rigid Rotation: 80 micro-radians

Figure 2.9a

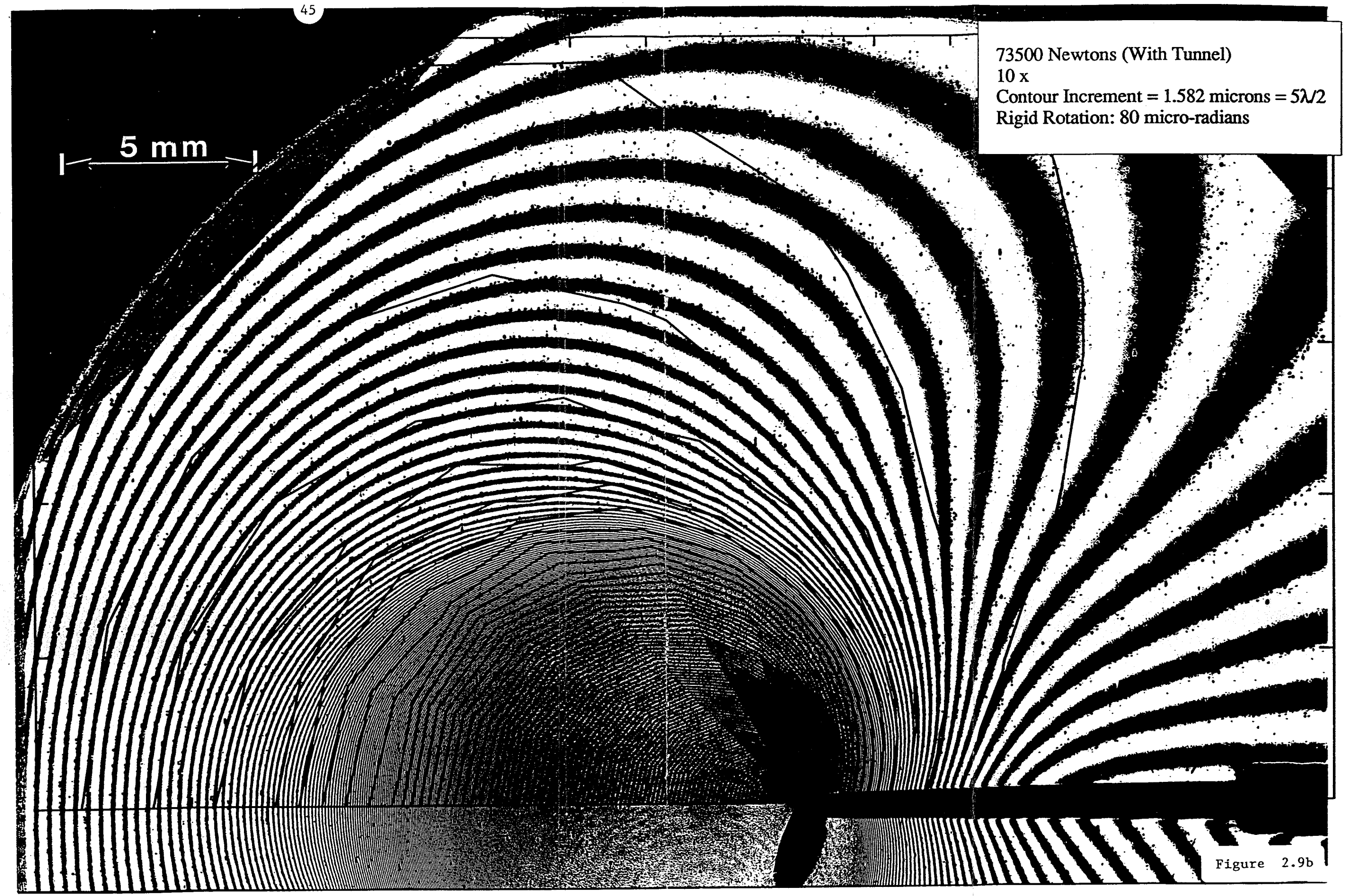

5 mm
73500 Newtons (With Tunnel)
10 x
Contour Increment = 1.582 microns = 5λ/2
Rigid Rotation: 80 micro-radians

Figure 2.9b

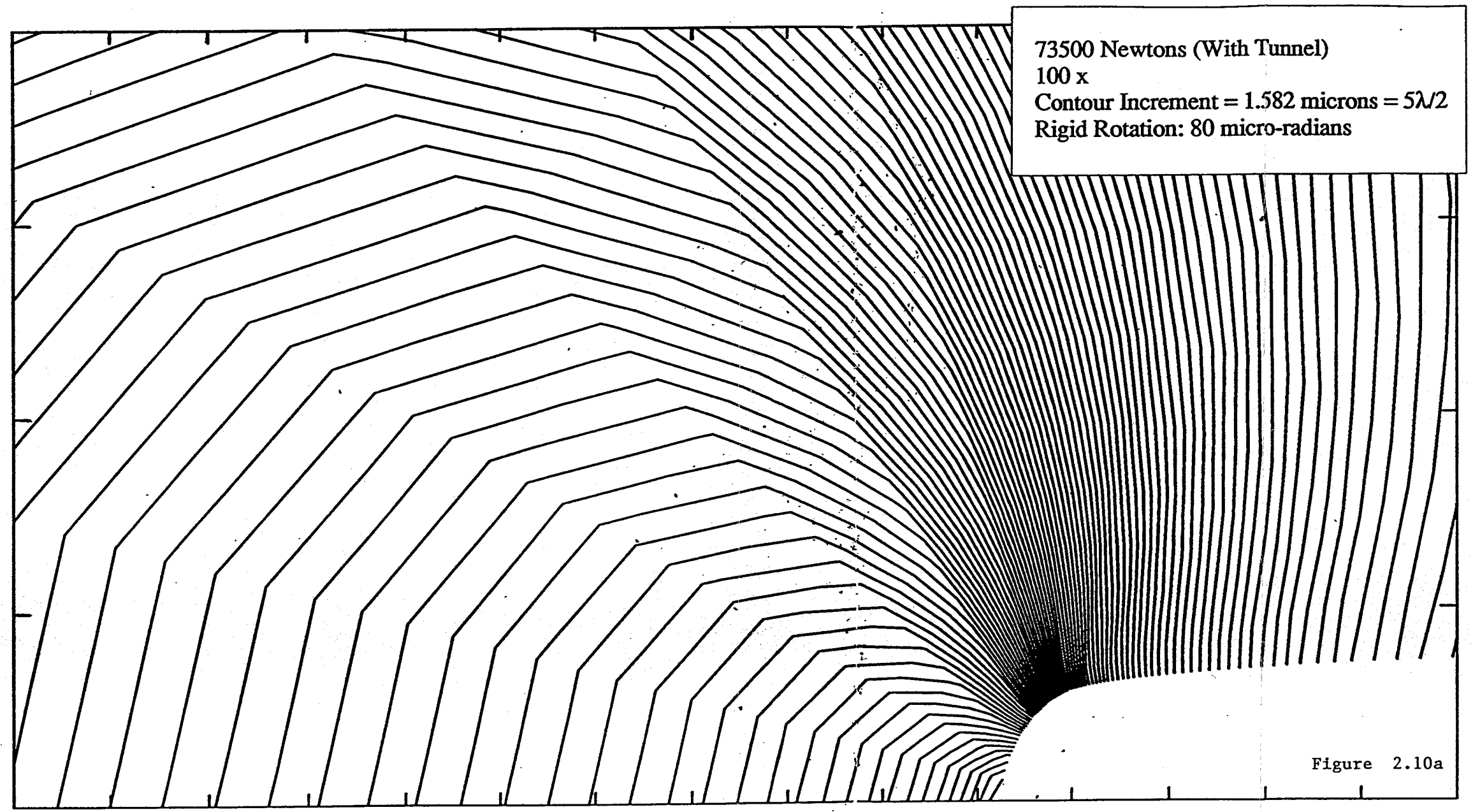

73500 Newtons (With Tunnel)
100 x
Contour Increment = 1.582 microns = 5λ/2
Rigid Rotation: 80 micro-radians

Figure 2.10a

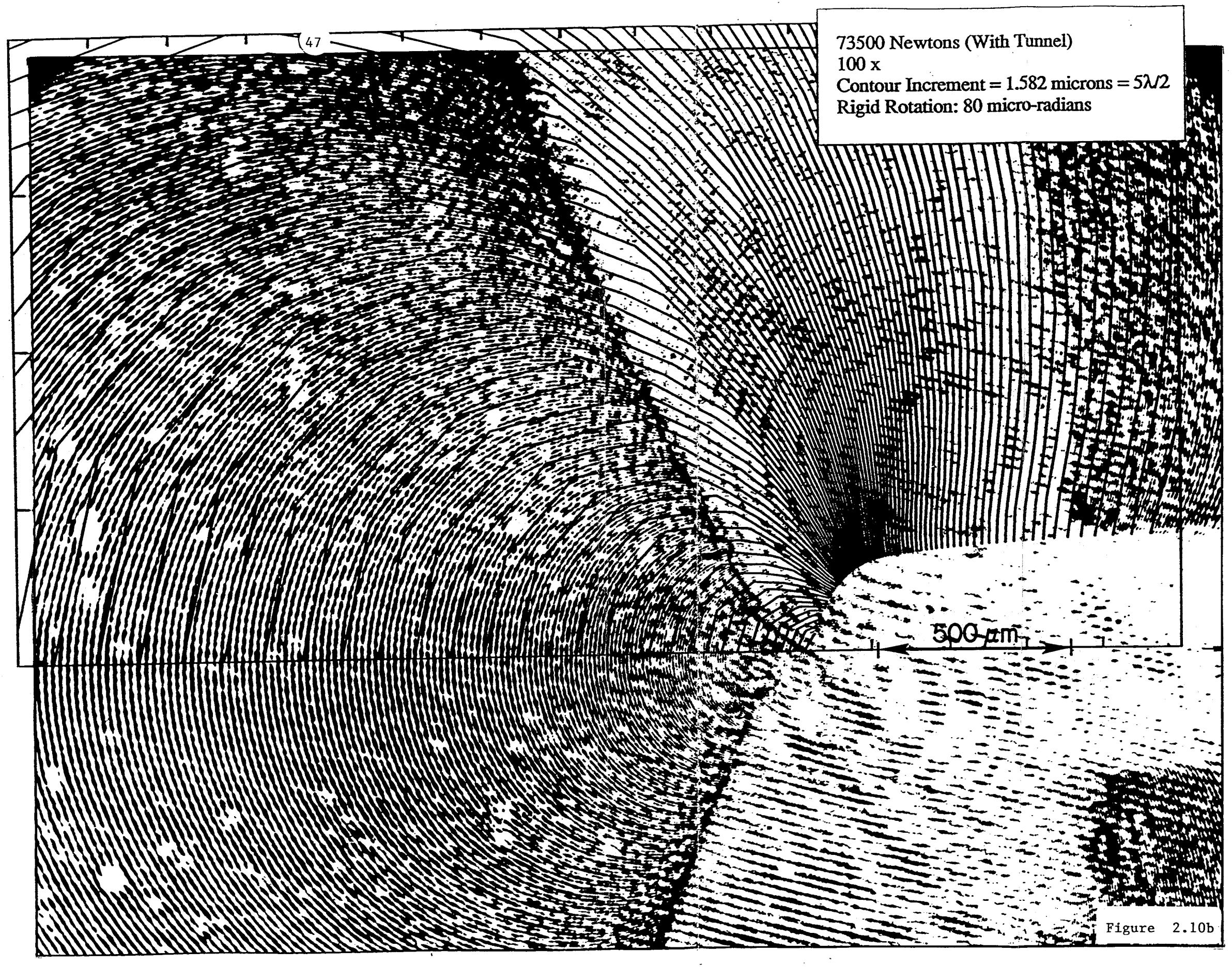

73500 Newtons (With Tunnel)
100 x
Contour Increment = 1.582 microns = 5λ/2
Rigid Rotation: 80 micro-radians
500 μm

Figure 2.10b

Chapter 3
High Speed Photography

The general rule of high speed cameras is that more is better——more frames/time, more data/frame, more total information storage, greater light collection capacity (numerical aperture), increased accuracy of time resolution, broader range of spectral sensitivity, better quantum data efficiency (photon efficiency), more powerful light sources, greater possible variation in magnification (image size/object size) ratios, etc. [17]. The limitations on these performance characteristics may be split, broadly speaking, into three categories: 1) Fundamental (theoretical) constraints; 2) Technical limitations; 3) Economic considerations. Such a division is subject to interpretation, and particularly dangerous is any thought process which unwittingly turns a technical limitation into a perceived theoretical constraint through a failure to view the problem broadly enough. However, it is possible to show that the performance of present day high speed cameras is very much less than what would be expected on the basis of physical limitations. The following discussion then resolves the major technical problems to allow improvements by several orders of magnitude over state-of-the-art cameras. During this process the various economic considerations will be introduced as they relate to the different performance improvements.

3.1 The Primary Theoretical Constraint

One can think of a high speed camera as a switching device, and as such, it shares certain constraints with any other "device" of this class. The basic time interval for an alteration of state of the switch is often determined by a limiting length scale and an associated wave speed for the switching phenomenon.

In this context, consider the example of rotating mirror cameras. As part of an optical system, a rotating mirror causes the image created by that optical system to be swept in space-time. The length scale for this optomechanical switching is the optical spot size or diffraction resolution limit. The relevant speed is determined by the velocity with which the image of the object is swept out in the image plane

of the associated optical spot size. In the case of the rotating mirror camera, the underlying sweep rate derives from the burst speed for a rotating solid body (in this case, a rotating mirror). This burst speed is determined by the strength to density ratio of the material which is limited by the nature of interatomic bonds, to a fixed fraction of the stiffness to density ratio of the substance. In turn, the square root of the stiffness to density ratio is the material's mechanical wave speed.

The diffraction limit for the angular resolution of an optical system with aperture D and operating at a monochromatic wavelength of λ is $\Delta\Theta = 1.22\lambda/D$ [16]. The angular sweep rate of an optical ray reflected from the surface of a rotating mirror is dependent on: 1) the angle between the axis of rotation of the mirror and the normal to the surface of the mirror, and 2) the angle between the axis of rotation of the mirror and the incoming optical ray. The sweep rate is maximized when both of these angles are $\pi/2$, in which case the sweep rate is twice the rotation rate of the mirror ($\frac{\Delta\Theta}{\Delta t} = 2\omega$, where ω is the mirror rotation rate). Thus the rotating mirror's optical "switching time," Δt, is

$$\Delta t = (0.6)\lambda/\omega D \tag{1}$$

Let us now relate the rotating mirror's spin rate (ω) and optical aperture (D) to its internal stress state. In this process one would also like to maximize the allowable ωD by finding that mirror shape with the highest "figure of merit." This work has been accomplished by Brixner [18] by considering the class of mirrors of prismatic shape with regular polygonal cross-section. It is not too difficult to see intuitively that the cross-sectional shape with the largest allowable ωD (i.e., the smallest switching time, Δt) will be an equilateral triangle when D is taken as the dimension of the reflecting face.

To estimate the maximum allowable ωD for a given material selection, one can turn to the critical rupture value for a rotating circular disk. The rupture value for an equilateral triangle will be only fractionally different from that of a circular disk.

From [19], the linear elastic plane stress solution for a rotating circular (solid) disk of homogeneous isotropic material with a radius of $\frac{D}{2}$ and a rotational speed

of ω is

$$\sigma_{rr} = \rho\omega^2 \left(\frac{D}{2}\right)^2 \frac{3+\nu}{8}\left[1-\left(\frac{r}{D/2}\right)^2\right]$$

$$\sigma_{\theta\theta} = \rho\omega^2 \left(\frac{D}{2}\right)^2 \frac{3+\nu}{8}\left[1-\frac{1+3\nu}{3+\nu}\left(\frac{r}{D/2}\right)^2\right]$$

$$\text{Hence,} \quad (\omega D)_c = 4\sqrt{\frac{E}{\rho}}\left[\frac{2}{(1-\nu)(3+\nu)}\right]^{1/2}\left[\frac{\epsilon_{\theta\theta}+\epsilon_{rr}}{2}|_{r=0}\right]^{1/2}$$

Recall that $\sqrt{\frac{E}{\rho}}$ is the familiar bar wave speed, and to explore design limitations, consider several material choices:

Case #1: Common structural materials (steel, titanium, aluminum, fused silica)

$$\left(\sqrt{\frac{E}{\rho}} \approx 5000\,\text{m/s}, \nu \sim \frac{1}{3}\right); \left[\frac{\epsilon_{\theta\theta}+\epsilon_{rr}}{2}|_{r=0}\right] < 10^{-2} \Rightarrow (\omega D)_c < 1900\,\text{m/s}$$

Case #2: Beryllium

$$\left(\sqrt{\frac{E}{\rho}} \approx 13,000\,\text{m/s}, \nu \approx 0.04\right); \left[\frac{\epsilon_{\theta\theta}+\epsilon_{rr}}{2}|_{r=0}\right] < 4\times10^{-4} \Rightarrow (\omega D)_c < 860\,\text{m/s}\ddagger$$

Case #3: Synthetic Sapphire

$$\left(\sqrt{\frac{E}{\rho}} \approx 9600\,\text{m/s}, \nu \sim 0.0\right)\dagger; \left[\frac{\epsilon_{\theta\theta}+\epsilon_{rr}}{2}|_{r=0}\right] < 10^{-2} \Rightarrow (\omega D)_c < 3100\,\text{m/s}$$

Case #4: Single Crystal Diamond

$$\left(\sqrt{\frac{E}{\rho}} \approx 17,000\text{m/s}, \nu \sim 0.2\right)\dagger; \left[\frac{\epsilon_{\theta\theta}+\epsilon_{rr}}{2}|_{r=0}\right] < 4\times10^{-2} \Rightarrow (\omega D)_c < 12,000\text{m/s}$$

By 1962 Anderson [20] had achieved a rotation rate of 32,000 rev/sec for a triangular mirror constructed of steel with a mirror face dimension of 8 mm. This

‡ This speed may be compared with a state-of-the-art rotating mirror, a Cordin Model 300, which has a triangular mirror with a top rotation rate (ω_{max}) of 8,333 revolutions/second and a mirror face dimension (D) of 17.0 mm leading to a $\omega_{max}D = 890$m/sec.

† Indicates average properties

corresponds to a ωD of $2\pi(32 \times 10^{-3}/\text{sec})(8 \times 10^{-3}\text{m}) = 1600$ m/sec. Since that time, commercial camera manufacturers have chosen mirrors made of beryllium, because that material has a bar wave speed $2\frac{1}{2}$ times higher than the common structural materials. It is important to note that the rotating mirror must not only survive the high angular velocity but must also remain sufficiently undeformed in order that perform its function as an optical component.

Technical issues aside, it is fairly clear that commercial rotating mirror cameras are within an order of magnitude of the maximum possible ωD which is allowed by chemical bonding energy; more importantly, this statement has been true for over three decades.

Returning to the calculation of the switching time for a rotating mirror by taking a reasonable value for ωD of 1200 m/sec and a visible optical wavelength for λ of 0.5×10^{-6} m and placing these values into equation (1), $\Delta t = 0.6\lambda/\omega D$, one obtains a value for Δt of 250×10^{-12} sec. Similarly, if one uses the upper limit for ωD of near 12,000 m/sec, and an ultraviolet wavelength for λ of 0.2μm (the lower transmission limit for ultraviolet grade synthetic fused silica is approximately 0.18μm), then $\Delta t \sim 10 \times 10^{-12}$ sec.

As a basis of comparison: a Cordin Model 179-LLL Image Converter Streak Camera (using cathode ray tube modulation) has a time resolution of 10×10^{-12} sec; by 1989 the faster laboratory semiconductor transistors had achieved switching times as small as $\sim 10 \times 10^{-12}$ sec [21]; Kerr and Pockel cells used as electro-optic light modulators have been used successfully at temporal periods of $\sim 30 \times 10^{-12}$sec [16]. These three types (cathode ray tubes, transistors, and electro-optic cells) are among the most developed and widely used switching devices.

There are, of course, other switching phenomena, governed by higher limiting wave speeds, which result in higher switching speeds. In order to place these considerations into perspective, let us examine the possibility of high speed cameras in the 10-100 billion frames/sec range consistent with the present potentials of the aforementioned devices: rotating mirrors, cathode ray tubes, transistors, and

electro-optic cells.

Beyond constraints on the high speed data recording device, there are always limitations associated with the time rate of change of the phenomena being observed. Ultimately it is the time scale of this event which forces the design of the recording device. In this sense, one can best gain the approximate perspective on a 10-100 billion frames/sec camera by recourse to a consideration of the class of physical events that one might hope to observe.

3.2 Additional Theoretical Considerations

3.2.1 Consideration 1: Image limitations with increased time resolution

Certain experimental techniques have readily identifiable and inherent theoretical constraints on their maximum achievable data collection rate. For example, consider the very simple and obvious constraint on the many coherent light interference techniques (holography, moiré interferometry, Twyman-Green interferometry, etc.) which arises from their use of a pulsed light source. The pulse has to be long enough to be coherent. While the frequency of visible light is indeed high (600×10^{12} Hz for $\lambda = 500$ nm), it is such that visible wavelength pulses with durations in the low femtosecond regime are, by necessity, white light pulses, i.e., noncoherent.

A somewhat less obvious, but analogous, constraint for these same coherent light techniques is inherent in the relationship between the interference fringe density, the extent (or size) of the fringe field, and the necessary pulse length. The easiest way to see this is to consider the interference fringes created by two collimated plane waves crossed at an angle to one another (cf. figure (3.1)). Because the two beams have a finite pulse length, they coincide in space-time only over a finite diamond shaped region. This diamond shaped region moves forward along the bisector of the directions of propagation of the two component beams, interacting with a recording device on which it is impinging.

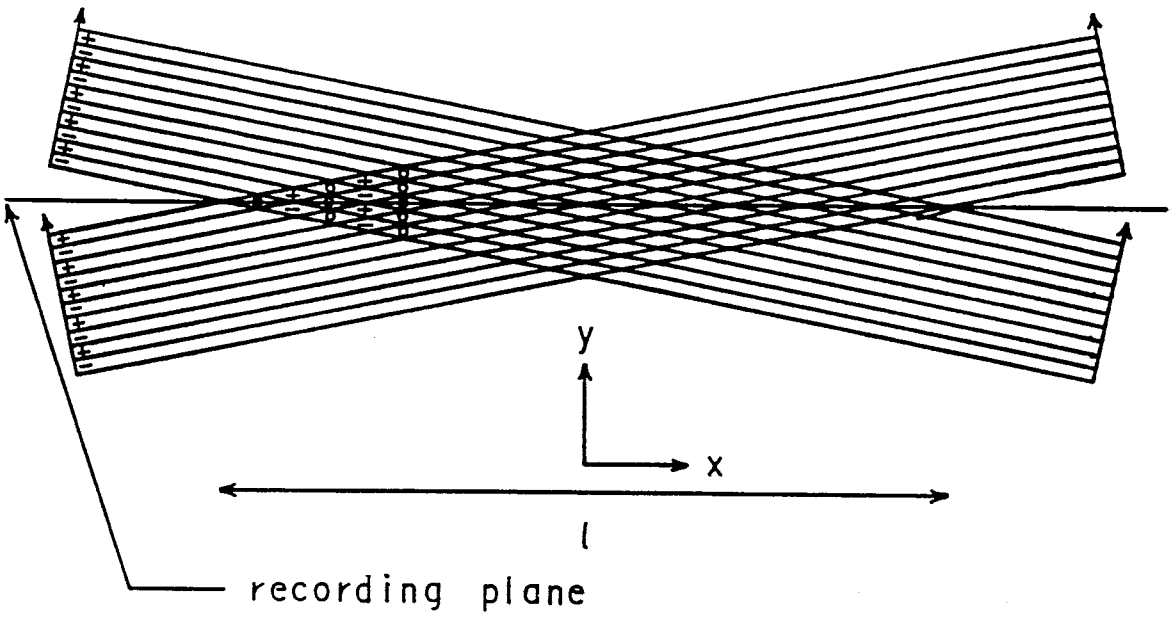


Figure (3.1)

If one looks at the variation in recorded energy for the recording plane indicated in the diagram, then the standing wave which results from the interference of the two beams exists only within the limits of the intersection of the recording plane and the diamond shaped region. The other radiant intensity (outside of the diamond shaped area) to which the recording medium is exposed does not show this interference effect. The ratio of interfered to total energy absorbed per unit area of detector falls linearly from one to zero as a point on the recording medium varies from $x = 0$ to $x = \pm l/2$.

In order to address the associated requirement for the total spatial resolution of the detector, note that there are as many standing wave nodes, N, across the half length of the diamond, $\frac{l}{2}$, as there are wavelengths in the pulse length of the component beams. That is,

$$N = (\Delta t)(c/\lambda)$$

where Δt is the pulse length, c the speed of light, and λ is the wavelength of light in the coherent pulse. For example, if the pulse length and frequency of the light were 100×10^{-12} sec and 600×10^{12} Hz, respectively, then $N = (100)(600) = 60 \times 10^3$.

Therefore, to capture a standing wave of such extent requires detector arrays

with linear dimension of 240,000 elements or more, depending on the acceptable level for the fractional uncertainty introduced by the discretization. Photographic images recorded on film emulsions corresponding to array sizes on the order of 200,000 by 200,000 are quite large but not uncommon. Holographic plates typically involve array sizes in this range, as one example. So, in the event that images of such resolution were required, recording times would be limited to 10^{-10} seconds or longer.

In any case, if one sets the discretization interval for acceptable image resolution such that the maximum necessary array dimension is $\sim 4N$, then the number of elements across the array, (#), is:

$$\# \sim 4\Delta t(c/\lambda) \tag{2}$$

$$\text{or} \quad (\#)/(\Delta t) \sim 4(c/\lambda).$$

3.2.2 Consideration 2: Velocity induced motion blur

For certain experiments it may be necessary to make interferometric measurements on an object which is moving with a large rigid body velocity or a large angular rotation rate. This need may arise, e.g., because the condition of interest in the test object has been generated by impacting it with a projectile moving at high velocity (5–10 km/sec). Under such conditions and depending on the nature of the interferometer, the absolute phase of the black and white fringes of the interferogram may be changing at a rate which is proportional to the velocity of the test object. That is, the overall shape of the fringe contours may not be changing, but their "blackness" or "whiteness" (whether they constitute constructive or destructive interference) may be changing so rapidly that only very short duration laser pulses can be used. Otherwise, the recording medium stores a time average of this phase shifting fringe pattern which constitutes an all gray image.

Consider an example where a test object is being viewed in a Twyman-Green interferometer with a stationary reference mirror and a test object which is moving

at approximately 10 km/sec (as a point of reference, the gas guns used in experimental geophysics at Caltech fire projectiles at speeds $\sim$ 7 km/sec). In a Twyman-Green interferometer, the length scale difference between a contour of constructive interference (white fringes) and a contour of destructive interference (black fringes) is $\lambda/4$, where λ is the wavelength of light used. If one hoped to keep the fringes stationary on the recording plane to within $\frac{1}{10}$ (or less) of this contour difference of $\lambda/4$, then the relative motion of the object over the duration of the laser pulse must be no more than $\lambda/40$. If the velocity difference between the object and the reference is v, then the allowed pulse time, Δt, is

$$\Delta t < (\lambda/40)/v \qquad \text{or} \qquad 4\Delta t(c/\lambda) < c/(10v) \tag{3}$$

where, again c is the speed of light.

Taking $v = v_0 = 10 \times 10^3$ m/sec and $\lambda = \lambda_0 = 0.5 \times 10^{-6}$ m yields

$$\Delta t|_{v_0,\lambda_0} < 1.25 \times 10^{-12} \text{ sec}$$

and

$$4\Delta t(c/\lambda)|_{v_0} < c/(10v)|_{v_0} = 3000.$$

By comparison with equation (2), it is apparent that the amount of detail which can be captured in an interferogram of a moving object can be inversely proportional to the velocity of the object. Of course, this restriction does not apply to all types of interferometry (transmission or Mach-Zehnder interferometry [16], for example). Also, the time between frames for the data recording device does not usually have to be as short as the "stop-action" laser pulse which is being used.

3.2.3 Consideration 3: Required experimental time resolution

The time resolution for the study of dynamic phenomena can often be established in a way analogous to the procedure used to find the time resolution for a high speed camera, namely, in terms of the ratio of the dynamic phenomenon's wave speed to its limiting (minimum) length scale. For example, studies of isotropic linear elastic media involve two associated wave speeds. However, for many isotropic

materials, these two wave speeds, the bulk and shear wave speeds, are only different by about a factor of two. Approximately speaking then, the relevant wave speed may be taken as the bar wave speed $[E/\rho]^{1/2}$. Referring to Section 3.1, recall that the bar wave speed varies from 5000 m/sec for the common structural materials to 17,000 m/sec for diamond. There are also, of course, a wide range of materials with lower wave speeds (such as polymers).

One possible limiting size scale is the optical spot size of the high speed camera at the object plane where the solid is being observed. For a diffraction limited optical system, the spot size (half width) is $1.22\lambda f$ [16], where, again, λ is the wavelength of light used and f is the "f number" of the objective lens ($f \sim 1$ would be considered typical for 20x–40x microscope objectives). For convenience, one may reasonably approximate this limiting size scale as 1.2λ. Finally, then, one possible value for the minimum time event associated with the time required for an elastic wave front to traverse the minimum optically resolvable spot would be $1.2\lambda/\sqrt{E/\rho}$. Note the similarity, obviously, between this limit and the limit calculated previously for the minimum switching time of a rotating mirror camera. Upon using values for λ of 0.25μm (diamond's ultraviolet transmission limit is $\sim 0.225\mu$m) and for $\sqrt{E/\rho}$ of 17,000 m/sec yields an estimate for the minimum time event of 18×10^{-12} sec.

From the foregoing discussion it is possible to see that restricting one's interest to high speed camera designs in the 10-100 billion frame/sec regime, or perhaps lower, may not be much of a restriction for studies in solid mechanics, or in general, where optical interferometric techniques are used. The techniques themselves are similarly limited.

3.3 Technical Issues

Having discussed some high speed camera performance characteristics in the context of framing rates $\sim 10 \times 10^9$ frames/sec, one is prompted to ask, "Do present high speed cameras collect data at 10 billion frames per second?" The short answer is, "No." The reasons distill down to engineering development costs. As already mentioned, several camera types have streak writing rates in the 100-10 billion

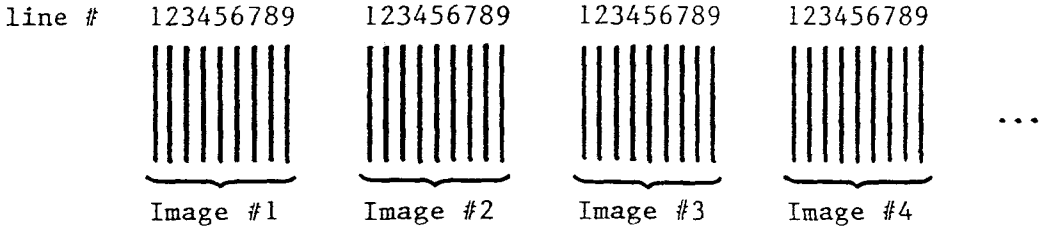
lines/sec range; the word "line" is being used here to indicate a one dimensional array of pixels. However, it has been customary in western industrialized countries to treat the two dimensional arrays or images being manipulated or recorded by high speed cameras as somewhat inviolable rectangular arrays of pixels which must be swept as a contiguous unit. Hence, a maximum sampling rate of 10×10^9 events/second ends up being divided by the number of lines/frame, so that a camera with 2000 lines/frame has an associated recording rate of 5 million frames/second.

In effect, one confronts here the question of how one creates a high speed camera which is an economic parallel processing device, so that images with larger numbers of pixels do not force longer time intervals between captured images. One thus asks how does one design a camera which captures successive two dimensional image arrays, but records them at a rate as if it were recording as rapidly as multiple streak cameras, with one streak camera for each line of the two dimensional array.

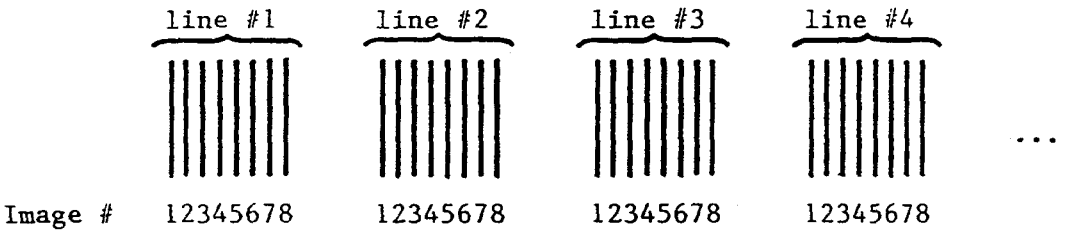
3.3.1 Raster Cameras

Soviet designers were the first to put significant engineering efforts into camera designs which might achieve framing rates in the billions/second with images containing arbitrarily large numbers of pixels/frame. These devices are usually referred to as raster cameras. A fair discussion of raster cameras (along with a number of other types) from the heyday of high speed camera design (1950's and 60's) can be found in Dubovik [22]. This period of great interest corresponds with the development period for nuclear weapons and their associated detonics.

Raster camera designs attempt, in one fashion or the other, to separate the many optical lines of data in each frame from one another spatially prior to recording. Consider first the schematic of the standard recording pattern of images on film:



By comparison, for a raster camera a schematic of the recorded image pattern on film is (usually) of the form:



As a rule, raster cameras did not really catch on, primarily because the Soviet designers were technologically limited in the 1960's to fairly cumbersome means for separating the lines of the optical image. As an example, one of the techniques used to physically separate the lines of data in each optical image utilizes a great many cylindrical lenses placed as an array of field lenses in an image plane.

3.3.2 Recent Technological Developments

Five important technical developments have occurred since the 1960's which now make the basic idea of raster cameras eminently more attractive. These developments will now be discussed:

3.3.2.1 Fiberoptic cables have undergone great improvements both in performance and in cost. This improvement now makes the separating of an optical array into a great many individual lines almost trivially easy by comparison with what one would have to have done in the 1960's. With this newer technology, an optical image is merely focussed onto the polished end of a fiberoptic bundle, which consists of a stack of fiberoptic ribbons (a ribbon, in this case, indicating a bonded coherent layer of individual optic fibers). These ribbons can then be separated, one from another, and snaked about to create configurations of input and output orientation with great design latitude.

3.3.2.2 Higher powered illumination sources are required to match correspondingly higher camera data recording rates. Since the most useful type for the

inquisitor illumination is often monochromatic coherent radiation, it is the rapid advances in laser technology which make higher data recording rates reasonable.

To understand the correspondence between data rates and illumination power requirements, one introduces the concept of quantum data efficiency. This characteristic of a recording medium is defined as the required number of photons to yield a quantum of data. For example, in ordinary use, modern photographic film requires about 100 visible frequency photons to generate the minimum recordable event—in this case, an exposed grain of silver halide which will develop into a globule of silver. This quantum data efficiency of approximately 0.01 for film is dependent on a number of conditions (exposure duration, exposure intensity, grain size, spectral distribution, etc.). By comparison, modern CCD detectors have quantum data efficiencies (QDE's) as high as the range from 0.25 to 0.80 out of a possible maximum of 1.0. A detailed discussion of the various influences which contribute to the net QDE for film and other media may be found in Dainty & Shaw [23].

Another characteristic of a detector type is its saturation level. For most types and certainly for the two most important, film and CCD's, this saturation level is typically taken as 256 or more times the minimum, or quantum detector response. Thus it is common to evaluate a recorded image in terms of 256 levels of gray as averaged over an associated area generally referred to as a pixel. The size of a pixel divided into the total area of the detector specifies the total resolution or array size for the detector.

The way in which the saturation level is tied to the resolution may be seen by recourse to a simple example. Consider a large photographic plate where one considers the entire film area as a single pixel in order to attain the highest possible saturation level. The minimum detectable exposure level is achieved when, after first development, a single exposed silver grain remains. The maximum detectable level (the saturation level) is reached when all of the original grains remain after development. Hence, if one were counting silver globules on the developed plate, the saturation level would be the total number of originally available silver halide

grains. However, if one does not need such a large gray scale, one can easily conceive of subdividing the surface area of the film plate into an ever increasing number of subareas (pixels) of proportionally smaller and smaller area per pixel. Obviously, in this process, the gray scale for each pixel will be proportional to the area of the pixel, smaller areas containing fewer silver halide grains.

To decide how many photons are required for a "picture," then, has much to do with the type of optical information that one attempts to record. The present commercial U.S. television standard corresponds to an array of 640×480 pixels with a gray scale of 256 in three primary colors. A fairly high quality 35 mm film image would have a similar gray scale with an associated array size of $10,000 \times 15,000$ (a 24 mm $\times$ 35 mm area with 400 pixels/mm, i.e., a 2.5μm $\times$ 2.5μm area per pixel). A large glass film plate used in the most advanced optical imaging systems would correspond to an array size on the order of $200,000 \times 200,000$,† with a gray scale of 256 or more.

To make the requirements for illumination power quantitative, let us consider three cases. But first, recall that

$$\text{energy per photon } = \hbar \frac{c}{\lambda}$$

where Planck's constant, $\hbar = 6.63 \times 10^{-34}$ joule/sec, the speed of light, $c = 3.00 \times 10^{8}$ m/sec, and where λ is the wavelength of light. For the visible spectrum ($\lambda \approx$ 420–700 nm) the energy per photon ranges between

$$4.7 \times 10^{-19} \text{ joule } > [\text{energy per photon}] > 2.8 \times 10^{-19} \text{ joule}$$

$$(\text{at } \lambda = 500\,\text{nm}) \text{energy per photon } = 4.0 \times 10^{-19} \text{ joule}$$

† The wide field Schmidt telescope at the Palomar Observatory is an $f/2.5$ device [16]. For a diffraction limited spot size (half width) of $\sim 1.22\lambda f$ and using $\lambda = 0.56\mu$m, this yields a visible wavelength spot size of $\sim 1.7\mu$m. This same telescope uses 35 cm square glass photographic plates as a recording medium. Thus the number of pixels across the 35 cm plate width is $350\text{mm}/(1.7 \times 10^{-3}\text{mm}) = 206,000$. The collector plate is equivalent, then, to an approximately $200,000 \times 200,000$ array.

Case 1: a) 640×480 array

b) gray scale of 256 with an average of 128

c) quantum data efficiency (QDE) of 0.25

d) energy per photon ($\hbar c/\lambda$) of 4.0×10^{-19} joule

e) framing rate of 30/sec

$\Rightarrow$ energy/frame $= 640 \times 480 \times 128(\frac{1}{0.25}) \times (4.0 \times 10^{-19}$ joule$) = 6.3 \times 10^{-11}$ joule

$\Rightarrow$ power requirement $= (6.3 \times 10^{-11}$ joule$) \times (30/\text{sec}) = 1.9$ nanowatts

Case 2: a) $10,000 \times 15,000$ array

b) gray scale of 256 with an average of 128

c) QDE of 0.01

d) $\hbar c/\lambda$ of 4.0×10^{-19} joule

e) framing rate of 10^6/sec

$\Rightarrow$ energy/frame $= 10,000 \times 15,000 \times 128 \times (\frac{1}{0.01}) \times (4.0 \times 10^{-19}$ joule$)$

$= 7.7 \times 10^{-7}$ joule

$\Rightarrow$ power requirement $= (7.7 \times 10^{-7}$ joule$) \times (10^6/\text{sec}) = 0.77$ watts

Case 3: a) $200,000 \times 200,000$ array

b) gray scale of 1024 with an average of 512

c) QDE of 0.01

d) $\hbar c/\lambda$ of 5.0×10^{-19} joule

e) framing rate of 4×10^9/sec

$\Rightarrow$ energy/frame $= 200,000 \times 200,000 \times 512 \times (\frac{1}{0.01}) \times (5.0 \times 10^{-19}$ joule$)$

$= 1.0 \times 10^{-3}$ joule

$\Rightarrow$ power requirement $= (1.0 \times 10^{-3}$ joule$) \times (4 \times 10^9/\text{sec}) = 4$ megawatts

These are the associated energy and power requirements for the recording

medium. The energy losses associated with the experimental technique itself can sometimes be quite large, as for example with holography, and these losses can boost the power requirement at the light source by orders of magnitude. However, there are a fair number of quite useful experimental techniques (Twyman-Green interferometry and moiré interferometry as examples) which have light efficiencies in the range of $\frac{1}{2}$ to $\frac{1}{10}$. If one assumes, for illustrative purposes, an efficiency of $\frac{1}{10}$, then the light source output power requirements for Case 2 would be approximately 8 watts. That value is exactly the present performance level for a large commercially available argon laser (20 watts all-line) operating in single line mode at its most efficient wavelength (514.5 nm).

If one considers Case 3 with an experimental light efficiency of $\frac{1}{10}$, then one arrives at the rather severe requirement that the light output for the laser be 40 megawatts. These are just the power levels being striven for and obtained with the development of free-electron lasers [24]. Free-electron lasers (FEL's) pose, as yet, a considerable expense. Stanford has established a single FEL user facility which emits trains of pulses several microseconds in duration at peak powers as high as two megawatts. However, because of their possible use in the Strategic Defense Initiative, and for plasma heating in controlled thermonuclear fusion, there is considerable interest in FEL's in the national laboratories. As a result, performance improvements in these devices are being developed fairly rapidly.

3.3.2.3 Extensive digital image processing has become considerably more reasonable in direct proportion to the rapid drop in the cost of computer hardware. Images are of markedly less value without the ability to extract quantitative information. It follows that, while not directly a part of high speed cameras, computer image processors make it a worthwhile pursuit to collect orders of magnitude more information.

Because human vision is quite sophisticated, it is easy to forget how much information can be contained in a single photograph. Consider a group of 100 photographs, each with an array size of $50,000 \times 50,000$ (approximately equal to

a 9 cm × 12 cm negative) and a gray scale of 256 or 1 byte. This amounts to 2.5 gigabytes per picture for a total of 250 gigabytes. By comparison, present day audio compact discs (125 mm) possess a data storage capacity of only 500 megabytes each.

Similarly, the current generation of mass produced flatbed document or image scanners have 8 bit gray scales with associated resolutions of 16 pixels/mm over a scan bed area of 212.5 mm × 350 mm; that capability corresponds to an array size of 3400×5600 at 8 bit gray scale, which is only $\frac{1}{130}$ of the area of a $50,000 \times 50,000$ array. However, some of the newer flatbed document scanners entering the market are at 32 pixels/mm over a similar area, corresponding to a 6800 × 9600 array. Document type scanners can be very reasonably priced because of the large market, but the demand for larger arrays falls significantly due to the obvious limitations of human vision. The diffraction limited performance of the eye is in the range of 1 minute of arc. With a 90° subtended angle, this would correspond to an array dimension of $\sim$ 5400 pixels. Except for murals, family portraits, etc., most images, of whatever size, generated for human viewing tend to be roughly $10,000 \times 10,000$ elements or less. At worst, maximum economic efficiency would require that larger images be segmented prior to digital capture, whether by flatbed scanner or CCD.

3.3.2.4 Quality control in lens manufacture has advanced to the point where one may order a 1000 unit production run of lenses and the variation between individual lenses will be sufficiently small so that from a design standpoint they can often be treated as being identical. The unit price for small lenses is not yet as low as, say, nuts and bolts, but increasing levels of automation continue to exert downward pressure. The technical achievement of inexpensive, functionally identical subcomponents brings major economic benefits in the manufacture of any device requiring many of these subcomponents. In spite of all the advances in the fabrication techniques (numerically controlled diamond machining, grinding, polishing, etc.) used for optical components, the process of lens manufacture is still exacting. At this time, it is usual for lens manufacturers to specify nominal lens focal length tolerances to no better than ±1%.

However, $\pm 1\%$ variation in lens focal length is too large to classify the lenses as identical for certain camera design purposes. On the other hand, much of the variation comes from batch to batch variations caused by production variables, like slow wearing of the master polishing tools, which lead to changes in the finished radii of curvature. One finds, though, that the variation in the radius of curvature, focal length, etc., between lenses of the same batch is probably no worse than about $\pm 0.1\%$. So, efficient design practice still requires careful coordination with the lens manufacturer.

As a case in point, a group of 144 symmetric convex, BK-7 glass lenses purchased (for improvement to one of the GALCIT high speed turbine cameras) from Melles-Griot in 1988 had the following nominal specifications:

lens diameter = 25.0 mm

focal length (f) $= 125.0\,\text{mm} \pm 1\%$

center thickness (t_c) $= 3.2\,\text{mm} \pm 0.1\,\text{mm}$

design index (at $\lambda_c = 546.1\text{nm}$) of refraction ($n_c$) $= 1.5187 \pm 0.001$

Measurements on the actual lenses using a precision interferometer provided a statistical description of the 144 lenses as:

Average surface radius of curvature (r)	$= 130.35\,\text{mm}$
Average lens center thickness (t_c)	$= 3.24\,\text{mm}$
Average focal length (f)	$= 125.75\,\text{mm}$
Standard deviation for r	$= 0.11\,\text{mm}$
Standard deviation for t_c	$= 0.04\,\text{mm}$
Standard deviation for f	$= 0.073\,\text{mm}$
Fractional standard deviation in r	$= 0.0008$
Fractional standard deviation in f	$= 0.00058$

One may note that the average (batch) focal length of 125.75 mm is indeed 0.6% larger than the manufacturer's nominal focal length. However, twice the fractional standard deviation in the batch focal length is only 0.1%.

Not all high speed camera designs require the use of a large number of identical

lenses, but a number of important ones do. Further, one of the distinguishing characteristics of those designs that utilize multiple lenses is that generally some aspect of their performance will vary linearly with the number of lenses used, so that an order of magnitude improvement in performance will require an order of magnitude more lenses of equivalent quality.

3.3.2.5 Optical Equipment and Design have seen many advancements. The optical designer's "bag of tricks" has been enlarged considerably since the 1960's by very substantial developments in the technology of optics. A high speed camera consists of more than that part which does the optical switching, and the usefulness of the overall design can be significantly degraded if the non-switching portions of the design impose additional restrictions on performance. We have already discussed the obvious necessity of a coherent light source powerful enough to supply the requisite photon rates for interferometric studies in the billions of frames per second. There are a great many other pertinent technical difficulties for which relevant technology has been developed in the last three decades. The field of modern optics is indeed immense, so it is only practiceable to touch on a few examples to illustrate the way in which the basic switching portion of a high speed camera interacts with the other elements of the overall camera. Examples:

3.3.2.5.A Frame Array Aspect Ratio. Depending upon the subject matter, the optimal height to width ratio (aspect ratio) for the photographic array may vary rather widely. The standard 35 mm film format is 3 to 2 (36 mm $\times$ 24 mm). Fractography studies and certain other subject matter are best suited, however, by more extreme aspect ratios. In dynamic crack propagation studies it is quite common for the researcher's interest to be limited to observations within a fairly narrow strip of material surrounding the crack propagation path.

It would clearly be beneficial if changing the aspect ratio of the photographic image (while keeping the total data per frame constant) could be readily achieved without major modifications to the overall high speed camera design. A number of possibilities suggest themselves. One technique would be to use the presently avail-

able square rods of glass made up of millions of silica fiberoptic filaments and their claddings fused together as a single conglomerate solid. These coherent fiberoptic image guides can be bent and formed using a heat source, such as a Bunsen burner, in a manner very much like that used with homogeneous glass rods. Consider, for illustration, how one would reconfigure an image with a 9 to 1 aspect ratio to an image with a 1 to 1 aspect ratio.

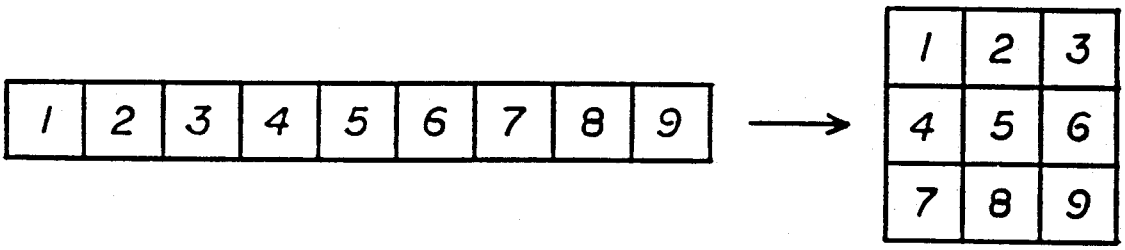


Figure (3.2)

The image with the 9 to 1 aspect ratio is focussed onto a grouping of 9 square fiberoptic image guides arranged in a single row. The 9 rods are then bent in groups of 3 into a folded arrangement which makes a 3×3 exit array of 1 to 1 aspect ratio (as illustrated in figure (3.2)).

This means of altering an image aspect ratio is limited to those electromagnetic frequencies where transmitting fiberoptic filaments are available. The same general idea of folding a long narrow array into a square array can also be accomplished using mirrors and lenses (or imaging reflectors). In any event, it is now more readily possible to design modular units which tailor the image aspect ratio to that for which the high speed camera was designed. These modular units would then be placed at an intermediate image plane (or field) between the collection optics of the high speed camera and the high speed recording device proper.

It should be pointed out that there are certain disadvantages to folding an image array using fiberoptics, as opposed to segmented mirrors. A fiberoptic bundle captures a discretized representation of an image. In order to avoid losing infor-

mation from the image, the discretization interval associated with each fiber (its diameter, roughly) must be at least several times smaller than the optical spot size for the image. The appropriate ratio of fiber size to optical spot size for the image can be arranged easily enough, so that no information is lost in the discretization of the image. However, an associated and unfortunate effect occurs when the captured image light exits the fiberoptic bundle. It was Abbe [16] who first set forth the important relationship between optical spot size and numerical aperture; the spot size is inversely proportional to the sine of the half-angle of the cone of light making the optical spot. Hence, the angular convergence of the light entering the fiber bundle is that angle associated with the optical spot size of the image. However the angular expansion of the light exiting the bundle is the greater angle associated with the smaller spot size (diameter) of the fibers. The camera is limited by its resolving capacity to the total spatial image array associated with the image incoming to the fiber bundle. In effect, the camera has insufficient angular aperture to capture all of the light exiting the fiber bundle at the greater expansion angle of the light leaving the bundle. Roughly, to avoid degrading the image quality, some light intensity is lost. Where available light is at a premium, it will be preferable to segment and fold the image using the somewhat more cumbersome techniques possible with segmented mirrors.

3.3.2.5.B Panning. Human vision makes extensive use of panning motions because the high resolution portion of the retina, the *fovea centralis*, is quite small ($\sim$ 0.3 mm in diameter) and consequently encompasses a rather small angular field of view ($\sim$ 0.8°). This technique is so effective that it was adopted more or less without modification for motion pictures. Unfortunately for high speed photography, it is just not physically reasonable to mechanically move the entire high speed camera (either linearly or rotationally) at the speeds which are generally necessary to effectively track phenomena moving at high velocity and in close proximity to the camera. This does not mean that the idea of panning—relieved of the restriction of accomplishing it by physically moving the camera—cannot be used in high speed

photography.

One means of accomplishing the equivalent of a panning operation suitable for high speed photography may be illustrated by consideration of the Cranz-Schardin type high speed camera. The Cranz-Schardin design consists of multiple illumination sources and multiple still cameras arranged in such a way that each still camera can receive light emanating from only one of the illumination sources (which are pulsed in sequence). One could consider combining the light received by the collecting (imaging) optics of the individual still cameras using a group of beamsplitters such that at one output it would be possible to view the superposed images from each of the lenses. This combined output could then be fed into a single input high speed switching type camera. If all of the images but one were dark at any one time, then, effectively, no double exposures would occur, and a number of time-sequenced images could be obtained by the high speed switching camera through each of the still camera lenses. The still camera lenses and illumination sources could be arranged in a wide variety of configurations, one of which would simulate panning through discrete, overlapping (or adjacent) viewing area.

While this idea is simple in concept, it becomes involved when considered in conjunction with the use of some experimental techniques (interferometry, for example). However, the alternative to panning—building a high speed camera which captures a very large array—is a choice which gets proportionately more expensive with the increasing number of array elements (pixels).

For economic reasons, one should like the optical and electronic hardware associated with the panning device to be adaptable and modular, so that it might be reconfigured for different applications, rather than purchasing completely new items for each application. For this reason, it is suitable to divide the general panning activity into a number of separate functions performed by individual groups of components, which are then assembled into a variety of complete devices for performing the entire panning function under different conditions.

3.3.2.5.B.i *A device for switching a single input laser beam between multiple exit paths.* Having a setup with separate, pulsed light sources follows the technique used in the original Cranz-Schardin camera. This approach is sometimes unreasonable and sometimes too expensive, depending on the type of light required. For some types of laser it is most reasonable to use a switching device on the output of a single large laser rather than attempting to get narrow, high peak power pulses out of a number of smaller lasers. In any event, in cases where switching is preferable, one might expect to fabricate a fairly versatile switching network using electro-optic materials (as variable polarizers) as the active switching elements in front of polarizing beamsplitter cubes.* Today polarizing beamsplitters are generally made by enhancing the Brewster's angle effect using dielectric coatings on the 45 degree surface of a glass beamsplitter, and the extinction ratio continues to improve, so that these devices are quite effective as discriminators between light of orthogonal polarizations. Electro-optic materials used as variable polarizers are commonly the mechanism for Q switching in lasers. As such, there is a well developed technology for manufacturing the electro-optic elements and for fabricating high voltage electronics which will rapidly alter the electric field applied to the electro-optic material. Figure (3.3) is a schematic example of a (single input) 8 output device using seven electro-optic elements and seven polarizing beamsplitters.

The physical size of the device can be tailored over a wide range, so the intensity of the laser light being switched can always be kept below the damage threshold of the materials and surface coatings. Obviously, as the laser beam is made smaller in size, the proportionally smaller electro-optic device can be made to alter its state more rapidly, which is always preferable. However, for very high power applications,

---

* The electro-optic material is placed in front of a polarizing beamsplitter cube and by appropriate modulation of the applied voltage to the electro-optic element the polarity of light entering the element can be switched between two orthogonal polarizations at exit from the element. The cube then sends the light in either of two directions depending on the polarity of the light (on the applied voltage).

it will be difficult to build an electro-optic element which can completely alter state in the 10^{-10} second range. As it turns out, one can design the overall panning device so it will be sufficient that the switching of any single electro-optic element can be performed at a rate which is substantially (an order of magnitude or more) less than the framing or switching rate for the high speed camera.

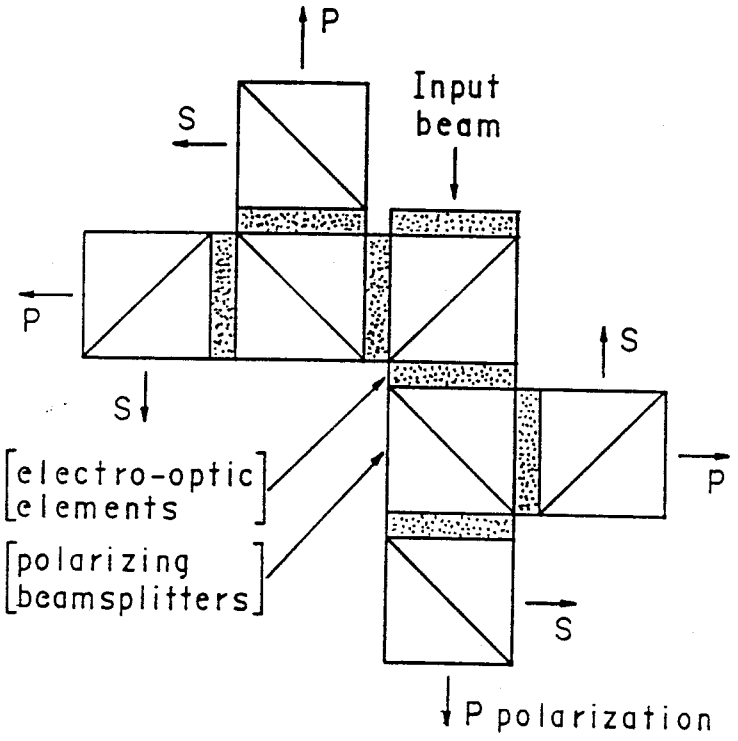


Figure (3.3)

There are obviously other means for making a switching device for the output of a single laser. Most of the physical effects which can be used to perform the high speed switching in the high speed camera itself can also be used with the illumination light from the laser. One of the reasons for choosing electro-optic materials in this latter capacity rests in the flexibility which can be built into the electronic driver circuits for the electro-optic elements. Panning is more valuable when it can be carried out with some flexibility. One can easily envision how difficult it would be to use the constant rate panning appropriate to phenomena moving at constant velocity for those cases where acceleration of the photographic object is involved.

The stop-start motion of a crack interacting with stress waves is an example of a rapid change in the velocity of the photographic object.

3.3.2.5.B.ii *Distribution and collimation system.* It certainly would be easiest if the different exit beams from the previously discussed switching device could be focussed into individual, single-mode, polarization preserving, fiberoptic cables. This can be achieved only for low power operation at present, but there are a great many manufactured items which make beam steering with mirrors almost as flexible, and beam steering with mirrors can be accomplished at very high power levels merely by insuring that the beam is sufficiently expanded to keep the intensity below the prescribed limits. There is some additional difficulty associated with using mirrors for studies involving short time scales. In 10^{-10} seconds, light only travels 3 cm. In order to maintain accuracy in the time scale of events, appropriate lengths must be maintained between mirrors, and this becomes more constraining as the optical arrangement becomes more elaborate. It is perhaps worth noting that for ultrashort time events, it is unnecessary to use an electro-optic switching assembly in the panning activity. A single, narrow, laser pulse can be split using a cascade of ordinary beamsplitters. One can achieve the appropriate timeshift between illumination at each of the collimators by varying the subsequent path length to each of the individual collimators. Choices in technique for beam expansion and collimation are largely economic—the smaller the allowable tolerances on the optical components, the smaller the deviation from the ideal collimated beam, and the higher the cost of the optics.

3.3.2.5.B.iii *Elimination of multiple exposures (isolation of the light of each of the individual illuminators from all but its respective collection optics).* This function is the key to the performance of a Cranz-Schardin type camera. Unfortunately, the specific way in which this particular function is carried out in the standard Cranz-Schardin camera is not entirely suitable for many optical techniques, as the angle of illumination changes markedly from one of the individual collimated beams to the next. It is often preferable, if not necessary, that the angle of illumination

remain constant with only a translation of the area of illumination between subsequent collimated beams.

An alternative arrangement which will allow the more preferable form of illumination may be illustrated by reference to a schematic in Figure (3.4). The way out of the difficulty is to record two fields of view simultaneously (rather than one as with the usual Cranz-Schardin camera) with the two fields of view being adjacent to one another or slightly overlapping, as necessary. If one numbers the fields of view in sequence, then the high speed camera will always be recording an adjacent odd numbered field and even numbered field simultaneously. There is now a wide separation between all of the odd (or even) numbered fields, and the collection optics for each of the individual odd (or even) fields will have no viewing area in common.

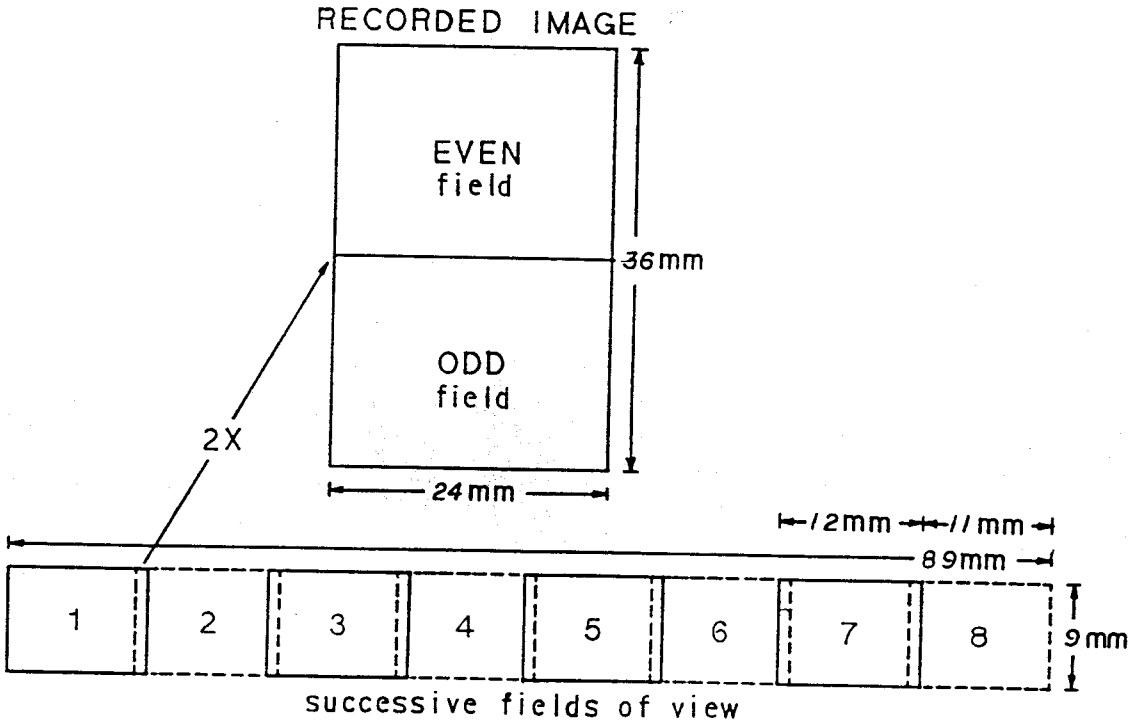


Figure (3.4)

In order that the collection optics have physical apertures which are as large or larger than the dimension of their record fields (cf. figure (3.5)) a long beamsplitter rod (merely a lengthened version of the standard beamsplitter cube retaining the

splitting surface at 45° to the optical faces) is used. All of the odd numbered collection optics view the experiment through one of the extended exit faces of the beamsplitter. All of the even numbered collection optics view the experiment through the alternate exit face at 90° to the odd numbered optics.

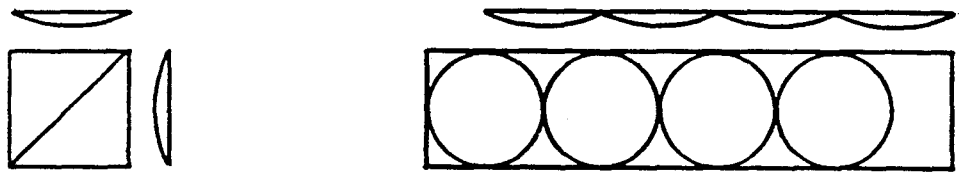


Figure (3.5)

Similarly, the illumination optics are also placed on alternate sides of a beam-splitter rod. There is now only the problem of illumination light "leaking" from the even fields to the odd fields (or vice versa) from the regions where the edges overlap slightly.

This leakage can be eliminated entirely, but the way that it is best eliminated will depend somewhat on the nature of the overall experimental technique into which the panning technique is being fitted. Because of diffraction effects, it is poor practice to have the edges of the illumination beam visible within the area of view where the experiment is being observed. One can illuminate both an odd and an even viewing field simultaneously with the same collimated beam while keeping the illuminating beam just small enough to avoid overlapping into the next adjacent fields. In this arrangement, one edge each of the odd and even viewing fields will be dark. These darkened edges will be the extreme edges (the two furthest from the viewing area of interest) so this solves both optical problems of concern (light "leakage" between viewing fields and the appearance of illuminating beam edges in the area of interest), but it introduces a constraint on the electro-optic switching network which is best avoided. As presently described, the optical setup requires that the light be switched from one collimated illuminator to the next in the time

between recorded frames.

To allow for a longer switching time for the electro-optic driver circuits, there are tricks which can be used where the experiment allows them. For example, in many cases, the experiment preserves (or almost completely preserves) the light polarity between the incoming and the exiting light. In such circumstances the beamsplitter rods discussed can quite usefully be polarizing beamsplitters. The illuminating beams are then made only slightly larger than a single odd (or even) viewing field and the odd fields are illuminated with one polarity of light while the even fields are illuminated with the opposite polarity of light. Both an odd and an even field are illuminated at the same time, then, but using two separate beams of opposite polarity. The two groups of beams of opposite polarity are switched alternately. In this way, even if the switching time for the electro-optic network is significantly longer than the time between frames of the high speed camera there will always be at least one field (either odd or even) in which viewing will be undisturbed. The importance of such a slower switching time for the light source was mentioned in subsection **3.3.2.5.B.i**. The other advantage to using polarizing beamsplitters in this capacity is that it eliminates the light waste which would occur when using ordinary beamsplitters.

3.3.2.5.B.iv *Superposition of the view from the individual collection (imaging) lenses.* It is selfdefeating to superpose the view from too many of the individual collection (imaging) optics. As one may recall, this superposition was proposed so that all of the captured images might be recorded by one high speed camera. If it becomes necessary to extend this discrete panning operation over a great many successive, adjacent fields of view (in the illustrative schematic there are only 8 adjacent fields of view) then it becomes more reasonable to use more than one high speed camera, each optically coupled to only 8 or 16 adjacent fields of view. The reasons for following such a guideline become clear when considering how the superposition of views might be accomplished.

Just as the illumination light can be switched between multiple exit paths

using a binary cascade of beamsplitters, so the successive views of the collection optics can be superposed through an inverted binary cascade of beamsplitters. The difference is that the illumination light beam is of the simplest possible form, that is, it is in Gaussian 00 mode, whereas the collected light will contain all of the higher frequency spatial modes which constitute its information content concerning the imaged object, which has caused these higher order modes to come into existence. The substance of diffraction theory is that the higher the order of the mode the greater the beam waist dimension for a given angular spread to the mode. For this reason, it is not reasonable to use electro-optic materials as switching elements in the collection cascade as a means of avoiding light loss. This electro-optic element would have all of the design requirements associated with a high speed camera itself, but one would be operating the camera switching process in reverse in this case.

3.3.2.5.C Insets; Split Screens (taking two pictures of different magnifications simultaneously). It is standard practice to represent certain forms of data using insets. For example, in mapmaking the space between roads in cities is much smaller than in rural areas, so it is usual to make road maps with one magnification to represent the interstate highway system and another magnification to represent localized city primaries. This way of representing the data saves one from making the entire map to the magnification of the city scale map. However, when confronted with the collection (as opposed to the representation) of data, most cameras are arranged to take one picture at a time of one magnification at a time. Rather than buying several cameras in order to take several pictures at several magnifications, it is sometimes sufficient to retain only one camera, but use its data collection capacity more efficiently. Certainly, the human eye uses this approach. The effective spacing between detector elements in the periphery of the retina is 6 times greater than in the fovea centralis. Experimental studies in fracture mechanics are a good example of a case where the resolution requirements are different for different parts of the overall field of view. One should like to make the pixel size at the object quite small in the region nearer the crack tip where the gradients in strain are largest.

Further from the crack tip region, the smaller pixel size is unnecessary.

There is an obvious drawback to this approach which requires compensation in those cases where the available amount of illumination light is limited. One may recall that under nighttime conditions the high resolution color vision of the fovea centralis receives too few photons to each of its detector elements. The solution in high speed photography is conceptually no different than for the eye - shine a spot light on the region that is to be resolved at high resolution. As the size of the pixel element on the object decreases, perforce the intensity of light should be increased, but the brighter light need only illuminate the high resolution region.

Since this idea of capturing two images of differing magnifications simultaneously is such a useful and basic idea, it is something equally basic which has limited its potential application until quite recently—geometric lens aberrations. Consider figure (3.6) , which depicts two fields of view, one within the other, to be captured with different magnifications. If one directs one's attention first to an optical arrangement suitable to imaging the smaller field of view of 3 mm $\times$ 4 mm, the appropriate lens choice would seem to be a low power microscope objective of approximately $5X$. However, how then does one fit the imaging optics for the larger field of view of 12 mm $\times$ 16 mm around the physical obstruction of the typical $5X$ microscope objective? There is insufficient working distance associated with the microscope objective to be able to fit an appropriately sized beamsplitting surface between the objective and the object so that the larger field of view may be imaged simultaneously. Alternatively, it is not possible to image the larger field of view through the $5X$ microscope objective (using a beamsplitter after the microscope objective) since it has insufficient field of view. These difficulties become unimportant if one need not reduce the geometric aberrations associated with the design of the microscope objective through the expedient of making the lens physically smaller, which reduces the working distance and field of view. A number of improvements in modern optical technology now make it possible to implement longer working distance microscope objectives. Future improvements in working distance

and field of view are to be expected with the full development of promising new optical technologies.

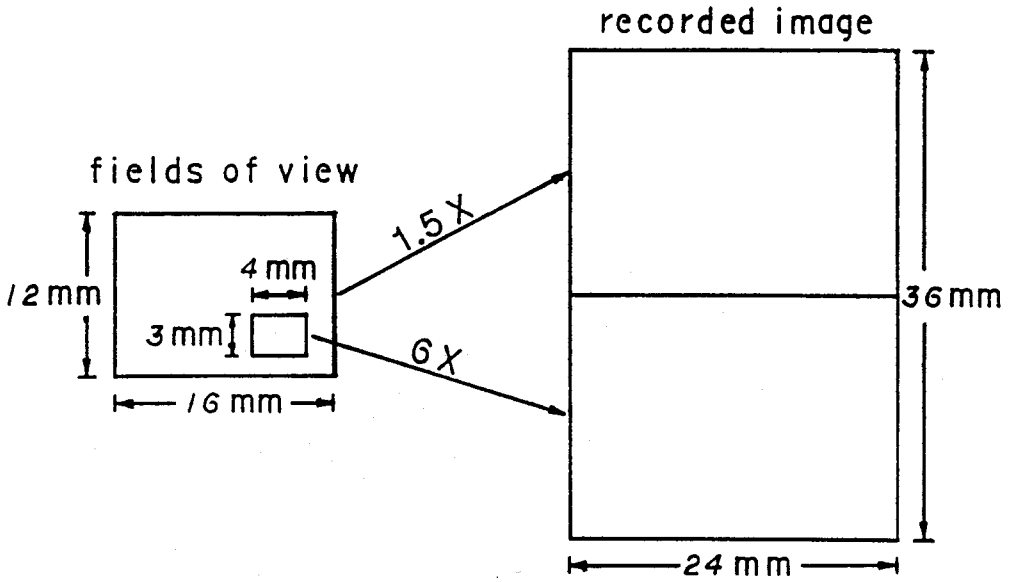


Figure (3.6)

3.3.2.5.D Objective Lens Interchangeability. It is not possible to use one multi–element lens for all possible imaging purposes. Most modern optical devices (microscopes, 35 mm SLR cameras, etc.) are really well thought through optical systems with a multitude of readily interchangeable lenses. To be of significant use, a high speed camera ought to have this same capability, and, to be economic, the interchangeable lenses should not be specific in design to the high speed camera, but rather should be those very lenses optimized so laboriously for microscopes, SLR cameras, enlargers, etc. Such interchangeability requires a standard interface for the connection of the various objective lenses to the high speed camera.

In speaking here of an interface, one is not overly concerned about the mechanical attachment of the lens. Standardization on a particular attachment type (bayonet mount, Royal Microscope Standard thread, etc.) or conversion between types is mostly a matter of simple mechanical adaptors. Of primary concern is the optical interface which will make the objective lens compatible with the recording

device. Familiarity with the modern microscope exemplifies the way in which this may be done. Inside the body of the microscope there will be a transfer lens system. The microscope objective will form an intermediate image within the body of the microscope and the transfer lens will "transfer" this image to a final image plane at the plane of the recording medium. In order to avoid using a different transfer lens system for each objective lens, the overall microscope is designed such that there is a standard "optical interface" between the objective lens and the transfer lens. The intermediate image created by each objective lens has to conform to certain common standards.

Similarly one would hope to design for a high speed camera a transfer lens system which would make it compatible with a wide range of objective lenses. In practice, in order to attain sufficient flexibility, it seems necessary to have some interchangeable components in the transfer lens system as well. This necessity is the case for microscopes and more flexibility is required of a high speed camera than of a microscope. It is not unexpected that one will wish to use both microscope objectives and $\sim$ 1000 mm focal length telephoto lenses on a high speed camera. No one expects to attach 35 mm format camera lenses onto the nose piece of their microscope.

Some of the difficulty of designing a transfer lens system has been reduced by improvements in the various objective lens types and by the availability of specialized components which can be used in the transfer system. Most of the difficulties center around the image deforming aberrations of Petzval field curvature and distortion and the special requirements of specular light imaging systems. Consider figure (3.7), a simplified schematic of a tandem optics system consisting of an objective lens, a field lens placed at a field or intermediate image plane, and a transfer lens.

The field lens is not needed for imaging purposes, per se, but assures that the light rays collected by the objective lens are incident upon the transfer lens and can be imaged through the high speed camera. Without the field lens, the required numerical aperture for the transfer lens system and for the high speed

camera would be much higher than necessary for purposes of optical resolution alone. This excessive demand on the high speed camera (in the absence of the field lens) is particularly true for optical techniques involving specular light, that is, where there is only one light ray emanating from each point on the object, as opposed to diffuse light where light rays emanate in all directions from each point on the object. Twyman-Green interferometry and diffraction moiré interferometry are examples of specular light techniques. The great utility of a field lens in tandem imaging systems may be noted in the design of microscope eyepieces and in the viewing and focussing optics of a 35 mm format SLR camera, both cases where field lenses are fairly universal.

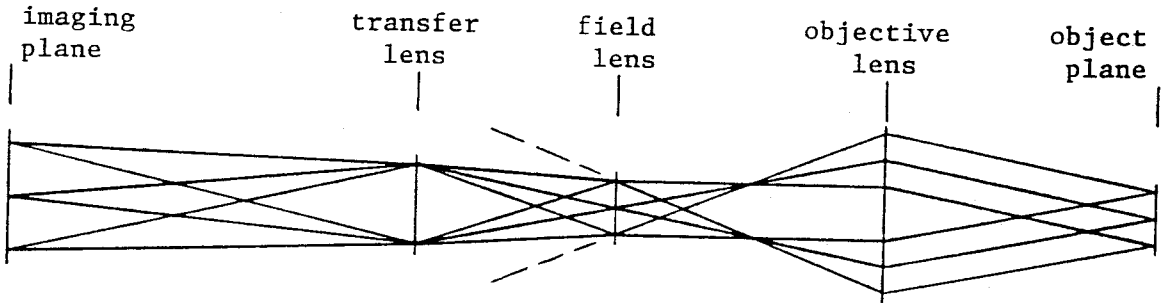


Figure (3.7)

Since a field lens is at or near an image plane, it has very little effect on the image deteriorating aberrations, but it can have a very pronounced effect on field curvature and on image distortion. One can readily appreciate this from the way in which negative lenses are often used in this capacity (placed just before the image plane) as *field flatteners* for the positive primary lens element. One can also readily appreciate how severe is the uncorrected field curvature of a simple positive lens element by observation of the human eye, where the image detector, the retina, has been placed on the curved image surface.

In order that our example optical system should perform optimally, the following conditions should be met:

i) Both the objective lenses and the transfer lens should be individually op-

tically corrected. To the extent that all of the lenses to be used are nearly ideal optical transform devices, one is allowed a considerable freedom in combining different types of lenses in tandem. Microscope objectives were not formerly individually corrected. The aberrations of the objective were compensated for by negating aberrations in the eyepiece. Fortunately, most modern microscope objectives can now be individually corrected due to advances in materials and design techniques. Similarly, it is not always the case that 35 mm format cameras are completely corrected for flatness of field. This has generally been of no great concern because of the typically three dimensional scenes which were being photographed, where much of the frame will be out of focus in any case. Some of the modern camera lenses available are now more or less fully corrected for flatness of field at reasonably large numerical aperture for the specialized purpose of photographing paintings and other planar objects.

ii) The size of the image customarily (when used as designed) made by all of the separate objective lenses should be reasonably similar, or the transfer lens system should have the capacity of variable magnification. For example, the image size generated by most microscope objectives is a 20 mm diameter circle, and the minimum dimension of the 35 mm format is 24 mm. These dimensions are not too different, but they are different. The technology of zoom lens elements continues to make advances, so, at worst, the transfer lens may require a zoom element as part of its construction or at least some aspect of variable geometry which would allow variable magnification.

iii) The field lens should be an interchangeable element available in a broad range of closely spaced focal lengths. This allows one to pick the more or less perfect focal length field lens for the photographic task at hand. Since the field lens is a nonmoving, simple lens, there is no great expense in owning a multitude of them. Certainly, new ones can be purchased as necessary.

iv) The optical aberrations of field curvature and image distortion associated with the field lens must be minimized. One may think of the ideal field lens,

and it then becomes clear how to minimize any deleterious effects which a real lens might have. Given that the objective lens and transfer lens have completely flat image surfaces (no field curvature) and no image distortions themselves, then if the field lens is infinitely thin and is placed exactly at the intermediate image plane between the objective lens and the transfer lens, then this ideal field lens will introduce absolutely no field curvature aberration or image distortion. It is well known how to make a lens as thin as possible (without resorting to a Fresnel lens, which is unsuitable in this case) for a given focal length: the diameter of the lens is made no larger than necessary; the edges of the lens are made no thicker than absolutely required; the index of refraction of the lens is maximized. These first two requirements are satisfied by having sufficient variety in available lens geometries. The third requirement is generally incompatible with lens design because chromatic aberration is so difficult to avoid that lens material is chosen for a maximum Abbé number, rather than a maximum index of refraction. However, a field lens, because of its location at an image plane, has virtually no effect on chromatic aberration. For this reason, one may take advantage of some of the more exotic optical materials which have unusually high indices of refraction, although in most cases that is really not necessary.

The lens system discussed here has been implemented for use on one of the GALCIT high speed turbine cameras and has proved quite versatile in actual use. There is no particular reason why this general sort of design cannot be adapted for use with a number of different types of high speed camera. The ability to use a great many lens elements in series this way has been made possible by the development of multilayer dielectric optical coatings with very low reflectance. For example, the standard coating used by Pentax in its multielement camera lenses has a reflectance of 0.2% per surface through the visible spectrum, as opposed to the 4% per surface of uncoated fused silica. This decrease corresponds to a rather impressive 400 fold improvement $\left(\frac{4\%}{0.2\%} \times \frac{4\%}{0.2\%}\right)$ in signal to noise ratio due to the reduction in forward transmitted ghost reflections for a single element lens. For multielement lenses the

improvement is even more pronounced.

3.4 Potential Design of a Raster Camera

Having explored various aspects of the components which might make up a billion frame/sec high speed camera with potentially very large frame array size, an example design for the primary data capture (or switching) portion of the camera is now considered. In this example design, the switching element itself has been chosen to be a rotating mirror. Such a choice might seem at first to be outmoded, but in fact a rotating mirror offers a clear advantage for certain applications. The principle advantages are two. First, a metallic surface reflects and can therefore be used to manipulate a very wide range of electromagnetic wavelengths. Second, a rotating mirror operates in free-space. This characteristic allows one to use this same switching device to simultaneously switch many (in this case $> 10^8$)inputs. It is this enormous throughput capacity of many free-space devices which makes them attractive, rather than an exceptionally fast fundamental switching time.

A schematic of this example raster camera design using a rotating mirror is found in figure (3.8). Following that, figure (3.9) superimposes an illustrative optical ray tracing over the schematic of the raster camera. In the illustration, the optical components (lens, beamsplitter, etc.) indicated are only those associated with one of the eight sides of the octagonal cylinder which constitutes the rotating mirror. The image is separated into eight parts, each of which is manipulated by one of the eight faces of the rotating mirror. The separation of the image into eight parts can be accomplished with the same device which is used to subsequently separate the image into widely separated individual lines suitable for rastering. Namely, the optical image may be focussed onto the end of a fiberoptic bundle which consists of a stack of fiberoptic ribbons, as discussed earlier.

The design as envisioned would have the following characteristics:

1. 128 frames
2. 1×10^9 frames/sec
3. $200,000 rpm$ mirror turbine speed
4. $4,096 \times 4,096$ pixels/frame
5. 6μm diameter for individual optical fibers
6. 16,384 optical fibers per optical ribbon forming the individual raster lines.
7. 1 to 1 magnification for raster line imaging lens used in double pass mode where light reflected off of rotating mirror passes back through lens
8. 3μm spot size (half width) for diffraction limited performance of raster line imaging lens with $f$number of ~ 4.8.
9. semi-slave or semi-asynchronous camera operation, where the 128 captured frames may be captured at any point within a length of time which is equivalent to $0.8 \times 128 \times (128)$ frames.

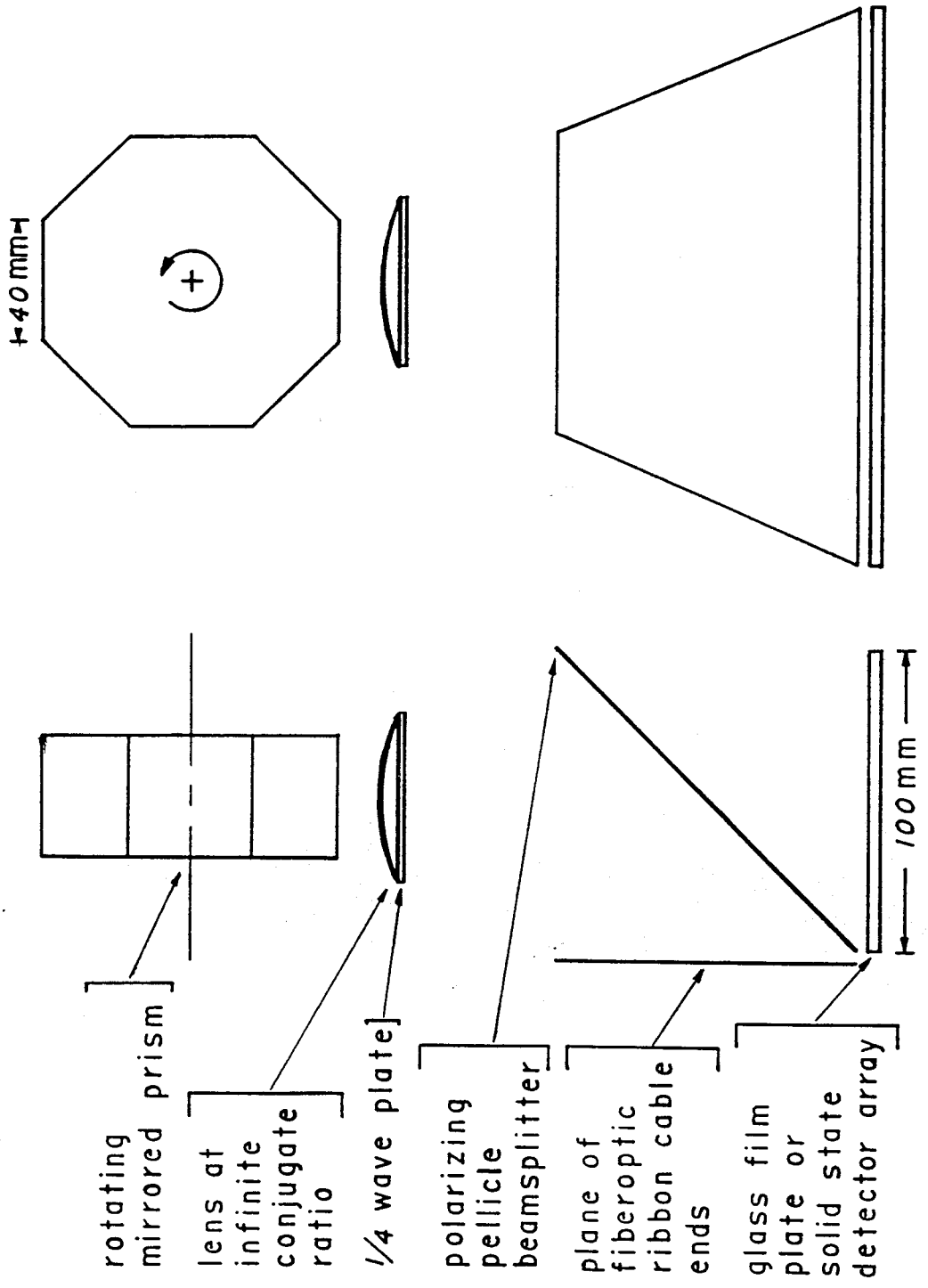

40 mm
rotating mirrored prism
lens at infinite conjugate ratio
1/4 wave plate
polarizing pellicle beamsplitter
plane of fiberoptic ribbon cable ends
glass film plate or solid state detector array
100 mm

Figure (3.8)

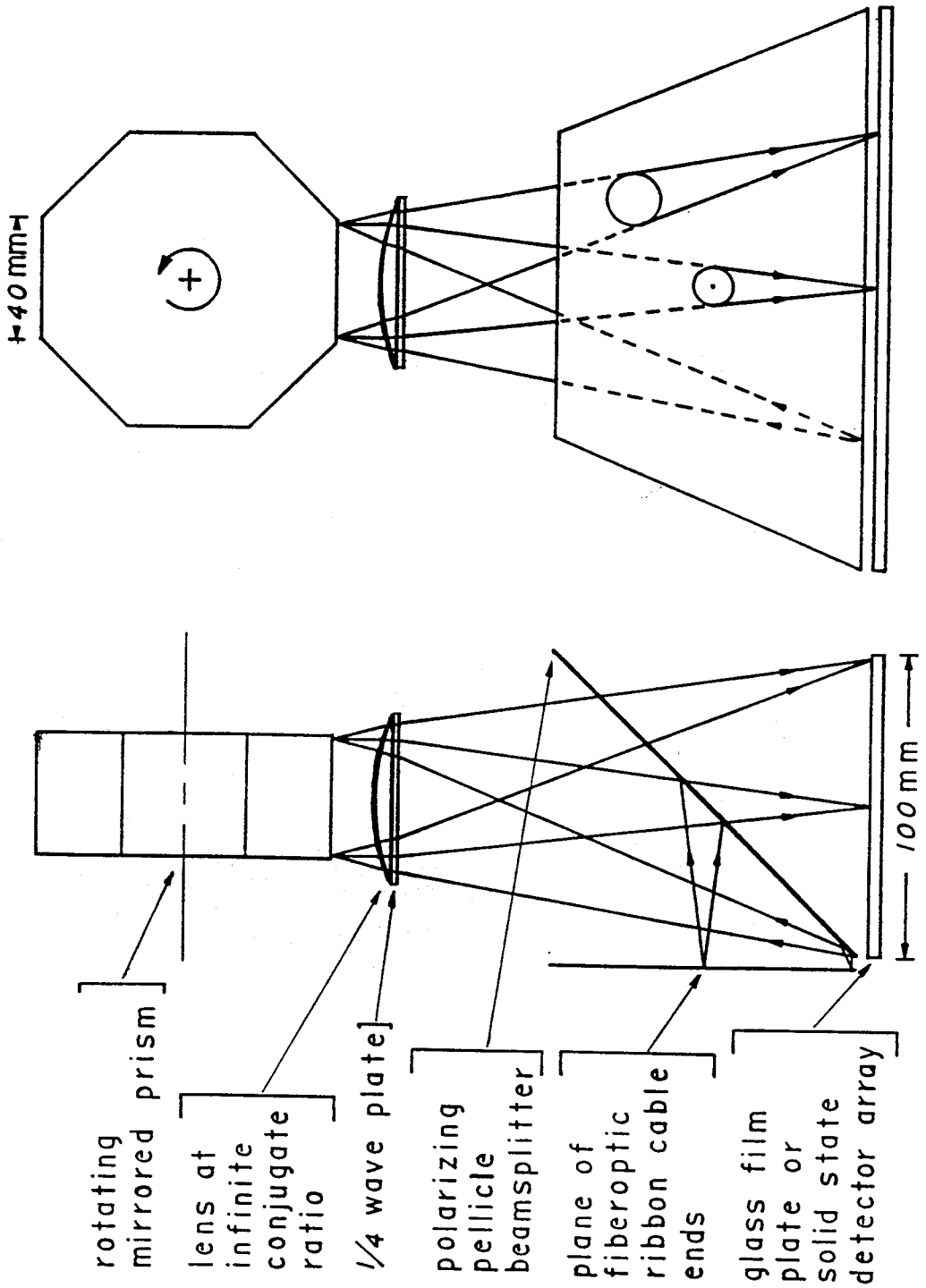

40 mm
rotating mirrored prism
lens at infinite conjugate ratio
1/4 wave plate
polarizing pellicle beamsplitter
plane of fiberoptic ribbon cable ends
glass film plate or solid state detector array
100 mm

Figure (3.9)

3.4.1 Manufacturing Details

There are a great many subtleties which need to be observed in the manufacture of such a raster camera. Three of these aspects will be discussed here briefly for illustrative purposes.

3.4.1.1 Mirror Rotation and Translation. The mirrored surfaces of the polygonal cylinder are not rotating about an axis within the plane of each mirror surface but about the center axis of the polygon. As an optical component, then, these mirrored surfaces are not only rotating but translating as well. The component of translation of the mirror in the direction of its surface normal will have a defocussing effect on the imaging system of which it is a part in all but one case. This special case, where normal translation of the mirror has no defocussing effect, is the one illustrated in the example raster camera. Since the lens is being used at an infinite conjugate ratio, a normal translation of the mirror will not defocus the optics.

3.4.1.2 Thermal expansion can obviously have a deleterious effect on almost any optical system by altering the geometric arrangement of the components in small but undesirable ways. Since a raster camera cannot use a rolled up photographic emulsion as a recording medium, it becomes progressively more bulky as the total recording capacity increases. If one considers the present example, the maximum diameter of the entire device is less than 0.5 m, but this dimension is still quite sizable given that it is desirable to limit thermally induced changes in geometry to levels of negligible optical importance ($< 1\,\mu$m). If one considers a possible temperature range of $\pm 5°$C, a characteristic dimension of 0.2 m, and an acceptable thermal expansion or contraction for this characteristic dimension of $< \pm 1\,\mu$m, then one requires a coefficient of thermal expansion of $< 10^{-6}/°$C. Such a requirement will necessitate the use of materials more exotic than aluminium or steel. Given that the modern grades of Super–Invar (Fe-Ni-Co alloys) have coefficients of thermal expansion of $\approx 2 \times 10^{-7}/°$C, this availability is another distinct improvement over the design limitations incurred by earlier developers of raster cameras.

3.4.1.3 Fiberoptic Component. The most specialized component in a raster camera is the device which separates the closely spaced pixels of an optical image into very widely spaced lines of pixels prior to rastering by the switching portion of the camera. It is important to establish a reliable manufacturing process for fabrication of this highly specialized component which is economical for very small lot manufacturing. There are several possible approaches to this manufacturing problem. Two of these possibilities are outlined here based on two of the different precursors from which the final fiberoptic device might be fabricated. Fiberoptic filament may be purchased in bulk in many forms, two of which are: spools consisting of one continuous filament; and fiberoptic ribbon cable (a line array of fiberoptic filaments adhesively bonded to form a thin ribbon). One may contemplate manufacturing processes utilizing either of these precursor forms.

3.4.1.3.a Ribbon based design: As purchased, lengths of fiberoptic ribbon cable are already separated from each other. One only needs to fuse together all of the separate segments of ribbon (lines) at one end into a large two–dimensional array of closest packed filaments, leaving the other end of all of the ribbon segments free (not fused together). In this regard, the adhesive which holds together the ribbons needs to be minimized in the final two–dimensional filament array. Further, it is important that filaments in the array be as regularly spaced as possible with a minimum of dislocations and of fiber crossovers. Only a small portion of the relevant manufacturing technology will be mentioned. In [25], Ashbee describes devices by which it is possible to achieve super-high fiber volume fraction in unidirectional fiber composites using bi–dimensional compression molding on a very matrix–rich uncured form of the composite layup. The final compacted composite attains a packing factor consistent with filaments which are touching in a hexagonal closest–packed configuration. Quite low levels of dislocated or crossed–over filaments were achieved. In effect, the process of extruding the viscous matrix from between the filaments (forcing it to flow longitudinally, parallel to the filaments until it may exit the bundle of filaments where the bundle and the bi–dimensional press both

terminate) has a stabilizing influence on the spacing between filaments. The viscous drag on the flowing matrix is lowest where the space between filaments is largest and hence the matrix flows or is extruded fastest from the widest gaps between adjacent filaments. This aspect of viscosity as it influences the extrusion of excess matrix leads to a natural centering of the individual filaments with relation to their original neighboring filaments. A modification of these techniques might be used in the present context.

3.4.1.3.b Single filament based design: Starting with a single continuous fiberoptic filament wound onto a spool suggests an alternative technique for fabricating the final "image array splitting device" made of fiberoptic filaments to be used in the raster camera. Familiarity with a spool of thread or an electrical transformer for that matter, gives one some notion of the extreme regularity with which modern machinery can wind successive layers of filament without over-crossing a filament in the process. If the winding spool is of large diameter, then a fairly long, coherent (no exchange of next nearest filament neighbors) fiberoptic bundle can be fabricated by winding up a bundle on the spool and then severing the resulting hoop of filaments.

One can envision a slightly more complex winding machine which inserts a very thin separating sheet of material between each new layer of filaments for a fraction of the total circumference of the spool. After winding, the resulting hoop of filaments could be severed into two pieces at diametrically opposite points of the hoop. One of the severing points would coincide with the location where the many layers of separating sheet had been inserted between the layers of the bundle. The other severing point would occur at a location where the bundle was free of separating sheets and the filaments were closest packed. In this way, and using binding adhesives to stabilize the wound configuration of the filaments, two of the fiberoptic "image array splitting devices" might be fabricated from each wound hoop of fibers.

References

1. Ashby, M.F. and Jones, R.H.J., *Engineering Materials 1, An Introduction to their Properties and Applications,* Pergamon Press, Oxford, 1980.

2. Broek, D., *Elementary Engineering Fracture Mechanics, 4th rev. ed.*, Section 4.4, Martinus Nijhoff Pub., 1987.

3. Knops, R.J. and Payne, L.E., *Uniqueness Theorems in Linear Elasticity,* pp. 90-91, Volume 19 of Springer Tracts in Natural Philosophy, 1971.

4. Pfaff, R.D., Ravichandar, K. and Knauss, W.G., *Dynamic Fracture in Viscoelastic Solids*, GALCIT SM Report 82-2, California Institute of Technology, Pasadena, California, 1982.

5. Pfaff, R.D. and Knauss, W.G., *Lorentz Force Loading with Ultrahigh Current Flux Densities*, GALCIT SM Report 91-00, California Institute of Technology, Pasadena, California, 1991.

6. Rice, J.R., "A Path Independent Integral and the Approximate Analysis of Strain Concentration by Notches and Cracks," *Journal of Applied Mechanics,* June 1968, pp. 379-386.

7. Shih, C.F., "*J*-Dominance Under Plane Strain Fully Plastic Conditions: The Edge Crack Panel Subject to Combined Tension and Bending," *International Journal of Fracture,* Vol. 29, pp. 73-84, 1985.

8. Alblas, J.E., "Theory of the Three-Dimensional Stress State in a Plate with a Hole," Thesis, H.J. Paris, Amsterdam, 1957: NASA TT F-10, 451.

9. Youngdahl, C.K., Sternberg, E., "Three-Dimensional Stress Concentration Around a Cylindrical Hole in a Semi-Infinite Elastic Body," *Journal of Applied Mechanics*, Dec. 1966, pp. 855-865.

10. Nakamura, T. and Parks, D.M., "Three-Dimensional Field Near the Crack Front of a Thin Elastic Plate," presented at the 24th Annual Meeting of the Society of Engineering Science, September 21-23, 1987, Salt Lake City, Utah.

11. Benthem, J.P., "State of Stress at the Vertex of a Quarter-Infinite Crack in a Half-Space," *International Journal of Solids and Structures,* Vol. 13, 1977, pp. 479-492.

12. Kanninen, M.F., and Popelar, C.H., *Advanced Fracture Mechanics*, Section 3.4, Oxford University Press, Oxford, 1985.

13. Schultheisz, C.R., Doctoral Dissertation, California Institute of Technology, Pasadena, California, 1991.

14. *Optics Guide 4*, p. 1.24, Melles Griot, Irvine, CA 92714, 1988.

15. Born, M., and Wolf, E., *Principles of Optics, 6th ed.*, Pergamon Press, Oxford, 1980.

16. Hecht, E., *Optics, 2d ed.*, Addison-Wesley Publishing, 1987.

17. Hadland, J., "A Review of the Current State of the Art," Proc. of the 11th Inter. Congress on High-Speed Photography, pp.54-60, 1974.

18. Brixner, B., *Rev. Scient. Instr.*, Vol. 30, No. 11, pp. 1041-1048, 1959.

19. Timoshenko, S.P., and Goodier, J.N., *Theory of Elasticity, 3d ed.*, McGraw-Hill, Inc., 1970.

20. Anderson, T., "Model 200 Reflecting-Optics Sweep Camera," Proc. of the 5th Inter. Congress on High-Speed Photography, New York, 1962.

21. Science and the Citizen, *Scientific American*, May, 1988, p. 26.

22. Dubovik, A.S., *Photographic Recording of High-Speed Processes*, Science Publishing House, Moscow, 1964, NASA Technical Translation, NASA TT F-377.

23. Dainty, J.C., and Shaw, R., *Image Science*, Academic Press, London, 1974.

24. Freund, H.P., and Parker, R.K., "Free Electron Lasers," *Scientific American*, April, 1989, pp. 84-89.

25. Ashbee, K.H.G., "Bi–Dimensional Compression Moulding of Super–High Fibre Volume Fraction Composites," *Journal of Composite Materials*, March, 1986, pp. 114-124.

26. Neuber, H., *Theory of Notch Stresses, Principles for Exact Stress Calculation*, Translated by F.A. Raven, J.A. Edwards, Ann Arbor, Michigan, 1946.

27. Green, A.E., "The Elastic Equilibrium of Isotripic Plates and Cylinders," Proceedings of The Royal Society (London), Series A, 195, 1949, p. 533.

Appendix A: 2-D Full-Field Elastic Solution for Notched Plates

Some of the most common configurations of fracture test specimens (4 point bend, 3 point bend, compact tension) generate the high stresses in the neighborhood of the notch, or crack, tip through the application of boundary tractions which create significant bending stresses in the body. However, as a first order analytic approximation to the stress fields in the aforementioned specimen configurations, it has been common in the fracture mechanics community to use functions which are solutions to problems with the simplest possible far field conditions (for example: the elliptic hole in an infinite plate under simple tension). This simplification is not the result of a paucity of available solutions to relevant problems with more suitable far field stess conditions. Neuber, in his book, *Theory of Notch Stresses** [26], gives solutions for potential functions satisfying the biharmonic equation for infinite planar bodies with elliptic or hyperbolic notches and with stress boundary conditions at infinity representing pure bending.

Neuber's solutions for the case of pure bending employ potential functions which are somewhat cumbersome. Perhaps for this reason, he does not perform the differentiation to yield the corresponding stress distribution in the body. It is a straightforward, if tedious process, to derive the stresses from these potential functions. For pure bending of an infinite plate containing an elliptic cutout, the potential function (in (ξ, η) elliptic coordinates) given by Neuber is:

$$\widehat{\phi_0} = S_b c^2 \frac{1}{96}\Big[\Big(-\cosh 3\xi - 3\cosh\xi + Ae^{-3\xi} + 6De^{-2\xi}\sinh\xi\Big)(-\cos 3\eta) + \Big(3\cosh 3\xi + 9\cosh\xi + Be^{\xi} + Ce^{-\xi} + 6De^{-2\xi}\sinh\xi\Big)\cos\eta\Big]$$

* translated from the German for the David Taylor Model Basin in 1946

where, for the particular problem at hand (pure bending at infinity),

$$A = e^{3\xi_0}(\cosh 3\xi_0 + 3\cosh \xi_0 - 12\cosh^2 \xi_0 \sinh \xi_0) \quad ;$$

$$B = -12\cosh^2 \xi_0 \quad ; \quad C = 0 \quad ; \quad D = 2e^{2\xi_0}\cosh^2 \xi_0 \quad ;$$

ξ_0 designates the boundary of the elliptic coutout ;

$$x = c\cosh \xi \cos \eta \quad , \quad y = c\sinh \xi sin\eta \quad .$$

The constant of proportionality, S_b, is defined from the far–field traction boundary condition, $(d\sigma_{yy}/dx)|_{\xi\to\infty} = S_b/c$.

From this potential function $\widehat{\phi_0}$, one derives the following stress field:

$$\frac{1}{2}(\sigma_{\xi\xi} + \sigma_{\eta\eta}) = \frac{1}{2}\nabla^2\widehat{\phi_0} = \frac{1}{2}\left[c^2(\cosh 2\xi - \cos 2\eta)\right]^{-1}\left[\frac{\partial^2}{\partial\xi^2} + \frac{\partial^2}{\partial\eta^2}\right]\widehat{\phi_0}$$

$$\sigma_{\xi\eta} = \frac{\left[-2\dfrac{\partial^2}{\partial\eta\partial\xi} + \dfrac{2\sin 2\eta}{\cosh 2\xi - \cos 2\eta}\dfrac{\partial}{\partial\xi} + \dfrac{2\sinh 2\xi}{\cosh 2\xi - \cos 2\eta}\dfrac{\partial}{\partial\eta}\right]\widehat{\phi_0}}{c^2(\cosh 2\xi - \cos 2\eta)}$$

$$\frac{1}{2}(\sigma_{\xi\xi} - \sigma_{\eta\eta}) = \frac{\left[\left(\dfrac{\partial^2}{\partial\eta^2} - \dfrac{\partial^2}{\partial\xi^2}\right) + \dfrac{2\sinh 2\xi}{\cosh 2\xi - \cos 2n}\dfrac{\partial}{\partial\xi} - \dfrac{2\sin 2n}{\cosh 2\xi - \cos 2n}\dfrac{\partial}{\partial\eta}\right]\widehat{\phi_0}}{c^2(\cosh 2\xi - \cos 2n)}$$

Carrying out the indicated operations there results, in detail,

$$\frac{1}{2}(\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1aa)} = S_b\Bigg[e^{2\xi_0}\cosh^2\xi_0\cos\eta\left(\frac{\sinh\xi}{\cosh 2\xi-\cos 2\eta}+(\sinh\xi-\cosh\xi)\right) + \frac{1}{2}\cosh\xi\cos\eta\Bigg]$$

$$\frac{1}{2}(\sigma_{\xi\xi}-\sigma_{\eta\eta})_{(*1aa)} =$$

$$S_b\Bigg[(\cosh\xi\cos\eta)\left(-\frac{1}{2}\right)(\cosh 2\xi\cos 2\eta-1)/(\cosh 2\xi-\cos 2\eta)$$

$$+e^{2\xi_0}\cosh^2\xi_0\sinh\xi\cos\eta(\cos 2\eta-1)\left(\frac{(\cosh 2\xi-\cos 2\eta)(\cosh 2\xi-\sinh 2\xi)+1}{(\cosh 2\xi-\cos 2\eta)^2}\right)$$

$$+\left(\frac{e^{2\xi_0}\cosh^2\xi_0\sinh\xi_0\cos\eta}{(\cosh 2\xi-\cos 2\eta)^2}\right)\left(\frac{1}{2}\left(e^{\xi-\xi_0}+e^{-\xi+\xi_0}\right)(1-\cosh 2\xi+\sinh 2\xi)\right.$$

$$\left.-e^{-\xi+\xi_0}\left\{[(\cosh 2\xi-\sinh 2\xi)(2\cos 2\eta-1)-1](\cosh 2\xi-\cos 2\eta)+\sinh 2\xi\right\}\right)\Bigg]$$

$$\sigma_{\xi\eta(*1aa)} =$$

$$S_b\Bigg[(\cosh\xi\cos\eta)\left(\frac{1}{2}\right)(\sinh 2\xi\sin 2\eta)/(\cosh 2\xi-\cos 2\eta)$$

$$+e^{2\xi_0}\cosh^2\xi_0\sinh\xi\sin\eta\left[\left(\frac{(\cosh 2\xi-\sinh 2\xi)\cos 2\eta-1}{\cosh 2\xi-\cos 2\eta}\right)+\frac{\sinh 2\xi}{(\cosh 2\xi-\cos 2\eta)^2}\right]$$

$$+\left(\frac{e^{2\xi_0}\cosh^2\xi_0\sinh\xi_0\sin\eta}{(\cosh 2\xi-\cos 2\eta)^2}\right)\left(\frac{1}{2}\left(e^{\xi-\xi_0}-e^{-\xi+\xi_0}\right)(1+\cosh 2\xi-\sinh 2\xi)\right.$$

$$\left.-e^{-\xi+\xi_0}\left\{[(\cosh 2\xi-\sinh 2\xi)(2\cos 2\eta+1)-1](\cosh 2\xi-\cos 2\eta)+\sinh 2\xi\right\}\right)\Bigg]$$

where the subscripts identify the solution to this particular component problem for later reference.

With this solution in hand it is possible to express the two dimensional elastic stress fields for the common fracture test specimen configurations accurately and in closed form. The additional functions which are necessary for describing the stress fields of such finite geometries may be derived from solutions found in the standard text, "Theory of Elasticity" [19].

The two primary of these stress fields are introduced immediately following; they correspond to the problems of an infinite plate with an elliptic cutout with far field loading of biaxial stress and of pure distortional stress, respectively. Additional elementary stress fields are introduced as necessary for formulating a complete description for the stress field of a notched 3-point bend specimen, a process which will be illustrative of a general approach applicable to other standard test geometries.

A.1 Infinite plate with an elliptical hole under uniform (biaxial) stress

(problem $*1ab$):

$$\frac{1}{2}(\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1ab)} = S_u\left(\frac{\sinh 2\xi}{\cosh 2\xi - \cos 2\eta}\right)$$

$$\frac{1}{2}(\sigma_{\xi\xi}-\sigma_{\eta\eta})_{(*1ab)} = S_u\left(\frac{-\sinh 2\xi}{\cosh 2\xi - \cos 2\eta} + \frac{(\cosh 2\xi - \cosh 2\xi_0)\sinh 2\xi}{(\cosh 2\xi - \cos 2\eta)^2}\right)$$

$$\sigma_{\xi\eta(*1ab)} = S_u\left(\frac{(\cosh 2\xi - \cosh 2\xi_0)\sin 2\eta}{(\cosh 2\xi - \cos 2\eta)^2}\right)$$

A.2 Infinite plate with an elliptical hole under pure distortional far–field stress:

With $x = c\cosh\xi\cos\eta$ and $y = c\sinh\xi\sin\eta$, define the rotated coordinate frame

$$x' = x\cos\beta + y\sin\beta; \qquad y' = -x\sin\beta + y\cos\beta.$$

Take the far–field traction boundary conditions in the form

$$\sigma_{x'x'}|_{\xi\to\infty} = S_d; \qquad \sigma_{y'y'}|_{\xi\to\infty} = -S_d; \qquad \sigma_{x'y'}|_{\xi\to\infty} = 0.$$

$$\frac{1}{2}(\sigma_{\xi\xi} + \sigma_{\eta\eta}) = S_d e^{2\xi_0}\left[\cos 2\beta\left(1 - \frac{\sinh 2\xi}{\cosh 2\xi - \cos 2\eta}\right) - \sin 2\beta\frac{\sin 2\eta}{\cosh 2\xi - \cos 2\eta}\right]$$

$$\frac{1}{2}(\sigma_{\eta\eta} - \sigma_{\xi\xi} + 2i\sigma_{\xi\eta}) =$$

$$S_d e^{2\xi_0}\left[-e^{i2\beta}\left(\frac{\sinh 2\xi}{\cosh 2\xi - \cos 2\eta} + \frac{(\cosh 2\xi - \cosh 2\xi_0)(-\sinh 2\xi + i\sin 2\eta)}{(\cosh 2\xi - \cos 2\eta)^2}\right)\right.$$

$$\left. + \left(\frac{e^{i2\beta}\frac{1}{2}e^{2\xi_0}\left(1 - e^{-2\xi}e^{-i2\eta}\right) + e^{-i2\beta}\frac{1}{2}e^{-2\xi_0}\left(1 - e^{2\xi}e^{i2\eta}\right)}{\cosh 2\xi - \cos 2\eta}\right)\right]$$

We shall have need for the special case of $\beta = \pi/2$.

$$\frac{1}{2}(\sigma_{\xi\xi} + \sigma_{\eta\eta})_{(*1ac)} = S_d e^{2\xi_0}\left(\frac{\sinh 2\xi}{\cosh 2\xi - \cos 2\eta} - 1\right)$$

$$\frac{1}{2}(\sigma_{\xi\xi} - \sigma_{\eta\eta})_{(*1ac)} = S_d e^{2\xi_0}\left(\frac{-\sinh 2\xi}{\cosh 2\xi - \cos 2\eta} + \frac{(\cosh 2\xi - \cosh 2\xi_0)\sinh 2\xi}{(\cosh 2\xi - \cos 2\eta)^2}\right.$$

$$\left. + \frac{\cosh 2\xi_0 - \cos 2\eta\cosh(2\xi - 2\xi_0)}{\cosh 2\xi - \cos 2\eta}\right)$$

$$\sigma_{\xi\eta(*1ac)} = S_d e^{2\xi_0}\left(\frac{(\cosh 2\xi - \cosh 2\xi_0)\sin 2\eta}{(\cosh 2\xi - \cos 2\eta)^2} + \frac{\sin 2\eta\sinh(2\xi - 2\xi_0)}{\cosh 2\xi - \cos 2\eta}\right)$$

A.3 Configuration of the tip region of an elliptic notch:

$$x = c\cosh\xi\cos\eta, \qquad y = c\sinh\xi\sin\eta.$$

$$\left(\frac{d^2x}{dy^2}\right)_{\xi=\xi_0} = -\frac{1}{c}\frac{1}{\cos^3\eta}\left(\frac{\cosh\xi_0}{\sinh^2\xi_0}\right)$$

$$\left[\left(\frac{d^2x}{dy^2}\right)_{\xi=\xi_0}\right]_{\eta=0} = -\frac{1}{c}\frac{\cosh\xi_0}{\sinh^2\xi_0} \equiv -\frac{1}{R_c}$$

where R_c is the radius of curvature at the elliptical slot tip.

$$R_c = c\frac{\sinh^2\xi_0}{\cosh\xi_0} = \left[c(\cosh\xi_0 - 1)\right]\left(1 + \frac{1}{\cosh\xi_0}\right)$$

Compare this with the distance from the focal point of the ellipse to the tip, which is $[c(\cosh\xi_0 - 1)]$.

$$R_c/[c(\cosh\xi_0 - 1)] = 1 + \frac{1}{\cosh\xi_0}$$

For $|\xi_0| << 1$, $(R_c/[c(\cosh\xi_0 - 1)])|_{\xi_0\to 0} = 2$, that is, for ξ_0 small, the radius of curvature of the elliptical slot tip is approximately twice the distance form the ellipse's focal point to its tip.

Let us consider further the shape of elliptical notches of slender aspect ratio (ξ_0 small).

$$\frac{x^2}{\cosh^2\xi_0} + \frac{y^2}{\sinh^2\xi_0} = c^2 = {R_c}^2\frac{\cosh^2\xi_0}{\sinh^4\xi_0}$$

Introducing $X = x - c$,

$$\Rightarrow \text{along } \xi = \xi_0, \; -\left(\frac{1}{\cosh\xi_0}\right)\left(\frac{R_c}{c}\right)\left(\frac{-X}{R_c}\right)^2 + 2\left(\frac{1}{\cosh\xi_0}\right)\left(\frac{-X}{R_c}\right) + 1 = \left(\frac{y}{R_c}\right)^2,$$

where $\cosh\xi_0 = \sqrt{1 + \frac{1}{4}(\frac{R_c}{c})^2} + \frac{1}{2}(\frac{R_c}{c})$.

Special cases:

$$y\Big|_{\substack{\xi=\xi_0\\X=0}} = R_c; \qquad \lim_{\xi_0\to 0}\left(\frac{y}{R_c}\right)^2 = 1 + 2\left(\frac{-X}{R_c}\right) \text{ along } \xi = \xi_0.$$

To express this last point in words, as the aspect ratio of the elliptical slot becomes extreme (i.e., when R_c/c becomes small), the shape of the notch tip region tends toward a parabola with a normalized difference which is of order R_c/c. This convergence indicates that if one intended to approximate a real slot tip of some particular shape with an ellipse of narrow aspect ratio, from a practical point of view, little flexibility is available since narrow ellipses are generally indistinguishable in the shape of the ellipse tip. Only the absolute size of the near parabolic shape, characterized by R_c, has any significance as a matching parameter.

A.4 Constants of proportionality for the embedded singularities generated by a traction free elliptic cutout in a plate:

Let us consider the behavior of

$$\frac{\sinh\xi\cos\eta}{\cosh 2\xi-\cos 2\eta} \qquad \text{and} \qquad \frac{\sinh 2\xi}{\cosh 2\xi-\cos 2\eta}$$

in the vicinity of the ellipse focal point, $\xi=0,\ \eta=0$. Define

$$\begin{aligned}
\widetilde{x} &\equiv \frac{x}{c} \equiv \cosh\xi\cos\eta;\\
\widetilde{y} &\equiv \frac{y}{c} \equiv \sinh\xi\sin\eta;\\
\widetilde{X} &\equiv \frac{X}{c} \equiv \frac{x-c}{c} = \widetilde{x}-1 = \cosh\xi\cos\eta-1.
\end{aligned}$$

Now

$$\frac{\sinh 2\xi}{\cosh 2\xi-\cos 2\eta} = 2\left(\frac{\sinh\xi\cos\eta}{\cosh 2\xi-\cos 2\eta}\right) \times\left(1-\frac{\sqrt{\widetilde{X}^2+\widetilde{y}^2}\left(\sqrt{1+\widetilde{X}+\frac{1}{4}(\widetilde{X}^2+\widetilde{y}^2)}-\frac{1}{2}\sqrt{\widetilde{X}^2+\widetilde{y}^2}\right)}{1+\widetilde{X}}\right)$$

and

$$\frac{\sinh\xi\cos\eta}{\cosh 2\xi-\cos 2\eta} = \frac{1}{4}\left[\frac{\widetilde{X}+\sqrt{\widetilde{X}^2+\widetilde{y}^2}+\frac{1}{2}(\widetilde{X}^2-\widetilde{y}^2)+\sqrt{\widetilde{X}^2+\widetilde{y}^2}\left(\sqrt{1+\widetilde{X}+\frac{1}{4}(\widetilde{X}^2+\widetilde{y}^2)}-1\right)}{(\widetilde{X}^2+\widetilde{y}^2)\left(1+\widetilde{X}+\frac{1}{4}(\widetilde{X}^2+\widetilde{y}^2)\right)}\right]^{1/2}$$

So,

$$\lim_{(\widetilde{X}^2+\widetilde{y}^2)\to 0}\left(\frac{\sinh\xi\cos\eta}{\cosh 2\xi-\cos 2\eta}\right) \longrightarrow \frac{1}{4}\left[\frac{\widetilde{X}+\sqrt{\widetilde{X}^2+\widetilde{y}^2}}{(\widetilde{X}^2+\widetilde{y}^2)}\right]^{1/2}$$

and

$$\lim_{(\widetilde{X}^2+\widetilde{y}^2)\to 0}\left(\frac{\sinh 2\xi}{\cosh 2\xi-\cos 2\eta}\right) \longrightarrow \frac{1}{2}\left[\frac{\widetilde{X}+\sqrt{\widetilde{X}^2+\widetilde{y}^2}}{(\widetilde{X}^2+\widetilde{y}^2)}\right]^{1/2}$$

Further, let us define:

$$\tilde{r}^2 \equiv \tilde{X}^2 + \tilde{y}^2; \qquad r \equiv c\tilde{r} = \sqrt{X^2 + y^2}$$

$$\cos\theta \equiv \frac{\tilde{X}}{\tilde{r}}; \qquad \sin\theta \equiv \frac{\tilde{y}}{\tilde{r}}; \qquad \text{i.e., } \theta = \arctan\left(\frac{\tilde{y}}{\tilde{X}}\right).$$

Hence,

$$\left[\frac{\tilde{X} + \sqrt{\tilde{X}^2 + \tilde{y}^2}}{(\tilde{X}^2 + \tilde{y}^2)}\right]^{1/2} = \sqrt{\frac{2}{\tilde{r}}}\cos\frac{\theta}{2} = \sqrt{c}\sqrt{\frac{2}{r}}\cos\frac{\theta}{2},$$

and

$$\lim_{\tilde{r}\to 0}\left(\frac{\sinh\xi\cos\eta}{\cosh 2\xi - \cos 2\eta}\right) \longrightarrow \frac{1}{2}\sqrt{\pi c}\left(\sqrt{\frac{1}{2\pi r}}\cos\frac{\theta}{2}\right);$$

$$\lim_{\tilde{r}\to 0}\left(\frac{\sinh 2\xi}{\cosh 2\xi - \cos 2\eta}\right) \longrightarrow \sqrt{\pi c}\left(\sqrt{\frac{1}{2\pi r}}\cos\frac{\theta}{2}\right).$$

A.5 Remarks on the accuracy of using the generalized plane stress solution as an approximation to the fully three dimensional problem of a traction free elliptic cutout in a plate.

Recall that for generalized plane stress (where the subscript a designates a thickness average function, i.e., a function in only the two in-plane variables, for example $\phi_a = \phi_a(x,y)$):

$$\sigma_{xx}(\phi) = \frac{\partial^2}{\partial y^2}(\phi); \quad \sigma_{xy}(\phi) = -\frac{\partial^2}{\partial x \partial y}(\phi); \quad \sigma_{yy}(\phi) = \frac{\partial^2}{\partial x^2}(\phi)$$

$$\sigma_{\xi\xi}(\phi) = \frac{\left[\left[\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2}\right] + \left[-\left(\frac{\partial^2}{\partial \xi^2} - \frac{\partial^2}{\partial \eta^2}\right) + \frac{2\sin 2\alpha}{\sin 2\eta}\frac{\partial}{\partial \xi} - \frac{2\sin 2\alpha}{\sinh 2\xi}\frac{\partial}{\partial \eta}\right]\right](\phi)}{[c^2(\cosh 2\xi - \cos 2\eta)]}$$

$$\sigma_{\xi\eta}(\phi) = \frac{\left[-2\frac{\partial^2}{\partial \eta \partial \xi} + \frac{2\sin 2\alpha}{\sinh 2\xi}\frac{\partial}{\partial \xi} + \frac{2\sin 2\alpha}{\sin 2\eta}\frac{\partial}{\partial \eta}\right](\phi)}{[c^2(\cosh 2\xi - \cos 2\eta)]}$$

$$\sigma_{\eta\eta}(\phi) = \frac{\left[\left[\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2}\right] - \left[-\left(\frac{\partial^2}{\partial \xi^2} - \frac{\partial^2}{\partial \eta^2}\right) + \frac{2\sin 2\alpha}{\sin 2\eta}\frac{\partial}{\partial \xi} - \frac{2\sin 2\alpha}{\sinh 2\xi}\frac{\partial}{\partial \eta}\right]\right](\phi)}{[c^2(\cosh 2\xi - \cos 2\eta)]}$$

where $\sin 2\alpha = \sinh 2\xi \sin 2\eta/(\cosh 2\xi - \cos 2\eta)$, and recalling that $x = c\cosh\xi\cos\eta$ and $y = c\sinh\xi\sin\eta$.

$$\sigma_{zz} = \sigma_{xz} = \sigma_{yz} = \sigma_{\xi z} = \sigma_{\eta z} = 0$$

$$\nabla^4\phi = 0 \qquad \text{and} \qquad \phi = \phi_a - \left[\frac{1}{2}\frac{\nu}{1+\nu}\frac{(t/2)^2}{3}\left(\left(\frac{\sqrt{3}z}{t/2}\right)^2 - 1\right)\right]\nabla^2\phi_a$$

Note that

$$\frac{1}{t}\int_{-t/2}^{t/2}\left[\frac{1}{2}\frac{\nu}{1+\nu}\frac{(t/2)^2}{3}\left(\left(\frac{\sqrt{3}z}{t/2}\right)^2 - 1\right)\right]dz = 0$$

$$\nabla^4 \phi_a = 0; \qquad \nabla^2 \phi_a = \frac{1}{t}\int_{-t/2}^{t/2} (\sigma_{\xi\xi} + \sigma_{\eta\eta}) dz.$$

Consider the case

$$\nabla^2 \phi_a = \frac{1}{t}\int_{-t/2}^{t/2} (\sigma_{\xi\xi} + \sigma_{\eta\eta}) dz = 2S\left(\frac{\sinh 2\xi}{\cosh 2\xi - csn2}\right)$$

$$\sigma_{\xi\xi} = \sigma_{\xi\xi_a}(\xi,\eta) - 2S\left[\left(\frac{\sqrt{3}z}{t/2}\right)^2 - 1\right]\left[\frac{2\nu}{1+\nu}\left(\frac{t/2}{c}\right)^2\right] \times \left[\frac{\sinh 2\xi(1 - \cosh 2\xi \cos 2\eta + \sin^2 2\eta}{(\cosh 2\xi - \cos 2\eta)^4}\right]$$

$$\sigma_{\xi\eta} = \sigma_{\xi\eta_a}(\xi,\eta) - 2S\left[\left(\frac{\sqrt{3}z}{t/2}\right)^2 - 1\right]\left[\frac{2\nu}{1+\nu}\left(\frac{t/2}{c}\right)^2\right] \times \left[\frac{\sin 2\eta(1 - \cosh 2\xi \cos 2\eta - \sinh^2 2\xi}{(\cosh 2\xi - \cos 2\eta)^4}\right]$$

$$\sigma_{\eta\eta} = \sigma_{\eta\eta_a}(\xi,\eta) - 2S\left[\left(\frac{\sqrt{3}z}{t/2}\right)^2 - 1\right]\left[\frac{2\nu}{1+\nu}\left(\frac{t/2}{c}\right)^2\right] \times \left[\frac{-\sinh 2\xi(1 - \cosh 2\xi \cos 2\eta + \sin^2 2\eta}{(\cosh 2\xi - \cos 2\eta)^4}\right]$$

Utilizing $c = R_c \frac{\cosh \xi_0}{\sinh^2 \xi_0}$, where R_c is the radius of curvature for the tip of the ellipse $\xi = \xi_0$,

$$\left(\frac{1}{c}\right)^2 = \left(\frac{1}{2R_c}\right)^2 \frac{(2\sinh^2 \xi_0)^2}{1 + \sinh^2 \xi_0}$$

Similarly, $\sinh^2 \xi_0 = \left[\left(\frac{R_c}{c}\right)\sqrt{1 + \frac{1}{4}\left(\frac{R_c}{c}\right)^2} + \frac{1}{2}\left(\frac{R_c}{c}\right)^2\right]$.

Rewriting the previous equations,

$$[\sigma_{\xi\xi} - \sigma_{\xi\xi_a}(\xi,\eta)] = 2S\left[\left(\frac{\sqrt{3}z}{t/2}\right)^2 - 1\right]\left[\frac{2\nu}{1+\nu}\left(\frac{t/2}{2R_c}\right)^2\right]\left(\frac{1}{1+\sinh^2 \xi_0}\right)\left[\left(\frac{\sinh^2 \xi_0}{\sinh^2 \xi + \sin^2 \eta}\right)^2\right] \times \left[\frac{\sinh 2\xi}{\cosh 2\xi - \cos 2\eta}\right]\left[1 - 2\sin^2 \eta - \frac{4(1 - \sin^2 \eta)\sin^2 \eta}{\sinh^2 \xi + \sin^2 \eta}\right]$$

$$[\sigma_{\xi\eta}-\sigma_{\xi\eta_a}(\xi,\eta)] =$$
$$2S\left[\left(\frac{\sqrt{3}z}{t/2}\right)^2-1\right]\left[\frac{2\nu}{1+\nu}\left(\frac{t/2}{2R_c}\right)^2\right]\left(\frac{1}{1+\sinh^2\xi_0}\right)\left[\left(\frac{\sinh^2\xi_0}{\sinh^2\xi+\sin^2\eta}\right)^2\right] \cdot$$
$$\times\left\{\sin 2\eta+\left[\frac{\sin 2\eta}{\cosh 2\xi-\cos 2\eta}\right]\left[3(1-2\sin^2\eta)-\frac{4(1-\sin^2\eta)\sin^2\eta}{\sinh^2\xi+\sin^2\eta}\right]\right\}$$

$$[\sigma_{\eta\eta}-\sigma_{\eta\eta_a}(\xi,\eta)] = -[\sigma_{\xi\xi}-\sigma_{\xi\xi_a}(\xi,\eta)]$$

where

$$\frac{\sinh 2\xi}{\cosh 2\xi-\cos 2\eta}=\frac{\sinh\xi\cosh\xi}{\sinh^2\xi+\sin^2\eta}; \qquad \frac{\sin 2\eta}{\cosh 2\xi-\cos 2\eta}=\frac{\sin\eta\cos\eta}{\sinh^2\xi+\sin^2\eta}.$$

Note that

$$\sin^2\eta=\left[\frac{\sinh^2\xi-\frac{1}{c^2}[c\cosh\xi\cos\eta-c][(c\cosh\xi\cos\eta-c)+2c)]}{1+\sinh^2\xi}\right]$$

so that, while

$$\left[1-2\sin^2\eta-\frac{4(1-\sin^2\eta)\sin^2\eta}{\sinh^2\xi+\sin^2\eta}\right]_{\substack{\xi=\xi_0\neq 0\\ \eta=0}}=1$$

$$\left[1-2\sin^2\eta-\frac{4(1-\sin^2\eta)\sin^2\eta}{\sinh^2\xi+\sin^2\eta}\right]_{\substack{\xi=\xi_0\\ (c\cosh\xi\cos\eta-c)=0}} =$$
$$\left[-1+\frac{2\sinh^2\xi_0}{(2+\sinh^2\xi_0)(1+\sinh^2\xi_0)}\right]$$

One may illustrate the dependence of the magnitude of the necessary correction to generalized plane stress at larger values of η by observing that

$$\left[\left(\frac{t/2}{2R_c}\right)^2\left[\frac{1}{1+\sinh^2\xi_0}\left(\frac{\sinh^2\xi_0}{\sinh^2\xi+\sin^2\eta}\right)^2\right]_{\substack{\xi=\xi_0\\ (c\cosh\xi\cos\eta-c)=-t/2}}\right] =$$
$$\frac{1}{16}(1+\sinh^2\xi_0)^2\Bigg/\left[\left(1-\frac{1}{2}\frac{t/2}{c}\right)+\frac{R_c}{t/2}\sqrt{1+\sinh^2\xi_0}\left(1+\frac{1}{2}\sinh^2\xi_0\right)\right]^2$$

Generalized plane stress may be considered to be the first term in a series solution which solves the fully three–dimensional problem of a traction–free elliptic cutout in a plate. This series approach for satisfying traction–free edge conditions was formulated in generality by Green [27] and implemented by Alblas [8] for the case of a circular cutout in a plate. Alblas encountered convergence difficulties with this approach as the ratio of the plate thickness to the hole radius became large. This lack of convergence was expected because the residual boundary tractions associated with the generalized plane stress approximation are obtained from second order derivatives of the plane stress solution. In cases where the plane stress solution contains an embedded singularity, the associated residual boundary tractions will have an associated embedded singularity which is two orders more singular. The foregoing relations were obtained to exemplify this greater embedded singularity for the case of an elliptic notch. It may be observed that this effect is represented by the term

$$\left[\frac{2\nu}{1+\nu}\left(\frac{t/2}{2R_c}\right)^2\right]$$

This difficulty was overcome by Sternberg [9] for the case of a circular hole in an infinite halfspace using a completely different method, where the plane strain solution is a better first approximation to the fully three–dimensional problem. For cases where the ratio of the plate thickness to the hole radius is large, but not infinite, neither plane strain nor generalized plane stress make a good first approximation. In considering an approach for solving the problem of an elliptic cutout in a plate, where the ratio of plate thickness to notch tip radius is large, one is confronted with this foregoing difficulty plus the additional complexity associated with the generalization from cylindrical coordinates to elliptical coordinates.

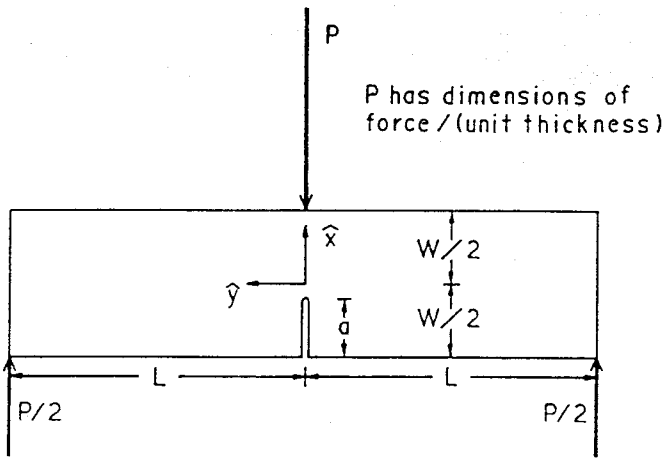
As a specific example of the inefficiency of generalized plane stress as a first order approximation, observe that

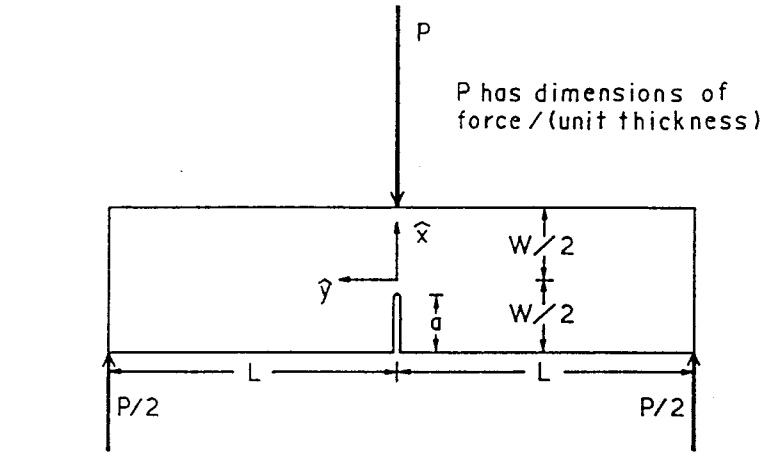
$$\left[\frac{2\nu}{1+\nu}\left(\frac{t/2}{2R_c}\right)^2\right]\Bigg|_{\substack{t=10\text{mm}\\ R_c=1/20\text{mm}\\ \nu=28/100}} = \frac{4375}{4}$$

A.6 Example Problem: Three-point-bend Configuration

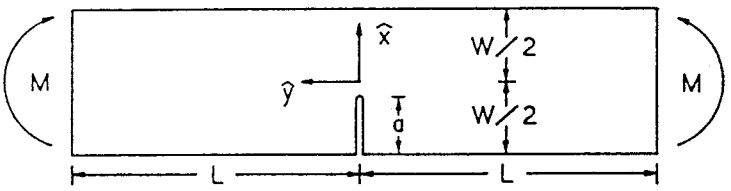
Having established the two–dimensional stress fields for an elliptic cutout in an infinite plate under far–field bending ($*1aa$), uniform biaxial loading ($*1ab$), and pure distortional loading where the principal axes of the far–field stress field are aligned with the axes of the ellipse ($*1ac$), one is in a position to formulate a description for the stress field of a notched three-point-bend specimen. This being a linear problem, extensive use of the superposition principle is permissible and, as well, becomes an organizing framework to delineate primary and secondary aspects of the stress field.

The geometry of the three-point-bend configuration is illustrated below and the separation of this loading condition into a series of component problems is illustrated immediately thereafter.

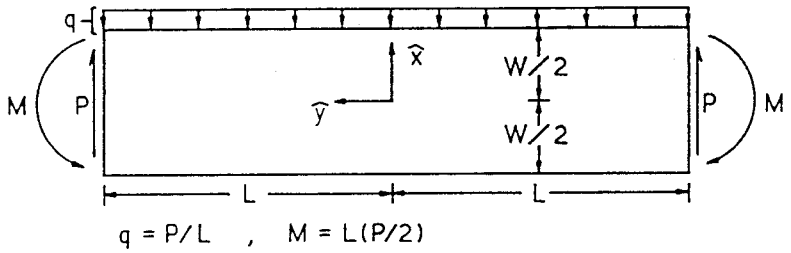



P
P has dimensions of
force / (unit thickness)
x̂
ŷ
W/2
W/2
a
L
L
P/2
P/2

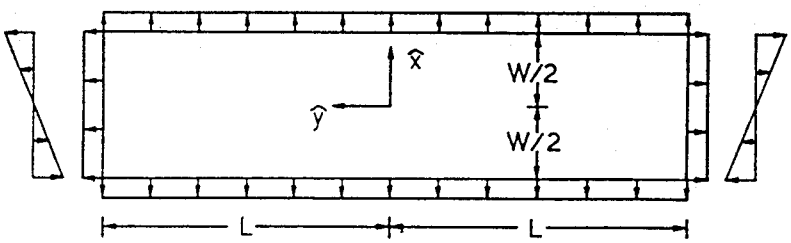
≡


M
M
x̂
ŷ
W/2
W/2
a
L
L

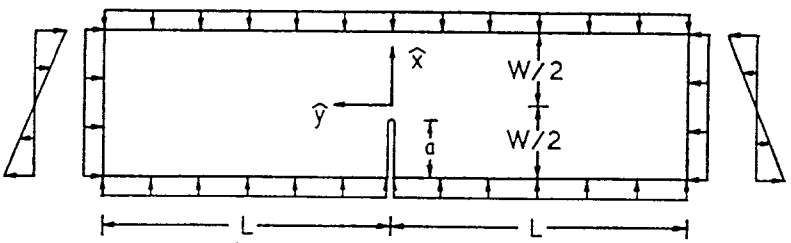
(∗1) +


q
M
P
P
M
x̂
ŷ
W/2
W/2
L
L
q = P/L , M = L(P/2)

(∗2) +

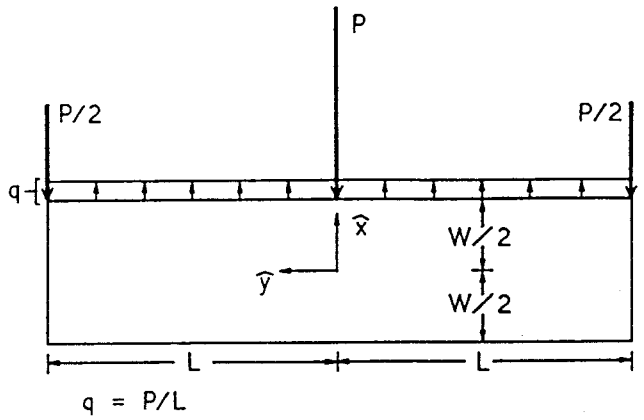

x̂
ŷ
W/2
W/2
L
L

(∗2a) +


x̂
ŷ
W/2
W/2
a
L
L

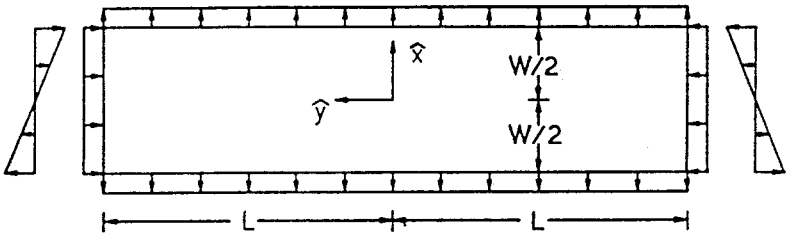
(∗2an) +

(*3)


P
P/2
P/2
q
x̂
ŷ
W/2
W/2
L
L
q = P/L

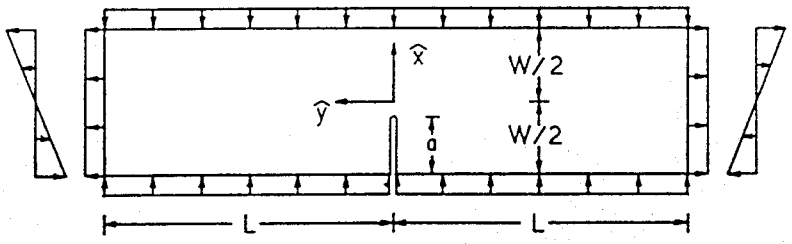
+

(*3a)


x̂
ŷ
W/2
W/2
L
L

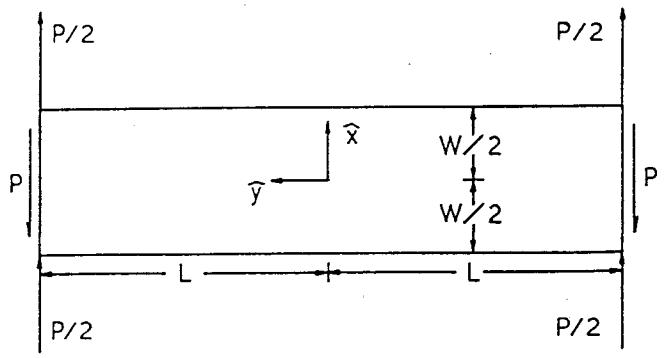
+

(*3an)


x̂
ŷ
W/2
W/2
a
L
L

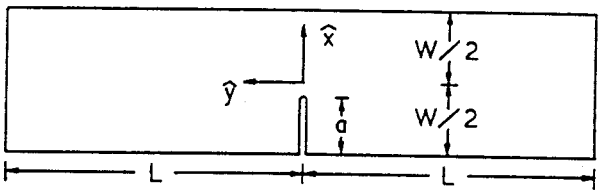
+

(*4)


P/2
P/2
P
P
x̂
ŷ
W/2
W/2
L
L
P/2
P/2

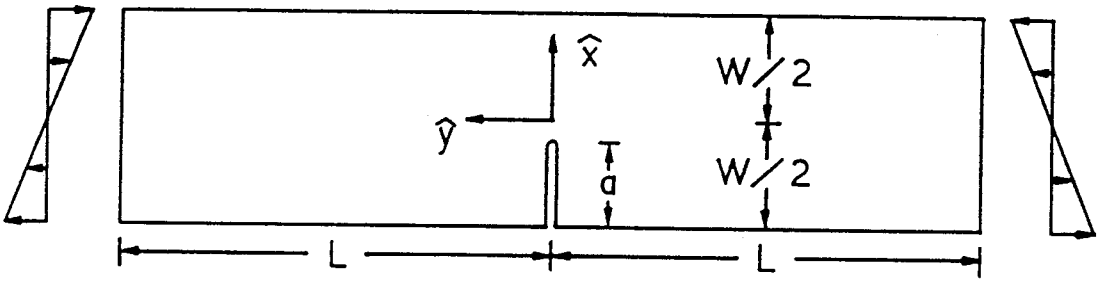
+

(*5)



+

The traction boundary conditions for problem (*5) consist of residual tractions on the notch boundary resulting from boundary value problems (*2), (*2a), (*3), (*3a) and (*4). Boundary value problems (*2a) and (*3a) have traction conditions chosen specifically to minimize the boundary tractions associated with problem (*5) and limit these tractions to a weak, higher order, traction distribution.



(∗1):

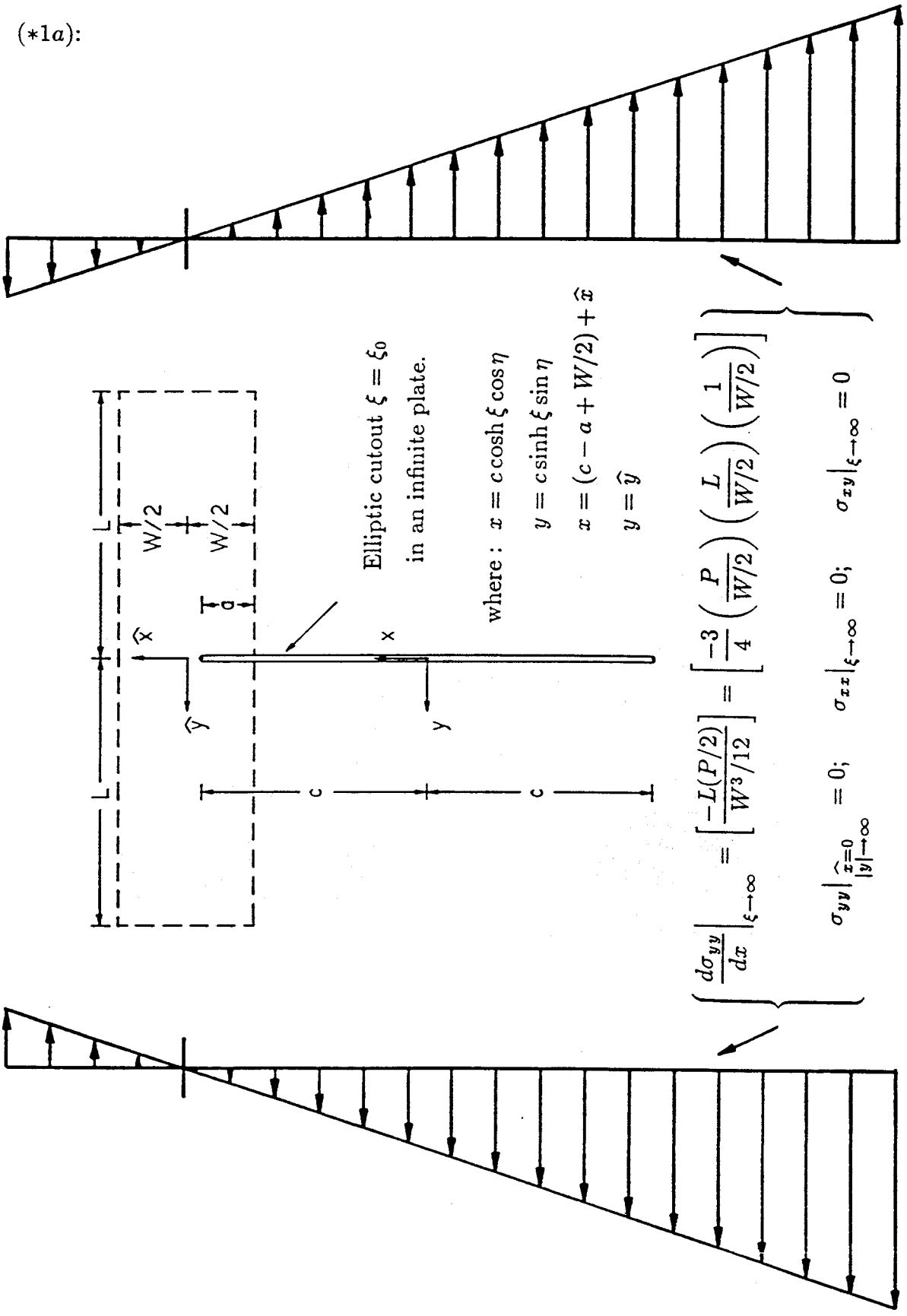
Boundary Condition:

$$\sigma_{\widetilde{yy}}|_{\widehat{y}=\pm L} = \frac{-L(P/2)}{W^3/12}(\widehat{x}) = -\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left(\frac{\widehat{x}}{W/2}\right)\Big)$$

otherwise traction free.

The solution to the boundary value problem (∗1) may be made up from the summation of boundary value problems (∗1a) and (∗1b).

$(*1a)$:


Elliptic cutout $\xi = \xi_0$
in an infinite plate.
where : $x = c\cosh\xi\cos\eta$
$y = c\sinh\xi\sin\eta$
$x = (c - a + W/2) + \hat{x}$
$y = \hat{y}$
$\left.\dfrac{d\sigma_{yy}}{dx}\right|_{\xi\to\infty} = \left[\dfrac{-L(P/2)}{W^3/12}\right] = \left[\dfrac{-3}{4}\left(\dfrac{P}{W/2}\right)\left(\dfrac{L}{W/2}\right)\left(\dfrac{1}{W/2}\right)\right]$
$\sigma_{yy}\Big|_{\substack{\hat{x}=0\\|y|\to\infty}} = 0;$ $\sigma_{xx}\Big|_{\xi\to\infty} = 0;$ $\sigma_{xy}\Big|_{\xi\to\infty} = 0$
L
W/2
W/2
a
c
c
x
y
$\hat{x}$
$\hat{y}$

The solution to boundary value problem $(*1a)$ consists of the superposition of problem solutions $(*1aa)$, $(*1ab)$ and $(*1ac)$.

$$\left\{\begin{array}{l} \sigma_{xx(*1a)}\big|_{\xi\to\infty} = 0 \\ \sigma_{xx(*1aa)}\big|_{\xi\to\infty} = 0 \\ \sigma_{xx(*1ab)}\big|_{\xi\to\infty} = S_u \\ \sigma_{xx(*1ac)}\big|_{\xi\to\infty} = -S_d \end{array}\right\} \Rightarrow S_d = S_u$$

$$\left\{\begin{array}{l} \sigma_{yy(*1a)}\Big|_{\substack{\widehat{x}=0 \\ |y|\to\infty}} = 0 \\ \sigma_{yy(*1aa)}\Big|_{\substack{\widehat{x}=0 \\ |y|\to\infty}} = \dfrac{S_b}{c}\left(c+\dfrac{W}{2}-a\right) \\ \sigma_{yy(*1ab)}\Big|_{\substack{\widehat{x}=0 \\ |y|\to\infty}} = S_u \\ \sigma_{yy(*1ac)}\Big|_{\substack{\widehat{x}=0 \\ |y|\to\infty}} = S_d \end{array}\right\} \Rightarrow S_u + S_d = -\frac{S_b}{c}\left(c+\frac{W}{2}-a\right)$$

$$\left\{\begin{array}{l} \dfrac{d\sigma_{yy(*1a)}}{dx}\Big|_{\xi\to\infty} = \left[\dfrac{-L(P/2)}{W^3/12}\right] \\ \dfrac{d\sigma_{yy(*1aa)}}{dx}\Big|_{\xi\to\infty} = \dfrac{S_b}{c} \\ \dfrac{d\sigma_{yy(*1ab)}}{dx}\Big|_{\xi\to\infty} = \dfrac{d\sigma_{yy(*1ac)}}{dx}\Big|_{\xi\to\infty} = 0 \end{array}\right\} \Rightarrow \frac{S_b}{c} = \left[\frac{-L(P/2)}{W^3/12}\right]$$

Summarizing:

$$\frac{S_b}{c} = \left[\frac{-L(P/2)}{W^3/12}\right] = \left[-\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left(\frac{1}{W/2}\right)\right];$$

$$S_d = S_u = -\frac{1}{2}\frac{S_b}{c}\left(c+\frac{W}{2}-a\right) = -\frac{1}{2}\left(c+\frac{W}{2}-a\right)\left[\frac{-L(P/2)}{W^3/12}\right].$$

Let us illustrate the combined stress function resulting from this superposition by considering :

$$\frac{1}{2}(\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1a)} = \frac{1}{2}\left[(\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1aa)} + (\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1ab)} + (\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1ac)}\right] .$$

$$= S_b\left[e^{2\xi_0}\cosh^2\xi_0\cos\eta\Big(\left[\sinh\xi/(\cosh 2\xi-\cos 2\eta)\right]-e^{-\xi}\Big)+\frac{1}{2}\cosh\xi\cos\eta\right]$$

$$+ S_u\left[\sinh 2\xi/(\cosh 2\xi-\cos 2\eta)\right] + S_d e^{2\xi_0}\left[\sinh 2\xi/(\cosh 2\xi-\cos 2\eta)-1\right]$$

Utilizing $S_u = S_d = -\frac{1}{2}\left(c+\frac{W}{2}-a\right)\left[\frac{-L(P/2)}{W^3/12}\right]$ and $S_b = c\left[\frac{-L(P/2)}{W^3/12}\right]$,

$$\frac{\frac{1}{2}(\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1a)}}{\left[\frac{-L(P/2)}{W^3/12}\right]ce^{\xi_0}\cosh\xi_0} = e^{\xi_0}\cosh\xi_0\left[\sinh\xi\cos\eta/(\cosh 2\xi-\cos 2\eta)\right]$$

$$-\frac{1}{c}\left(c+\frac{W}{2}-a\right)\left[\sinh 2\xi/(\cosh 2\xi-\cos 2\eta)\right]$$

$$+\frac{1}{2}\cosh\xi\cos\eta/(e^{\xi_0}\cosh\xi_0) - e^{\xi_0-\xi}\cosh\xi_0\cos\eta$$

$$+\frac{1}{2}\frac{1}{c}\left(c+\frac{W}{2}-a\right)/\cosh^2\xi_0$$

Recall that

$$\lim_{\widetilde{r}\to 0}\left(\frac{\sinh\xi\cos\eta}{\cosh 2\xi-\cos 2\eta}\right)\longrightarrow\frac{1}{2}\sqrt{\pi c}\left(\sqrt{\frac{1}{2\pi r}}\cos\frac{\theta}{2}\right);$$

$$\lim_{\widetilde{r}\to 0}\left(\frac{\sinh 2\xi}{\cosh 2\xi-\cos 2\eta}\right)\longrightarrow\sqrt{\pi c}\left(\sqrt{\frac{1}{2\pi r}}\cos\frac{\theta}{2}\right),$$

where

$\widetilde{r}=\sqrt{(\cosh\xi\cos\eta-1)^2+(\sinh\xi\sin\eta)^2}$; $r=c\widetilde{r}$; $\theta=\arctan\left(\dfrac{\sinh\xi\sin\eta}{\cosh\xi\cos\eta-1}\right)$.

Hence,

$$\lim_{\widetilde{r}\to 0}\left[\frac{1}{2}(\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1a)}\right]\longrightarrow\left(\frac{1}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\right)\left[\left(\frac{L(P/2)}{(W-a)^2/6}\right)\sqrt{\frac{\pi a(W-a)}{W}}\right]$$
$$\times\left[\left(1+(e^{\xi_0}\cosh\xi_0-1)\right)\sqrt{\frac{c}{a}\left(1-\frac{a}{W}\right)}\right]$$
$$\times\left[\frac{a}{W}\left(\frac{c}{a}\left(1-\frac{a}{W}\right)\right)\left(1-(e^{\xi_0}\cosh\xi_0-1)\right)+\left(1-2\frac{a}{W}\right)\left(1-\frac{a}{W}\right)\right]$$

Recall, further, that $\sinh^2\xi_0/\cosh\xi_0=R_c/c$, where R_c is the radius of curvature for the tip of the ellipse $\xi=\xi_0$. Equivalently,

$$\sinh^2\xi_0=\left[\frac{R_c}{c}\sqrt{1+\frac{1}{4}\left(\frac{R_c}{c}\right)^2}+\frac{1}{2}\left(\frac{R_c}{c}\right)^2\right],$$

$$\cosh\xi_0=\left[\sqrt{1+\frac{1}{4}\left(\frac{R_c}{c}\right)^2}+\frac{1}{2}\left(\frac{R_c}{c}\right)\right],$$

and

$$\left[e^{\xi_0}\cosh\xi_0-1\right]=\sinh\xi_0\cosh\xi_0+\sinh^2\xi_0.$$

So,

$$\lim_{r\to 0}\left[\frac{1}{2}(\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1a)}\right] \longrightarrow \left(\frac{1}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\right)\left[\left(\frac{L(P/2)}{(W-a)^2/6}\right)\sqrt{\frac{\pi a(W-a)}{W}}\right]$$
$$\times\left\{\frac{a}{W}\left[\sqrt{\frac{c}{a}\left(1-\frac{a}{W}\right)}\right]^3\left[1-\frac{5}{2}\left(\frac{R_c}{c}\right)^2 f_3\left(\frac{R_c}{c}\right)\right]\right.$$
$$+\sqrt{\frac{c}{a}\left(1-\frac{a}{W}\right)}\left(1-\frac{a}{W}\right)\left[\left(1-2\frac{a}{W}\right)\left[1+\frac{R_c}{c}f_1\left(\frac{R_c}{c}\right)\right]-\frac{a}{W}\left(\frac{R_c}{c}\right)\right]$$
$$\left.+\left(1-\frac{a}{W}\right)^{3/2}\left[\left(1-2\frac{a}{W}\right)\sqrt{\frac{R_c}{a}}-2\frac{a}{W}\left(\frac{R_c}{a}\right)^{3/2}\right]\right\},$$

where

$$f_3\left(\frac{R_c}{c}\right)\equiv 1+\left(\frac{R_c}{c}\right)^{1/2}+\frac{17}{20}\left(\frac{R_c}{c}\right)+\frac{1}{2}\left(\frac{R_c}{c}\right)^{3/2}+\frac{2}{5}\left(\frac{R_c}{c}\right)^2$$
$$+\frac{2}{5}\left(\frac{R_c}{c}\right)^{-1}\left[\sqrt{1+\frac{1}{4}\left(\frac{R_c}{c}\right)^2}-1-\frac{1}{8}\left(\frac{R_c}{c}\right)^2\right]$$
$$+\frac{4}{5}\left(\frac{R_c}{c}\right)^{-1/2}\left[\left(1+\frac{R_c}{c}\sqrt{1+\frac{1}{4}\left(\frac{R_c}{c}\right)^2}+\frac{1}{2}\left(\frac{R_c}{c}\right)^2\right)^{5/4}-1-\frac{5}{4}\frac{R_c}{c}-\frac{5}{8}\left(\frac{R_c}{c}\right)^2\right]$$
$$+\frac{4}{5}\left(\frac{R_c}{c}\right)\left[\sqrt{1+\frac{1}{4}\left(\frac{R_c}{c}\right)^2}-1\right]$$

and

$$f_1\left(\frac{R_c}{c}\right)\equiv 1+\frac{3}{4}\sqrt{\frac{R_c}{c}}+\frac{1}{2}\left(\frac{R_c}{c}\right)+\left[\sqrt{1+\frac{1}{4}\left(\frac{R_c}{c}\right)^2}-1\right]$$
$$+\left(\frac{R_c}{c}\right)^{-1/2}\left[\left[\sqrt{1+\frac{1}{4}\left(\frac{R_c}{c}\right)^2}+\frac{1}{2}\left(\frac{R_c}{c}\right)\right]^{3/2}-1-\frac{3}{4}\left(\frac{R_c}{c}\right)\right]$$

At this point one may choose to establish c/a as a function of a/W and R_c/a by requiring that the stress intensity factor derived from the above expression for problem $(*1a)$ be set equal to the stress intensity factor for problem $(*1)$.

Take the standard definition for the notch stress intensity factor, that is,

$$\lim_{\substack{\sim \\ r\to 0}}\left[\frac{1}{2}(\sigma_{\xi\xi}+\sigma_{\eta\eta})_{(*1a)}\right] \longrightarrow \left(\frac{1}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\right)K_{1(*1a)}.$$

The numerically established formula * for the stress intensity factor in a single edge notched, pure bending specimen may, after minor manipulation, be put in the form (where M is the moment per unit thickness):

$$K_{1(*1)} = \left[\frac{M}{(W-a)^2/6}\sqrt{\frac{\pi a(W-a)}{W}}\right]\left[\frac{2}{\sqrt{\pi}}\left(\frac{1}{1+2a/W}\right)\right] \times \frac{1}{2}\left[1.99-\frac{a}{W}\left(1-\frac{a}{W}\right)\left(2.15-3.93\frac{a}{W}+2.7\left(\frac{a}{W}\right)^2\right)\right]$$

Setting $K_{1(*1a)} = K_{1(*1)}$ leads to a relationship for $\frac{c}{a}$ as a function of $\frac{a}{W}$ and $\frac{R_c}{c}$:

$$\left[\sqrt{\frac{c}{a}\left(1-\frac{a}{W}\right)}\right]^3 C_3 + \left[\sqrt{\frac{c}{a}\left(1-\frac{a}{W}\right)}\right] C_2 + C_1 = 0$$

where

$$C_3 \equiv \frac{a}{W}\left[1-\frac{5}{2}\left(\frac{R_c}{c}\right)^2 f_3\left(\frac{R_c}{c}\right)\right]$$

$$C_2 \equiv \left(1-\frac{a}{W}\right)\left[\left(1-2\frac{a}{W}\right)\left[1+\frac{R_c}{c}f_1\left(\frac{R_c}{c}\right)\right]-\frac{a}{W}\left(\frac{R_c}{a}\right)\right]$$

$$C_1 \equiv -\frac{2}{\sqrt{\pi}}\left(\frac{1}{1+2a/W}\right)\frac{1}{2}\left[1.99-\frac{a}{W}\left(1-\frac{a}{W}\right)\left(2.15-3.93\frac{a}{W}+2.7\left(\frac{a}{W}\right)^2\right)\right] + \left(1-\frac{a}{W}\right)^{3/2}\left[\left(1-2\frac{a}{W}\right)\sqrt{\frac{R_c}{a}}-2\frac{a}{W}\left(\frac{R_c}{a}\right)^{3/2}\right]$$

---

* ASTM Standard E399-81,"Annual Book of ASTM Standards," Part 10, 1981, as quoted in "Deformation and Fracture Mechanics of Engineering Materials, Third Edition," by Richard W. Hertzberg, Appendix B, p. 656.

Special Cases:

$$\#1: \quad \left\{\frac{a}{W} \longrightarrow 0\right\} \Rightarrow \frac{c}{a}\Big|_{a/W=0} = \left[\left(\frac{1}{\sqrt{\pi}}(1.99) - \sqrt{\frac{R_c}{a}}\right) \Big/ \left[1 + \frac{R_c}{c} f_1\left(\frac{R_c}{c}\right)\right]\right]^2 .$$

$$\#2: \quad \left\{\frac{a}{W} \longrightarrow 1\right\} \Rightarrow$$

$$\left[\frac{c}{a}\left(1 - \frac{a}{W}\right)\right]\Big|_{a/W \to 1} \longrightarrow \left[\left(\frac{1}{3\sqrt{\pi}}(1.99)\right) \Big/ \left[1 - \frac{5}{2}\left(\frac{R_c}{c}\right)^2 f_3\left(\frac{R_c}{c}\right)\right]\right]^{2/3}$$

$$\#3: \quad \left\{ \begin{array}{c} \frac{a}{W} \neq 0;\ \frac{R_c}{c} << 1, \text{ i.e. } \left[1 - \frac{5}{2}\left(\frac{R_c}{c}\right)^2 f_3\left(\frac{R_c}{c}\right)\right] \approx 1 \\ \text{and } \left[1 + \frac{R_c}{c} f_1\left(\frac{R_c}{c}\right)\right] \approx 1 \end{array} \right\} \Rightarrow$$

$$\left[\frac{c}{a}\left(1 - \frac{a}{W}\right)\right]\Big|_{\substack{a/W \neq 0 \\ R_c/c << 1}} \approx \left[\sqrt[3]{-\frac{b_0}{2} + \sqrt{\frac{b_0^2}{4} + \frac{a_0^3}{27}}} + \sqrt[3]{-\frac{b_0}{2} - \sqrt{\frac{b_0^2}{4} + \frac{a_0^3}{27}}}\right]^2$$

where

$$a_0 = \left(\frac{a}{W}\right)^{-1}\left(1 - \frac{a}{W}\right)\left[\left(1 - 2\frac{a}{W}\right) - \frac{a}{W}\left(\frac{R_c}{a}\right)\right]$$

and

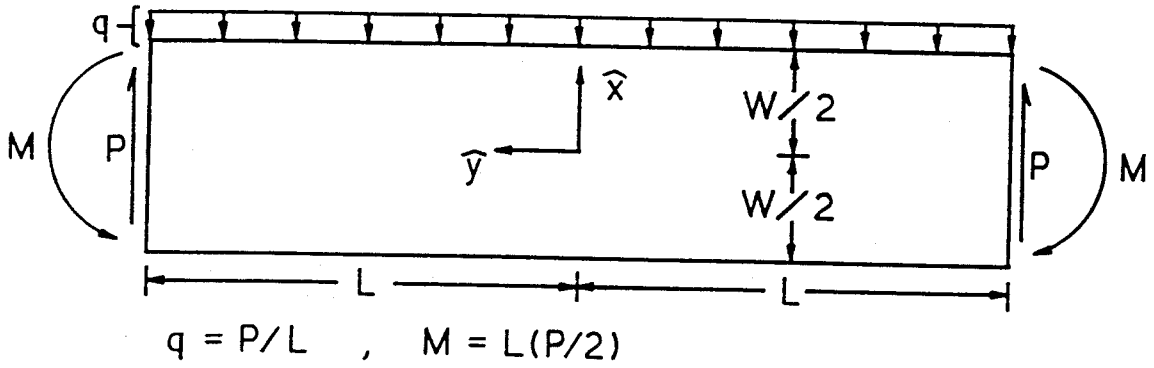
$$b_0 = \left(\frac{a}{W}\right)^{-1}\left\{\left(1 - \frac{a}{W}\right)^{3/2}\left[\left(1 - 2\frac{a}{W}\right)\sqrt{\frac{R_c}{a}} - 2\frac{a}{W}\left(\frac{R_c}{a}\right)^{3/2}\right]\right.$$

$$\left. - \frac{2}{\sqrt{\pi}}\left(\frac{1}{1 + 2a/W}\right)\frac{1}{2}\left[1.99 - \frac{a}{W}\left(1 - \frac{a}{W}\right)\left(2.15 - 3.93\frac{a}{W} + 2.7\left(\frac{a}{W}\right)^2\right)\right]\right\}$$

$$\#4: \quad \left\{\frac{R_c}{a} = \frac{R_c}{c} = 0;\ \frac{a}{W} = \frac{2}{5}\right\} \Rightarrow \frac{c}{a}\Big|_{\substack{R_c/a=0 \\ a/W=2/5}} = 1.7362$$

$$\#5: \quad \left\{\frac{R_c}{a} << 1;\ \frac{a}{W} = \frac{2}{5},\ \frac{R_c}{a} = \frac{1}{625}\right\} \Rightarrow \frac{c}{a}\Big|_{\substack{R_c/a=1/625 \\ a/W=2/5}} \approx 1.728$$

Problem ($*1a$) will satisfy the boundary conditions of problem ($*1$) only approximately. The difference in boundary conditions will establish the boundary tractions for problem ($*1b$). From the solution of problem ($*1a$), one will recall that the available free parameters were chosen so that the stress intensity factor associated with ($*1a$) would be the same as for ($*1$), and the far–field bending stress distribution in ($*1a$) would match the bending stress distribution at the ends of the bend specimen under loading condition ($*1$). In this case, the boundary tractions associated with problem ($*1b$) consist primarily of a higher order traction distribution along the long edges of the specimen which does not induce any stress intensity factor. Because problem ($*1b$) is of secondary importance, it is omitted for the sake of brevity.

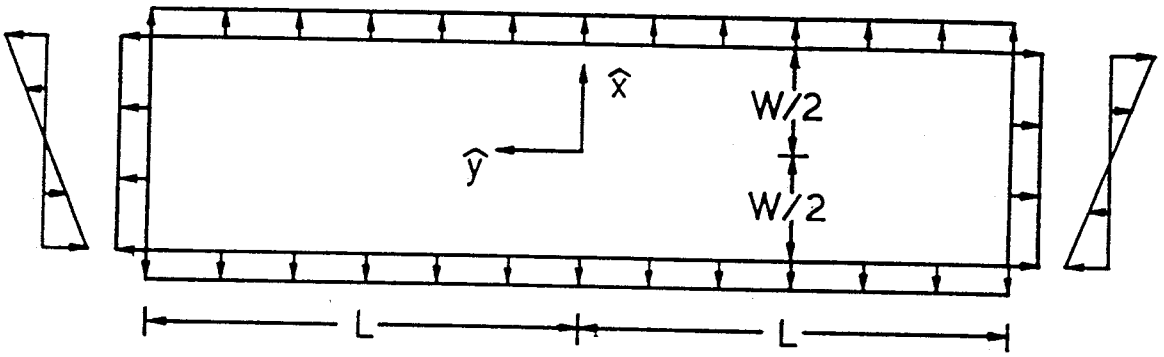


(*2):

$$\sigma_{\widetilde{xy}} = \frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left(\frac{W/2}{L}\right)\left[1-\left(\frac{\hat{x}}{W/2}\right)^2\right]\left(\frac{\hat{y}}{L}\right)$$

$$\sigma_{\widetilde{xx}} = -\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[\left(\frac{W/2}{L}\right)^2\right]\left[\frac{2}{3}+\left(\frac{\hat{x}}{W/2}\right)-\frac{1}{3}\left(\frac{\hat{x}}{W/2}\right)^3\right]$$

$$\sigma_{\widetilde{yy}} = \frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left(\frac{\hat{x}}{W/2}\right)\left[\left(\frac{\hat{y}}{L}\right)^2-\frac{2}{3}\left(\frac{W/2}{L}\right)^2\left[\left(\frac{\hat{x}}{W/2}\right)^2-\frac{3}{5}\right]\right]$$



$(*2a)$:

$$\sigma_{\widetilde{xy}} = 0$$

$$\sigma_{\widetilde{yy}(*2a)} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]\left[C_{1(*2a)} + \left(\frac{\hat{x}}{W/2}\right)C_{2(*2a)}\right]$$

$$C_{1(*2a)} = \left[\frac{1}{120}\right]\left[-\left(\frac{a}{W/2}\right)^3 + 7\left(\frac{a}{W/2}\right)^2 - 15\left(\frac{a}{W/2}\right) + 10\right]$$

$$C_{2(*2a)} = \left[\frac{1}{120}\right]\left[3\left(\frac{3}{2}\left(\frac{a}{W/2}\right)^2 - 5\left(\frac{a}{W/2}\right) + 4\right)\right]$$

$$\sigma_{\widetilde{xx}(*2a)} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]C_{3(*2a)}$$

$$C_{3(*2a)} = \left[\frac{1}{24}\right]\left[\left(\frac{a}{W/2}\right)^2\left(\frac{3}{2} - \frac{1}{2}\left(\frac{a}{W/2}\right)\right)\right]$$

The constants $C_{1(*2a)}$, $C_{2(*2a)}$ and $C_{3(*2a)}$ were chosen to satisfy the following three conditions:

$$\int_{-W/2}^{-(W/2-a)} \left(\sigma_{\widehat{yy}(*2)}\Big|_{\widehat{y}=0} + \sigma_{\widehat{yy}(*2a)}\Big|_{\widehat{y}=0} \right) d\widehat{x} = 0$$

$$\int_{-W/2}^{-(W/2-a)} \left(\widehat{x} + \frac{W}{2} - \frac{a}{2} \right) \left(\sigma_{\widehat{yy}(*2)}\Big|_{\widehat{y}=0} + \sigma_{\widehat{yy}(*2a)}\Big|_{\widehat{y}=0} \right) d\widehat{x} = 0$$

$$\left[\sigma_{\widehat{xx}(*2)}\Big|_{\substack{\widehat{x}=-(W/2-a) \\ \widehat{y}=0}} + \sigma_{\widehat{xx}(*2a)}\Big|_{\substack{\widehat{x}=-(W/2-a) \\ \widehat{y}=0}} \right] = 0$$

The variation of $C_{1(*2a)}$, $C_{2(*2a)}$ and $C_{3(*2a)}$ with differing values of a/W is illustrated in Table A.1.

Table [A.1]

$\breve{a}$	$-\breve{a}^3 + 7\breve{a}^2 - 15\breve{a} + 10$	$3(\frac{3}{2}\breve{a}^2 - 5\breve{a} + 4)$	$\breve{a}^2(\frac{3}{2} - \frac{1}{2}\breve{a})$
0.0	10.000	12.000	0.0000
0.1	8.569	10.545	0.0145
0.2	7.272	9.180	0.0560
0.3	6.103	7.905	0.1215
0.4	5.056	6.720	0.2080
0.5	4.125	5.625	0.3125
0.6	3.304	4.620	0.4320
0.7	2.587	3.705	0.5635
0.8	1.968	2.880	0.7040
0.9	1.441	2.145	0.8505
1.0	1.000	1.500	1.0000

The higher order distribution of residue tractions on the notch surface may be examined by consideration of:

$$\left\{\left[\sigma_{\widehat{yy}(*2)}\Big|_{\widehat{y}=0}+\sigma_{\widehat{yy}(*2a)}\Big|_{\widehat{y}=0}\right]\Big/\left(\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right)\right\}=$$
$$\left[\frac{1}{96}\breve{a}^2\right]\left(\frac{1}{2}\left[-\breve{a}\left(s^3-\tfrac{3}{5}s\right)+3(2-\breve{a})\left(s^2-\tfrac{1}{3}\right)\right]\right)$$

where

$\breve{a}\equiv\dfrac{a}{W}$; $\quad s\equiv\dfrac{\widehat{x}+(W/2)-(a/2)}{a/2}$; $\quad -1\le s\le 1$ is the domain of the notch.

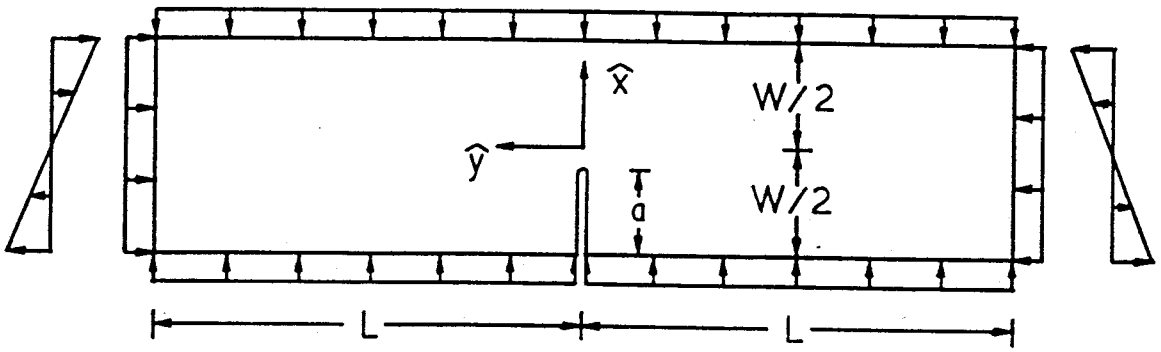
$$\int_{-1}^{1}\left[-\breve{a}\left(s^3-\tfrac{3}{5}s\right)+3(2-\breve{a})\left(s^2-\tfrac{1}{3}\right)\right]ds=0$$
$$\int_{-1}^{1}\left[-\breve{a}\left(s^3-\tfrac{3}{5}s\right)+3(2-\breve{a})\left(s^2-\tfrac{1}{3}\right)\right]sds=0$$

As an example case,

$$\left[\left[\frac{1}{96}\breve{a}^2\right]\left(\frac{1}{2}\left[-\breve{a}\left(s^3-\tfrac{3}{5}s\right)+3(2-\breve{a})\left(s^2-\tfrac{1}{3}\right)\right]\right)\right]_{\breve{a}=4/5}=$$
$$\frac{1}{1000}\left[-\tfrac{8}{3}\left(s^3-\tfrac{3}{5}s\right)+12\left(s^2-\tfrac{1}{3}\right)\right]$$

Table [A.2]

s	$[-\frac{8}{3}(s^3-\frac{3}{5}s)]$	$[12(s^2-\frac{1}{3})]$	$[-\frac{8}{3}(s^3-\frac{3}{5}s)+12(s^2-\frac{1}{3})]$
-1	16/15	8	136/15
$-3/4$	$-3/40$	11/4	107/40
$-1/2$	$-7/15$	-1	$-22/15$
$-1/4$	$-43/120$	$-13/4$	$-433/120$
0	0	-4	-4
1/4	43/120	$-13/4$	$-347/120$
1/2	7/15	-1	$-8/15$
3/4	3/40	11/4	113/40
1	$-16/15$	8	104/15



Boundary Condition ($*2an$):

$$\sigma_{\widetilde{yy}}\Big|_{\widehat{y}=\pm L} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]\left[-C_{1(*2a)} - \left(\frac{\widehat{x}}{W/2}\right)C_{2(*2a)}\right],$$

$$\sigma_{\widetilde{xx}}\Big|_{\widehat{x}=\pm W/2} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]\left(-C_{3(*2a)}\right),$$

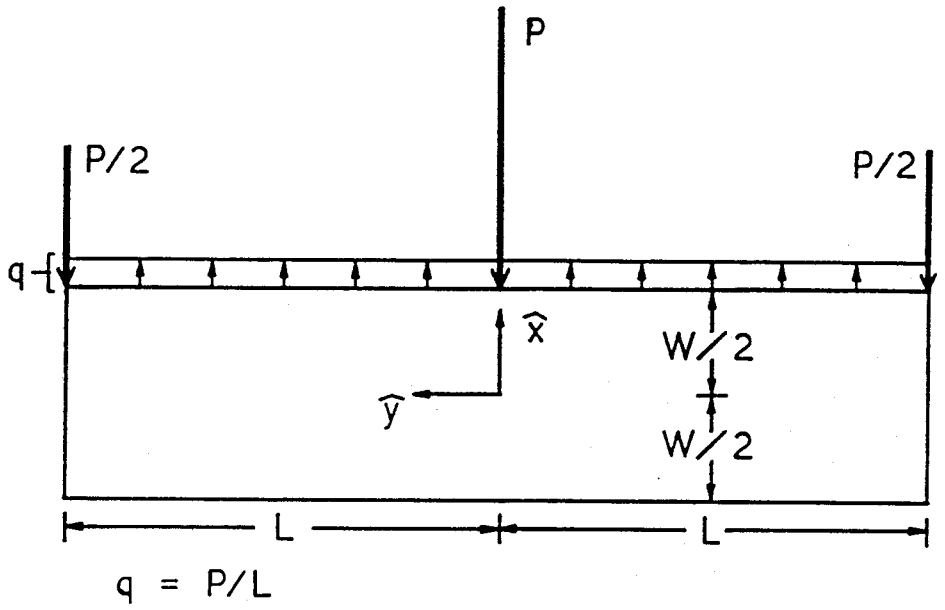
otherwise traction free. Recall,

$$C_{1(*2a)} = \left[\frac{1}{120}\right]\left[-\left(\frac{a}{W/2}\right)^3 + 7\left(\frac{a}{W/2}\right)^2 - 15\left(\frac{a}{W/2}\right) + 10\right]$$

$$C_{2(*2a)} = \left[\frac{1}{120}\right]\left[3\left(\frac{3}{2}\left(\frac{a}{W/2}\right)^2 - 5\left(\frac{a}{W/2}\right) + 4\right)\right]$$

$$C_{3(*2a)} = \left[\frac{1}{24}\right]\left[\left(\frac{a}{W/2}\right)^2\left(\frac{3}{2} - \frac{1}{2}\left(\frac{a}{W/2}\right)\right)\right]$$

Problem ($*2an$) is solved in the manner of problem ($*1$).



Boundary Condition (*3):

For $\hat{x} = -W/2$, $\sigma_{\widehat{xy}} = 0$ and $\sigma_{\widehat{xx}} = 0$.

For $\hat{y} = \pm L$, $\int_{-W/2}^{W/2} \sigma_{\widehat{xy}} d\hat{x} = 0$; $\int_{-W/2}^{W/2} \sigma_{\widehat{yy}} d\hat{x} = 0$; $\int_{-W/2}^{W/2} \sigma_{\widehat{yy}} \hat{x} d\hat{x} = 0$.

For $\hat{x} = W/2$,

$$\hat{y} = L, \qquad \sigma_{\widehat{xx}} = (P/2)\Big(-\delta(\hat{y} - L)\Big)$$

$$L > \hat{y} > 0, \qquad \sigma_{\widehat{xx}} = (P/L)$$

$$\hat{y} = 0, \qquad \sigma_{\widehat{xx}} = P\Big(-\delta(\hat{y})\Big)$$

$$0 > \hat{y} > -L, \qquad \sigma_{\widehat{xx}} = (P/L)$$

$$\hat{y} = -L, \qquad \sigma_{\widehat{xx}} = (P/2)\Big(-\delta(\hat{y} + L)\Big)$$

$P > 0$ when the point load is compressive.

Solution (*3):

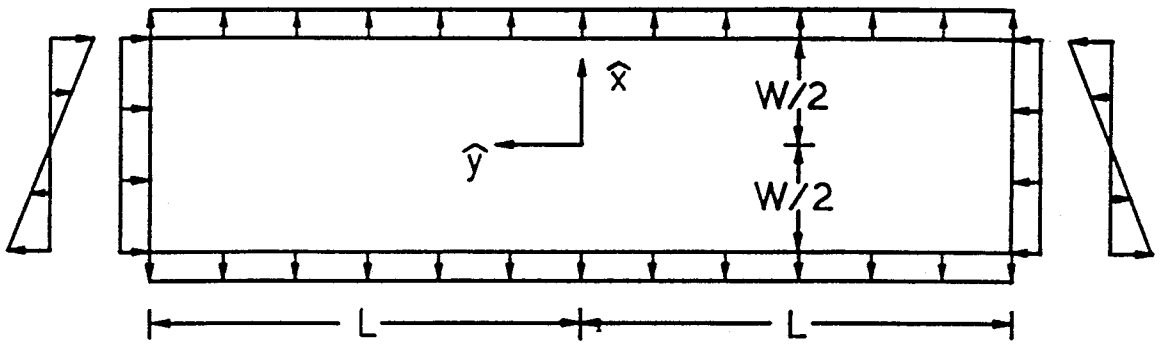
Define $\quad \breve{x} \equiv \left[\widehat{x}/(W/2)\right],$

and let $\quad \alpha = n2\pi\left(\frac{W/2}{L}\right) = n\frac{\pi}{2}4\left(\frac{W/2}{L}\right), \quad n$ a positive integer.

$$\sigma_{\widetilde{xx}} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right] \times \frac{1}{6}\sum_{n=1}^{\infty}\left(\cos n2\pi\frac{\widehat{y}}{L}\right)\left\{\frac{-\left[(\alpha\cosh\alpha+\sinh\alpha)\cosh\alpha\breve{x}-\alpha\breve{x}\sinh\alpha\breve{x}\sinh\alpha\right]}{[\sinh 2\alpha+2\alpha]} \frac{-\left[(\alpha\sinh\alpha+\cosh\alpha)\sinh\alpha\breve{x}-\alpha\breve{x}\cosh\alpha\breve{x}\cosh\alpha\right]}{[\sinh 2\alpha-2\alpha]}\right\}$$

$$\sigma_{\widetilde{xy}} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right] \times \frac{1}{6}\sum_{n=1}^{\infty}\left(\sin n2\pi\frac{\widehat{y}}{L}\right)\left\{\frac{\left[\alpha\cosh\alpha\sinh\alpha\breve{x}-\alpha\breve{x}\cosh\alpha\breve{x}\sinh\alpha\right]}{[\sinh 2\alpha+2\alpha]} \frac{+\left[\alpha\sinh\alpha\cosh\alpha\breve{x}-\alpha\breve{x}\sinh\alpha\breve{x}\cosh\alpha\right]}{[\sinh 2\alpha-2\alpha]}\right\}.$$

$$\sigma_{\widetilde{yy}} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right] \times \sum_{n=1}^{\infty}\left\{\frac{1}{\alpha^2}\frac{1}{4}\breve{x} + \frac{1}{6}\left(\cos n2\pi\frac{\widehat{y}}{L}\right)\left\{\frac{\left[(\alpha\cosh\alpha-\sinh\alpha)\cosh\alpha\breve{x}-\alpha\breve{x}\sinh\alpha\breve{x}\sinh\alpha\right]}{[\sinh 2\alpha+2\alpha]} + \frac{\left[(\alpha\sinh\alpha-\cosh\alpha)\sinh\alpha\breve{x}-\alpha\breve{x}\cosh\alpha\breve{x}\cosh\alpha\right]}{[\sinh 2\alpha-2\alpha]}\right\}\right\}$$



$(*3a)$:

$$\sigma_{\widehat{xy}} = 0$$

$$\sigma_{\widehat{yy}(*3a)} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]\left[-C_{1(*3a)} - \left(\frac{\hat{x}}{W/2}\right)C_{2(*3a)}\right],$$

$$\sigma_{\widehat{xx}(*3a)} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]\left[C_{3(*3a)}\right].$$

$$C_{3(*3a)} = \frac{1}{6}\sum_{n=1}^{\infty}\left\{\frac{\sinh 2\alpha[\alpha(2-\breve{a})\cosh\alpha\breve{a} + \sinh\alpha\breve{a}]}{\sinh^2 2\alpha - (2\alpha)^2} + \frac{2\alpha[-\alpha\breve{a}\cosh(\alpha(2-\breve{a})) - \sinh(\alpha(2-\breve{a}))]}{\sinh^2 2\alpha - (2\alpha)^2}\right\}$$

$$C_{1(*3a)} = \sum_{n=1}^{\infty}\Big((2-\breve{a})\alpha\Big)\left\{\frac{\frac{1}{3}\left(\frac{1}{2-\breve{a}}\right)\sinh(\alpha(2-\breve{a}))}{\sinh^2 2\alpha - (2\alpha)^2} - \frac{\left(\alpha\cosh 2\alpha + \left(\frac{1}{2} - \frac{1}{6}\right)\sinh 2\alpha\right)\frac{1}{\alpha\breve{a}}\sinh\alpha\breve{a}}{\sinh^2 2\alpha - (2\alpha)^2} + \frac{[2\alpha\cosh 2\alpha + (1+\alpha^2\breve{a})\sinh 2\alpha]\left(\frac{1}{\alpha\breve{a}}\right)^2\left(\cosh\alpha\breve{a} - \frac{1}{\alpha\breve{a}}\sinh\alpha\breve{a}\right)}{\sinh^2 2\alpha - (2\alpha)^2}\right\}$$

$$C_{2(*3a)} = \sum_{n=1}^{\infty} \left\{ \frac{1}{\alpha^2}\frac{1}{4} + 2\alpha \left(\frac{-\left(\alpha\cosh 2\alpha + \frac{1}{2}\sinh 2\alpha\right)\frac{1}{\alpha\breve{a}}\sinh\alpha\breve{a}}{\sinh^2 2\alpha - (2\alpha)^2} \right.\right.$$

$$\left.\left. + \frac{[2\alpha\cosh 2\alpha + (1+\alpha^2\breve{a})\sinh 2\alpha]\left(\frac{1}{\alpha\breve{a}}\right)^2\left(\cosh\alpha\breve{a} - \frac{1}{\alpha\breve{a}}\sinh\alpha\breve{a}\right)}{\sinh^2 2\alpha - (2\alpha)^2} \right) \right\}$$

where

$$\breve{a} = \frac{a}{W}; \qquad \alpha = n\frac{\pi}{2}4\left(\frac{W/2}{L}\right);$$

$$\sum_{n=1}^{\infty}\frac{1}{\alpha^2}\frac{1}{4} = \left[\pi 4\left(\frac{W/2}{L}\right)\right]^{-2}\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{1}{6}\left[4\left(\frac{W/2}{L}\right)\right]^{-2}.$$

As before, the constants $C_{1(*3a)}$, $C_{2(*3a)}$ and $C_{3(*3a)}$ were chosen to satisfy the following three conditions:

$$\int_{-W/2}^{-(W/2-a)} \left(\sigma_{\widehat{yy}(*3)}\Big|_{\widehat{y}=0} + \sigma_{\widehat{yy}(*3a)}\Big|_{\widehat{y}=0}\right) d\widehat{x} = 0$$

$$\int_{-W/2}^{-(W/2-a)} \left(\widehat{x} + \frac{W}{2} - \frac{a}{2}\right)\left(\sigma_{\widehat{yy}(*3)}\Big|_{\widehat{y}=0} + \sigma_{\widehat{yy}(*3a)}\Big|_{\widehat{y}=0}\right) d\widehat{x} = 0$$

$$\left[\sigma_{\widehat{xx}(*3)}\Big|_{\substack{\widehat{x}=-(W/2-a)\\ \widehat{y}=0}} + \sigma_{\widehat{xx}(*3a)}\Big|_{\substack{\widehat{x}=-(W/2-a)\\ \widehat{y}=0}}\right] = 0$$

$$\frac{\sigma_{\widehat{xx}(*3)}\Big|_{\substack{\widehat{x}=-(W/2-a)\\ \widehat{y}=0}}}{\left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]} = -C_{3(*3a)}$$

$$\frac{\frac{1}{a}\int_{-W/2}^{-(W/2-a)} \left(\sigma_{\widehat{yy}(*3)}\Big|_{\widehat{y}=0}\right) d\widehat{x}}{\left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]} = \left[\sum_{n=1}^{\infty} -\frac{1}{\alpha^2}\frac{1}{4}\left(1 - \frac{\breve{a}}{2}\right)\right] + C_{4(*3a)}$$

$$\frac{\frac{1}{a^3/12}\int_{-W/2}^{-(W/2-a)} \left(\widehat{x} + \frac{W}{2} - \frac{a}{2}\right)\left(\sigma_{\widehat{yy}(*3)}\Big|_{\widehat{y}=0}\right) d\widehat{x}}{\left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]} =$$

$$\left[\frac{1}{W/2}\right]\left[\left(\sum_{n=1}^{\infty} -\frac{1}{\alpha^2}\frac{1}{4}(-1)\right) + C_{5(*3a)} + C_{6(*3a)}\right]$$

$$\left\{ \frac{\frac{1}{a}\int_{-W/2}^{-(W/2-a)} \left(\sigma_{\widehat{yy}(*3)}\Big|_{\widehat{y}=0}\right) d\widehat{x}}{\left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]} \right.$$

$$\left. + \frac{\left[\frac{1}{a^3/12}\int_{-W/2}^{-(W/2-a)} \left(\widehat{x}+\frac{W}{2}-\frac{a}{2}\right)\left(\sigma_{\widehat{yy}(*3)}\Big|_{\widehat{y}=0}\right) d\widehat{x}\right]\left(\widehat{x}+\frac{W}{2}-\frac{a}{2}\right)}{\left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]} \right\} =$$

$$\left\{ C_{4(*3a)} + \left(1-\frac{\breve{a}}{2}\right)\left[C_{5(*3a)}+C_{6(*3a)}\right] + \breve{x}\left[\left(\sum_{n=1}^{\infty}\frac{1}{\alpha^2}\frac{1}{4}\right) + C_{5(*3a)} + C_{6(*3a)}\right] \right\}$$

$$C_{1(*3a)} = C_{4(*3a)} + \left(1-\frac{\breve{a}}{2}\right)\left[C_{5(*3a)}+C_{6(*3a)}\right]$$

$$C_{2(*3a)} = \left(\sum_{n=1}^{\infty}\frac{1}{\alpha^2}\frac{1}{4}\right) + C_{5(*3a)} + C_{6(*3a)}$$

$$C_{4(*3a)} = \sum_{n=1}^{\infty}\left\{ \frac{\frac{\alpha}{6}\left[2\sinh\left(\alpha(2-\breve{a})\right) + (2-\breve{a})(\sinh 2\alpha)\frac{1}{\alpha\breve{a}}\sinh\alpha\breve{a}\right]}{\sinh^2 2\alpha - (2\alpha)^2} \right\}$$

$$C_{5(*3a)} = \sum_{n=1}^{\infty}\left\{ \frac{2\alpha\left[-\left(\alpha\cosh 2\alpha + \frac{1}{2}\sinh 2\alpha\right)\frac{1}{\alpha\breve{a}}\sinh\alpha\breve{a}\right]}{\sinh^2 2\alpha - (2\alpha)^2} \right\}$$

$$C_{6(*3a)} =$$

$$\sum_{n=1}^{\infty}\left\{ \frac{2\alpha\left[2\alpha\cosh 2\alpha + (1+\alpha^2\breve{a})\sinh 2\alpha\right]\left(\frac{1}{\alpha\breve{a}}\right)^2\left(\cosh\alpha\breve{a} - \frac{1}{\alpha\breve{a}}\sinh\alpha\breve{a}\right)}{\sinh^2 2\alpha - (2\alpha)^2} \right\}$$

$$C_{1(*3a)}\Big|_{\substack{a/(W/2)=4/5\\(W/2)/L=1/4}} \approx 0.01283 \qquad C_{4(*3a)}\Big|_{\substack{a/(W/2)=4/5\\(W/2)/L=1/4}} \approx 0.058464$$

$$C_{2(*3a)}\Big|_{\substack{a/(W/2)=4/5\\(W/2)/L=1/4}} \approx 0.09060 \qquad C_{5(*3a)}\Big|_{\substack{a/(W/2)=4/5\\(W/2)/L=1/4}} \approx -1.04907$$

$$C_{3(*3a)}\Big|_{\substack{a/(W/2)=4/5\\(W/2)/L=1/4}} \approx 0.0715 \qquad C_{6(*3a)}\Big|_{\substack{a/(W/2)=4/5\\(W/2)/L=1/4}} \approx 0.973003$$

$$[-C_{1(*3a)} - \breve{x}C_{2(*3a)}] = -\left[\breve{x}\left(\sum_{n=1}^{\infty}\frac{1}{\alpha^2}\frac{1}{4}\right) + C_{4(*3a)} + \left(\frac{\widehat{x}+(W/2)-(a/2)}{a/2}\right)\frac{1}{2}\breve{a}\left(C_{5(*3a)} + C_{6(*3a)}\right)\right]$$

$$s = \left(\frac{\widehat{x}+(W/2)-(a/2)}{a/2}\right); \qquad -1 \le s \le 1 \text{ is the domain of the notch.}$$

The higher order distribution of residue tractions on the notch surface may be examined by consideration of:

$$\left\{\left(\sigma_{\widehat{yy}(*3)}\Big|_{\widehat{y}=0} + \sigma_{\widehat{yy}(*3a)}\Big|_{\widehat{y}=0}\right)\Big/\left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]\right\} \equiv C;$$

$$\left\{\left(\sigma_{\widehat{yy}(*3)}\Big|_{\widehat{y}=0}\Big/\left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]\right) - \sum_{n=1}^{\infty}\frac{1}{\alpha^2}\frac{1}{4}\breve{x}\right\} \equiv B;$$

$$C = B - C_{4(*3a)} - \left(\frac{\widehat{x}+(W/2)-(a/2)}{a/2}\right)\frac{1}{2}\frac{a}{W}\left(C_{5(*3a)} + C_{6(*3a)}\right).$$

As a specific example,

$$B\Big|_{(W/2)/L=1/4} \equiv B_*.$$

$$C\Big|_{\substack{a/(W/2)=4/5\\(W/2)/L=1/4}} \equiv C_*; \qquad s\Big|_{a/(W/2)=4/5} \equiv s_*.$$

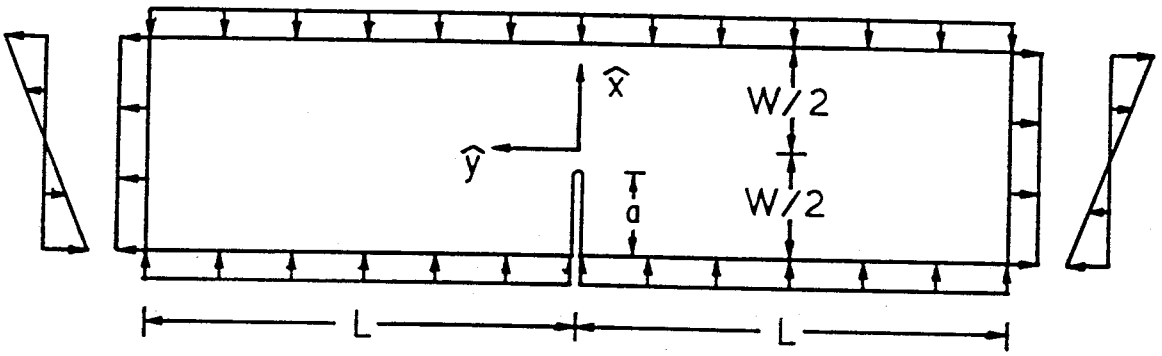
$$\left[\frac{1}{2}\frac{a}{W}\left(C_{5(*3a)} + C_{6(*3a)}\right)\right]_{\substack{a/(W/2)=4/5\\(W/2)/L=1/4}} \approx -0.03043$$

Table [A.3]

$\breve{x}$	B_*	C_*	s_*
-1	0.106287	0.0174	-1
$-9/10$	0.0859	0.0046	$-3/4$
$-4/5$	0.0706	-0.0031	$-1/2$
$-7/10$	0.0594	-0.0067	$-1/4$
$-3/5$	0.0515	-0.0070	0
$-1/2$	0.0460	-0.0049	$1/4$
$-2/5$	0.0425	-0.0007	$1/2$
$-3/10$	0.0407	0.0051	$3/4$
$-1/5$	0.0400	0.0120	1
$-1/10$	0.0404		
0	0.041521		

Table [A.4]

s_*	$[C_* - (0.022)({s_*}^2 - \frac{1}{3})]$	$[C_* - (0.022)({s_*}^2 - \frac{1}{3})$ $+(0.0068)({s_*}^3 - \frac{3}{5}s_*)]$
-1	0.0027	0.0000
$-3/4$	-0.0004	-0.0002
$-1/2$	-0.0013	-0.0001
$-1/4$	-0.0007	0.0002
0	0.0003	0.0003
$1/4$	0.0011	0.0002
$1/2$	0.0011	-0.0001
$3/4$	0.0001	-0.0001
1	-0.0027	0.0000



Boundary Condition ($*3an$):

$$\sigma_{\widetilde{yy}}|_{\hat{y}=\pm L} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]\left[C_{1(*3a)} + \left(\frac{\hat{x}}{W/2}\right)C_{2(*3a)}\right],$$

$$\sigma_{\widetilde{xx}}|_{\hat{x}=\pm W/2} = \left[\frac{3}{4}\left(\frac{P}{W/2}\right)\left(\frac{L}{W/2}\right)\left[16\left(\frac{W/2}{L}\right)^2\right]\right]\left[-C_{3(*3a)}\right].$$

otherwise traction free.

Problem ($*3an$) is solved in the manner of problem ($*1$).

Appendix B: 3-D Elastic Out-of-plane Displacement Approximation

Nakamura and Parks [10] present numerical data on the three–dimensional elastic field near the tip of a crack in a plate of thickness t. They have provided out-of-plane displacements for different values of Poisson's ratio and angular (θ) dependence in accordance with the following two tables:

$$\left[-u_3\big|_{x_3=t/2}\right] \Big/ \left[\frac{\nu}{\mathrm{E}} K_I \sqrt{t} \cos\frac{\theta}{2} \Big/ (\sqrt{2\pi}\sqrt{r/t})\right]$$

On $\theta = 0°$

r/t	$\nu = 0.15$	0.30	0.40	0.499
1.0000E-06	2.4127E-03	2.3415E-03	2.3176E-03	2.2935E-03
5.0000E-03	0.1683	0.1660	0.1647	0.1630
1.0000E-02	0.2374	0.2341	0.2321	0.2296
1.6305E-02	0.3013	0.2972	0.2947	0.2915
2.4255E-02	0.3644	0.3596	0.3566	0.3526
3.4285E-02	0.4285	0.4231	0.4195	0.4149
4.6930E-02	0.4944	0.4883	0.4843	0.4790
6.2880E-02	0.5624	0.5557	0.5511	0.5452
8.2995E-02	0.6321	0.6248	0.6197	0.6130
0.1084	0.7027	0.6948	0.6892	0.6818
0.1403	0.7726	0.7642	0.7581	0.7500
0.1807	0.8396	0.8307	0.8242	0.8155
0.2316	0.9007	0.8914	0.8845	0.8754
0.2957	0.9523	0.9428	0.9357	0.9262
0.3766	0.9910	0.9815	0.9744	0.9649
0.4787	1.014	1.005	0.9982	0.9890
0.6073	1.023	1.014	1.007	0.9989
0.7696	1.020	1.012	1.006	0.9978
0.9743	1.012	1.004	0.9991	0.9925
1.232	1.005	0.9982	0.9936	0.9878
1.558	1.003	0.9966	0.9929	0.9884
1.968	1.003	0.9966	0.9932	0.9893
2.486	1.005	1.000	0.9979	0.9957
3.139	0.9923	0.9917	0.9911	0.9902

$$\left[-u_3\big|_{x_3=t/2}\right] \Big/ \left[\frac{\nu}{\mathrm{E}} K_I \sqrt{t} \cos\frac{\theta}{2} \Big/ \left(\sqrt{2\pi}\sqrt{r/t}\right)\right]$$

For $\nu = 0.30$

r/t	$\theta = 0°$	40°	80°	120°	140°
1.0000E-06	2.3415E-03				
5.0000E-03	0.1660	0.1745	0.2073	0.3050	0.4378
1.0000E-02	0.2341	0.2451	0.2874	0.4144	0.5883
1.6305E-02	0.2972	0.3099	0.3589	0.5075	0.7130
2.4255E-02	0.3596	0.3735	0.4274	0.5926	0.8234
3.4285E-02	0.4231	0.4376	0.4947	0.6722	0.9225
4.6930E-02	0.4883	0.5032	0.5619	0.7468	1.011
6.2880E-02	0.5557	0.5704	0.6289	0.8163	1.087
8.2995E-02	0.6248	0.6389	0.6953	0.8794	1.150
0.1084	0.6948	0.7078	0.7603	0.9350	1.197
0.1403	0.7642	0.7756	0.8222	0.9811	1.224
0.1807	0.8307	0.8402	0.8793	1.016	1.232
0.2316	0.8914	0.8987	0.9292	1.039	1.218
0.2957	0.9428	0.9479	0.9694	1.050	1.186
0.3766	0.9815	0.9847	0.9978	1.048	1.140
0.4787	1.005	1.007	1.014	1.039	1.089
0.6073	1.014	1.015	1.018	1.026	1.043
0.7696	1.012	1.012	1.014	1.014	1.013
0.9743	1.004	1.005	1.007	1.006	1.001
1.232	0.9982	0.9992	1.002	1.004	1.000
1.558	0.9966	0.9976	1.000	1.004	1.004
1.968	0.9966	0.9974	0.9998	1.003	1.005
2.486	1.000	1.001	1.003	1.006	1.008
3.139	0.9917	0.9919	0.9922	0.9928	0.9938

This numerical data may, alternatively, be expressed in the form:

$$\left[-u_3\big|_{x_3=t/2}\right] \Big/ \left[\frac{\nu}{\mathrm{E}} K_I \sqrt{t}\right]$$

On $\theta = 0°$

r/t	$\nu = 0.15$	0.30	0.40	0.499
1×10^{-6}	0.9625	0.9341	0.9246	0.9150
5×10^{-3}	0.9495	0.9366	0.9292	0.9196
1×10^{-2}	0.9471	0.9339	0.9259	0.9160
0.016305	0.9413	0.9285	0.9207	0.9107
0.024255	0.9334	0.9211	0.9135	0.9032
0.034285	0.9232	0.9116	0.9038	0.8939
0.046930	0.9105	0.8992	0.8919	0.8821
0.062880	0.8947	0.8841	0.8768	0.8674
0.082995	0.8753	0.8652	0.8582	0.8489
0.1084	0.8515	0.8419	0.8351	0.8261
0.1403	0.8229	0.8139	0.8074	0.7988
0.1807	0.7880	0.7796	0.7735	0.7653
0.2316	0.7467	0.7389	0.7332	0.7257
0.2957	0.6986	0.6917	0.6865	0.6795
0.3766	0.6442	0.6381	0.6334	0.6273
0.4787	0.5847	0.5795	0.5756	0.5703
0.6073	0.5237	0.5191	0.5155	0.5114
0.7696	0.4639	0.4602	0.4575	0.4538
0.9743	0.4090	0.4058	0.4038	0.4011
1.232	0.3612	0.3588	0.3571	0.3550
1.558	0.3206	0.3185	0.3173	0.3159
1.968	0.2852	0.2834	0.2824	0.2813
2.486	0.2543	0.2530	0.2525	0.2519
3.139	0.2234	0.2233	0.2232	0.2230

and

$$\left[-u_3\big|_{x_3=t/2}\right] \Big/ \left[\frac{\nu}{\mathrm{E}} K_I \sqrt{t}\right]$$

For $\nu = 0.30$

r/t	$\theta = 0°$	40°	80°	120°	140°
1×10^{-6}	0.9341				
5×10^{-3}	0.9366	0.9251	0.8959	0.8604	0.8448
1×10^{-2}	0.9339	0.9188	0.8783	0.8266	0.8027
0.016305	0.9285	0.9098	0.8590	0.7928	0.7619
0.024255	0.9211	0.8991	0.8387	0.7590	0.7214
0.034285	0.9116	0.8860	0.8165	0.7241	0.6798
0.046930	0.8992	0.8708	0.7927	0.6876	0.6368
0.062880	0.8841	0.8527	0.7665	0.6493	0.5915
0.082995	0.8652	0.8314	0.7376	0.6089	0.5447
0.1084	0.8419	0.8059	0.7057	0.5665	0.4961
0.1403	0.8139	0.7763	0.6708	0.5225	0.4459
0.1807	0.7796	0.7410	0.6322	0.4768	0.3955
0.2316	0.7389	0.7001	0.5901	0.4307	0.3453
0.2957	0.6917	0.6535	0.5448	0.3852	0.2976
0.3766	0.6381	0.6015	0.4969	0.3406	0.2535
0.4787	0.5795	0.5456	0.4479	0.2995	0.2148
0.6073	0.5191	0.4883	0.3992	0.2626	0.1826
0.7696	0.4602	0.4325	0.3532	0.2306	0.1576
0.9743	0.4058	0.3817	0.3118	0.2033	0.1384
1.232	0.3588	0.3375	0.2759	0.1804	0.1229
1.558	0.3185	0.2996	0.2448	0.1604	0.1098
1.968	0.2834	0.2665	0.2178	0.1426	0.0977
2.486	0.2530	0.2380	0.1944	0.1273	0.0872
3.139	0.2233	0.2099	0.1711	0.1118	0.0765

One may fit this numerical data from Nakamura and Parks to an accuracy which is well within the accuracy of the data itself with the following type of formula:

$$\frac{-u_3|_{x_3=t/2}}{\frac{\nu}{\mathrm{E}}K_I\sqrt{t}} \approx f_a\Big(\frac{r}{t},\theta\Big)\Big[1+f_b\Big(\nu,\frac{r}{t},\theta\Big)\Big]$$

where f_a and f_b are chosen so that

$$\frac{-u_3|_{x_3=t/2}}{\frac{\nu}{\mathrm{E}}K_I\sqrt{t}}\Big|_{\nu=3/10} \equiv f_a\Big(\frac{r}{t},\theta\Big); \qquad f_b\Big(\nu,\frac{r}{t},\theta\Big)\Big|_{\nu=3/10} = 0.$$

Let

$$f_a\Big(\frac{r}{t},\theta\Big) = \left\{\left[\Big(1-e^{-c_1\sqrt{2\pi}\sqrt{r/t}\,[1+\pi r/t]}\Big)\frac{1}{\sqrt{2\pi}\sqrt{r/t}}\right] f_0\Big(\frac{r}{t}\Big) \times\left[\cos\frac{\theta}{2}+f_1\Big(\frac{r}{t},\theta\Big)\right]\right\}$$

with

$$f_0\Big(\frac{r}{t}\Big) \approx \Big(1-c_2e^{-f_2(r/t)}\Big)\Big(1+c_4e^{-c_6(r/t-c_5)^2}\Big)$$

where

$$-\widehat{u}_0 \equiv -\left[\Big(u_3\big|_{\substack{r=0\\x_3=t/2}}\Big)\Big/\Big(\frac{\nu}{\mathrm{E}}K_I\sqrt{t}\Big)\right]_{\nu=3/10}; \qquad c_2 \equiv 1-\frac{1}{c_1}\big[-\widehat{u}_0\big];$$

$$\big[-\widehat{u}_0\big] \approx \frac{187}{200}; \qquad c_1 \approx \frac{\pi-1}{2};$$

$$c_4 \approx \frac{33}{2000}; \qquad c_5 \approx \frac{3}{5}; \qquad c_6 \approx 10\,.$$

$$f_2\Big(\frac{r}{t}\Big) = \Big(\frac{1-c_2}{c_2}\Big)\left[\frac{1}{2}c_1\sqrt{2\pi}\sqrt{r/t}+\left[-1+\Big(\frac{1}{6}+\frac{1}{4}\Big(\frac{1-c_2}{c_2}\Big)\Big)c_1^2\right]\pi\frac{r}{t}+c_7\frac{r}{t}+f_3\Big(\frac{r}{t}\Big)\right];$$

$$c_7 \approx 2: \qquad f_3\Big(\frac{r}{t}\Big) \approx \left[70\Big(\frac{r}{t}\Big)^2+\Big(\frac{177}{8}\frac{r}{t}\Big)^7\right].$$

$$f_1\left(\frac{r}{t},\theta\right) = \left(1-\cos\frac{\theta}{2}\right)e^{-g_1(r/t)\,g_2(\theta,r/t)}$$

$$g_1\left(\frac{r}{t}\right) = \frac{117}{25}\left(\frac{r}{t}\right)^{3/5}\left(1+\frac{9}{50}e^{-[4(r/t-1/2)]^4}\right)$$

$$g_{1_2}\left(\theta,\frac{r}{t}\right) \approx \left[1-\frac{2}{33}\left(|\theta|-\frac{2\pi}{9}\right)|\theta|\left(\frac{9}{10}-\frac{r}{t}\right)\right]$$

$$f_b\left(\nu,\frac{r}{t},\theta\right) = \left\{\left[\frac{4}{159}\left(\frac{10}{3}\left(\frac{3}{10}-\nu\right)\right)-\frac{2}{55}\left(\frac{10}{3}\left(\frac{3}{10}-\nu\right)\right)^6\right]e^{-\frac{1}{2}r/t}\right\} \times\left(1+f_c\left(\nu,\frac{r}{t},\theta\right)\right)$$

where $f_c(\nu, r/t, \theta)$ is an as yet to be determined function.

The quality of the fit may be illustrated by subtracting the formula from the tables of numerical data: observe that the formula represents the data to within the accuracy of the data.

$$\left[\frac{-u_3|_{x_3=t/2}}{\frac{\nu}{E}K_I\sqrt{t}}\right]_{\substack{\text{numerical}\\ \text{data}}} - \left[\frac{-u_3|_{x_3=t/2}}{\frac{\nu}{E}K_I\sqrt{t}}\right]_{\text{formula}}$$

For $\nu = 0.3$

r/t	$\theta = 0°$	40°	80°	120°	140°
1×10^{-6}	-0.0013				
5×10^{-3}	0.0000	-0.0015	-0.0038	-0.0054	-0.0061
1×10^{-2}	0.0000	0.0007	-0.0023	-0.0045	-0.0062
0.016305	-0.0001	-0.0005	-0.0016	-0.0040	-0.0059
0.024255	0.0000	-0.0001	-0.0007	-0.0028	-0.0045
0.034285	0.0000	-0.0002	-0.0004	-0.0017	-0.0031
0.046930	0.0019	0.0020	0.0020	0.0009	0.0000
0.062880	0.0029	0.0028	0.0030	0.0022	0.0015
0.082995	0.0018	0.0019	0.0020	0.0015	0.0015
0.1084	-0.0004	-0.0004	-0.0004	-0.0004	0.0004
0.1403	-0.0026	-0.0026	-0.0028	-0.0028	-0.0017
0.1807	-0.0042	-0.0042	-0.0043	-0.0042	-0.0021
0.2316	-0.0042	-0.0040	-0.0037	-0.0025	0.0006
0.2957	-0.0025	-0.0022	-0.0014	0.0013	0.0053
0.3766	-0.0005	-0.0004	-0.0001	0.0012	0.0049
0.4787	0.0003	0.0003	-0.0001	-0.0010	0.0006
0.6073	-0.0001	-0.0001	-0.0007	-0.0024	-0.0029
0.7696	0.0000	-0.0004	-0.0011	-0.0034	-0.0051
0.9743	0.0000	0.0001	0.0000	-0.0015	-0.0028
1.232	-0.0007	-0.0005	0.0001	0.0000	-0.0008
1.558	-0.0011	-0.0008	-0.0002	0.0004	0.0003
1.968	-0.0010	-0.0007	-0.0001	0.0004	0.0004
2.486	0.000	0.0002	0.0006	0.0008	0.0007
3.139	-0.0019	-0.0017	-0.0014	-0.0008	-0.0005

$$\left[\frac{-u_3|_{x_3=t/2}}{\frac{\nu}{E}K_I\sqrt{t}}\right]_{\substack{\text{numerical}\\ \text{data}}}$$

$$-\left[\frac{-u_3|_{x_3=t/2}}{\frac{\nu}{E}K_I\sqrt{t}}\right]_{\substack{\nu=0.3\\ \text{numerical}\\ \text{data}}}\left[1+\left(\frac{4}{159}\left[\frac{10}{3}\left(\frac{3}{10}-\nu\right)\right]-\frac{2}{55}\left[\frac{10}{3}\left(\frac{3}{10}-\nu\right)\right]^6\right)e^{-\frac{1}{2}r/t}\right]$$

For $\theta = 0°$

r/t	$\nu = 0.15$	0.30	0.40	0.499
1×10^{-6}	0.0172	0	-0.0016	-0.0006
5×10^{-3}	0.0017	0	0.0005	0.0015
1×10^{-2}	0.0020	0	-0.0001	0.0005
0.016305	0.0017	0	0.0000	0.0005
0.024255	0.0014	0	0.0001	0.0001
0.034285	0.0008	0	-0.0002	0.0001
0.046930	0.0007	0	0.0001	0.0003
0.062880	0.0003	0	-0.0001	0.0003
0.082995	0.0001	0	0.0000	0.0002
0.1084	0.0000	0	-0.0001	0.0000
0.1403	-0.0001	0	-0.0001	-0.0001
0.1807	-0.0002	0	-0.0001	-0.0002
0.2316	-0.0001	0	-0.0001	-0.0002
0.2957	-0.0003	0	-0.0002	-0.0004
0.3766	-0.0003	0	-0.0002	-0.0003
0.4787	-0.0003	0	0.0000	-0.0002
0.6073	0.0000	0	-0.0004	-0.0001
0.7696	-0.0001	0	-0.0001	-0.0002
0.9743	0.0002	0	0.0001	0.0002
1.232	0.0001	0	-0.0001	0.0000
1.558	0.0003	0	0.0000	0.0003
1.968	0.0005	0	-0.0001	0.0000
2.4862	0.0004	0	0.0001	0.0003
3.139	-0.0005	0	0.0003	0.0006

Appendix C: Self-similar Normalization of Out-of-plane Surface Displacements for a Through-cracked Elasto-plastic Plate

Data for the out-of-plane surface displacements generated from a finite element method analysis by Schultheisz [13] of a notched (with a small tip radius) 3-point-bend fracture specimen of an elastic-plastic material were manipulated to explore the degree to which this surface displacement would exhibit some form of self-similar growth with increasing load. The basic description of the specimen and of the FEM analysis are as follows:

3-point-bend loading geometry; span (the distance between the centerlines of the two support indenters) to width ratio of 4; plate thickness to width ratio of 1/7.6; notch length to specimen width ratio of 0.4; notch shape consisting of straight and parallel sides with a semi-circular notch tip and a straight notch tip front through the thickness of the plate; notch tip radius to plate thickness ratio of 3/200; plate thickness of 10 mm; specimen material description emulating a 4340 steel of ductile heat treatment; Young's modulus of 192 GPa; Poisson's ratio of 0.28; uniaxial yield stress of 960 MPa; J_2 (second invariant of the deviatoric stress) plastic flow theory description; power law hardening plasticity with a hardening coefficient of 20; a small strain, incremental plasticity formulation; a finite element mesh with 10 elements thru the half-thickness of the plate, parabolically distributed, with the thinnest layer (at the traction free surface of the plate) having a thickness dimension of 1/20 of the plate half-thickness; elements which, in-plane, are made everywhere nearly square in shape by meshing largely coincident with cylindrical coordinate lines with mesh lines spaced roughly every 15 deg and with element size inversely proportional to the radial distance from the notch tip center; a resulting element aspect ratio for those elements surrounding the semicircular notch tip and adjacent to the traction free surface of the plate of roughly 1 x 1 x 6, where the long dimension is in the thickness dimension.

In considering the possibility of any self-similarity in the developing deformation field surrounding the notch tip, it must be recognized that there are aspects of

this specimen geometry which will introduce additional length scales that can have a dominating effect on the development of the deformation field with increasing load. First, the notch tip geometry—its shape and absolute size—has a primary effect on the character of the plastic deformation when the 3-point-bend specimen is only lightly loaded. Second, the change in the nature of the deformation field with a change in load level becomes quite pronounced as the extent of the shear lip zone nears that of the half thickness of the plate. However, if the initial notch tip were sufficiently small and the plate were sufficiently thick, then one might expect that these two length scales would lose their influence on the plastic deformation field expanding outward with increasing load from the "vertex" of what, locally, is effectively a quarter infinite crack where the notch intersects the traction free surface of the plate. In this case, one might expect the local strain field to develop self-similarity in polar coordinates centered at the "vertex."

Unfortunately, the out-of-plane surface displacement is an integration of the strain field which depends rather more critically on the overall behavior of the plate throughout the thickness. While the measurement of in-plane displacements on the traction free surfaces of the plate lead directly to the local in-plane strain field—thus making them readily interpretable but only locally important—the measurement of the out-of-plane displacements on the surface of the fracture specimens yield a function whose relationship to the internal strain state of the specimen is more indirect. For this reason, it is not immediately intuitively obvious how the out-of-plane surface displacement field might be representable by a "scaleable" reduced function. On the other hand, it is by no means precluded that such a "scaleable" representation might exist when the out-of-plane displacements have been properly normalized. The following approximate normalization is considered here:

$$w_s(r,\theta,P) \approx P w_{se}(r,\theta)[1+\alpha(P)f(r\beta(P),\theta)]$$

where (r,θ) are in-plane cylindrical coordinates;

P is the load variable;

$w_s(r,\theta,P) \equiv$ the out-of-plane surface displacement field associated with the elastic-plastic material description;

$Pw_{se}(r,\theta) \equiv$ the out-of-plane surface displacement field which would have been exhibited if the fracture specimen had remained linearly elastic at higher loads;

$\alpha(P)$ is some function of P;

$\beta(P)$ is some function of P;

$\alpha(P)f(r\beta(P),\theta)$ is the hypothetical self-similar function.

It is to be noted that $Pw_{se}(r,\theta)$ may be readily approximated using the functions elaborated in Appendix A and Appendix B.

In order to explore the nature of our hypothetical self-similar function the out-of-plane surface displacement data from the finite element analysis has been put into the form

$$(w_s(r,\theta,P) - Pw_{se}(r,\theta))/Pw_{se}(r,\theta)$$

This normalized data has been plotted in Figures (C.1) through (C.4) as contour lines of constant value on 4 plots with load (P) values of 18.5 kN, 35 kN, 52.3 kN, and 63.5 kN, respectively. Note that the coordinate axes have different constants of proportionality for each of the different load values and that the increment between displayed contour lines has a different value for each of the loads.

For this specimen geometry and material, the forward extent of the plastic zone has a dimension greater than the plate thickness for loads above approximately 65 kN. Also for loads of approximately 15 kN or less, the plastic zone extends beyond the notch tip by an amount which is less than the characterizing radius of the notch tip.

In Figures (C.5), (C.6), and (C.7) there are contour plots wherein the contour lines from the 35 kN data (dashed lines) are superimposed upon the contour lines for 18.5 kN, 52.3 kN and 63.5 kN, each set of contour lines being in a scaled coordinate frame appropriate for comparison of the degree of similarity in the pattern of contour lines for different load values. As would be expected, the degree of dissimilarity is

greater for the comparison with the lowest load, 18.5 kN, where the notch size effect becomes important and with the highest load, 63.5 kN, where the plastic zone size has become equal to the plate thickness.

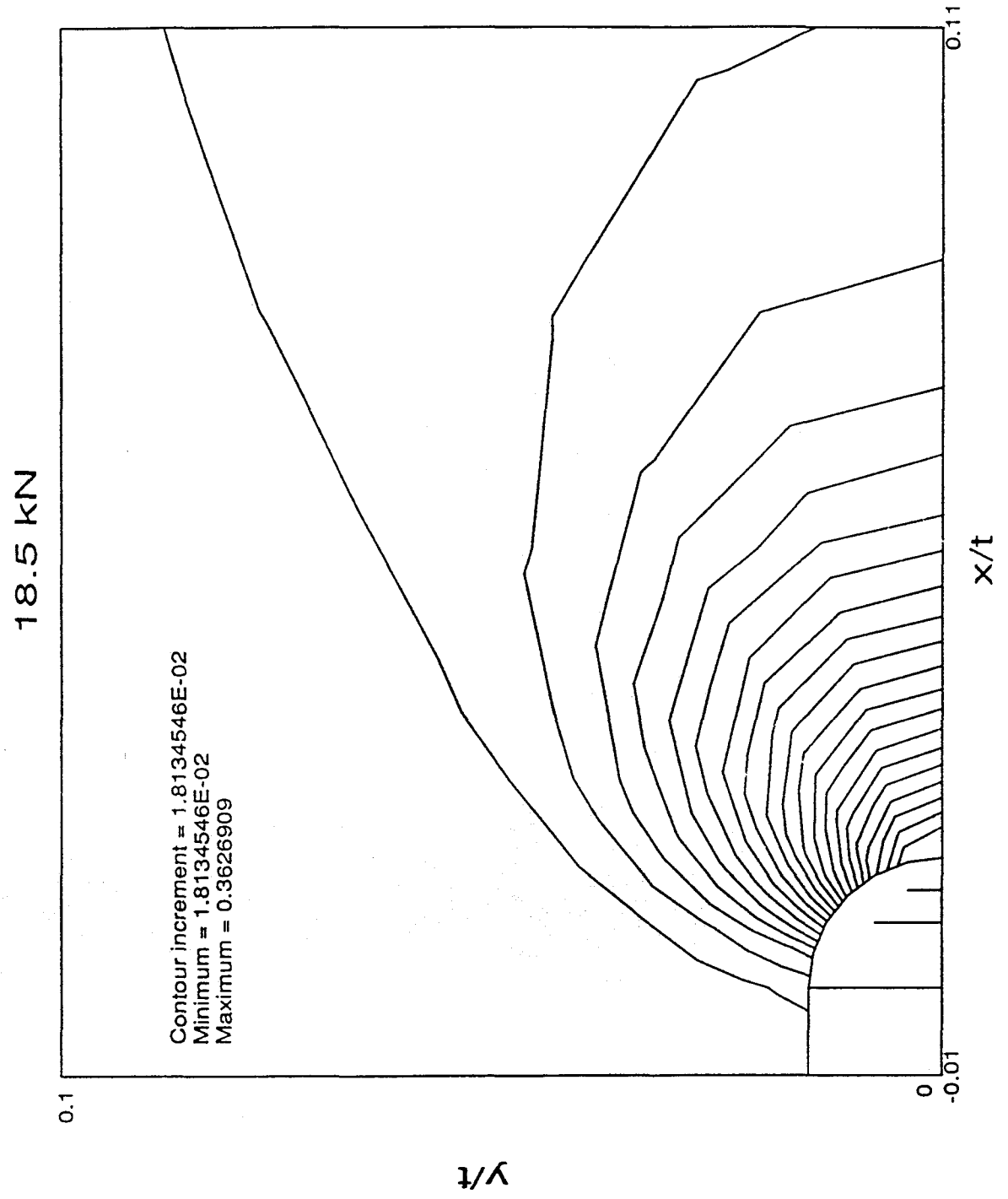

18.5 kN
Contour increment = 1.8134546E-02
Minimum = 1.8134546E-02
Maximum = 0.3626909
0.1
0
-0.01
0.11
x/t
y/t

Figure (C.1)

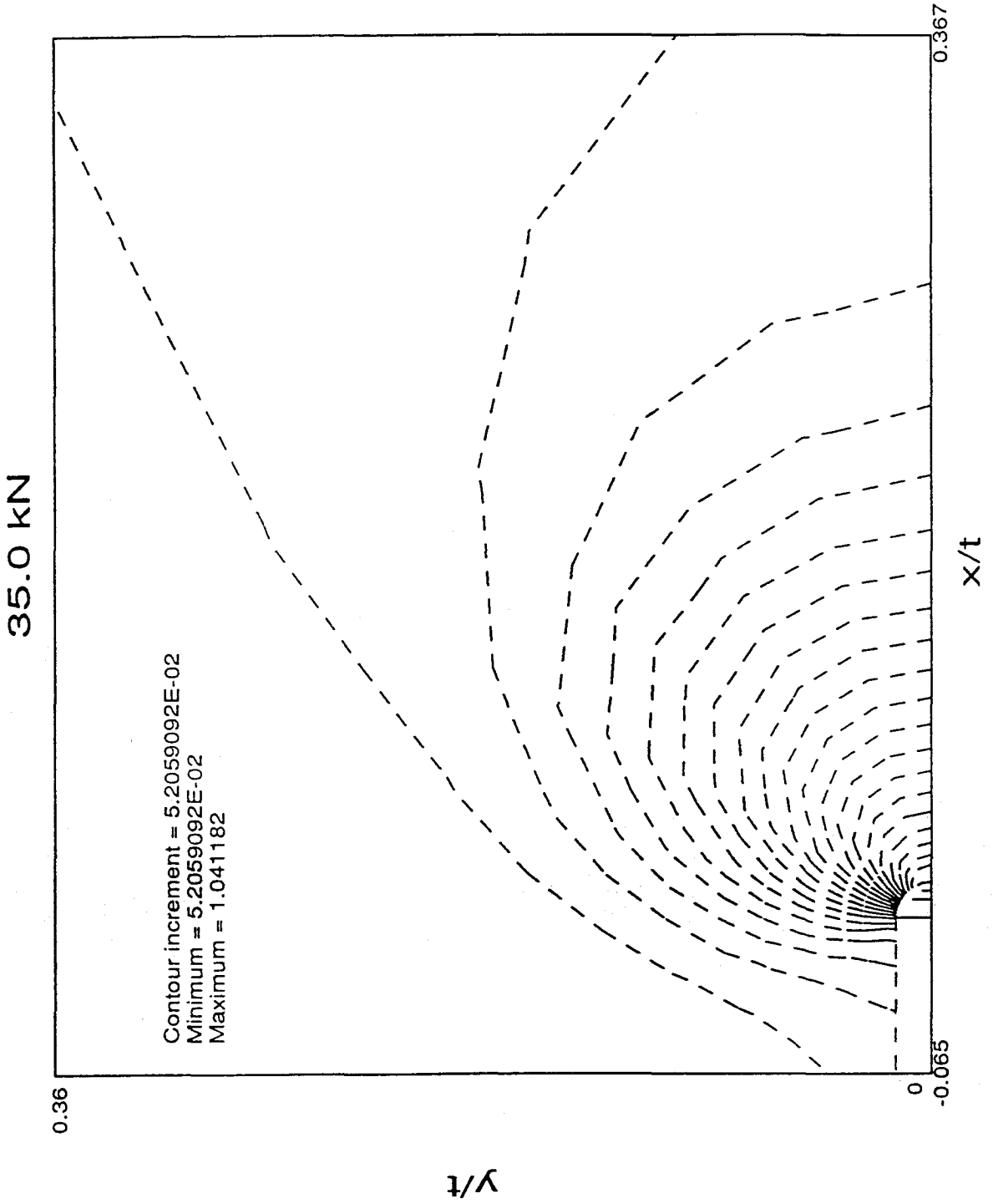

35.0 kN
Contour increment = 5.2059092E-02
Minimum = 5.2059092E-02
Maximum = 1.041182
0.36
0
-0.065
0.367
x/t
y/t

Figure (C.2)

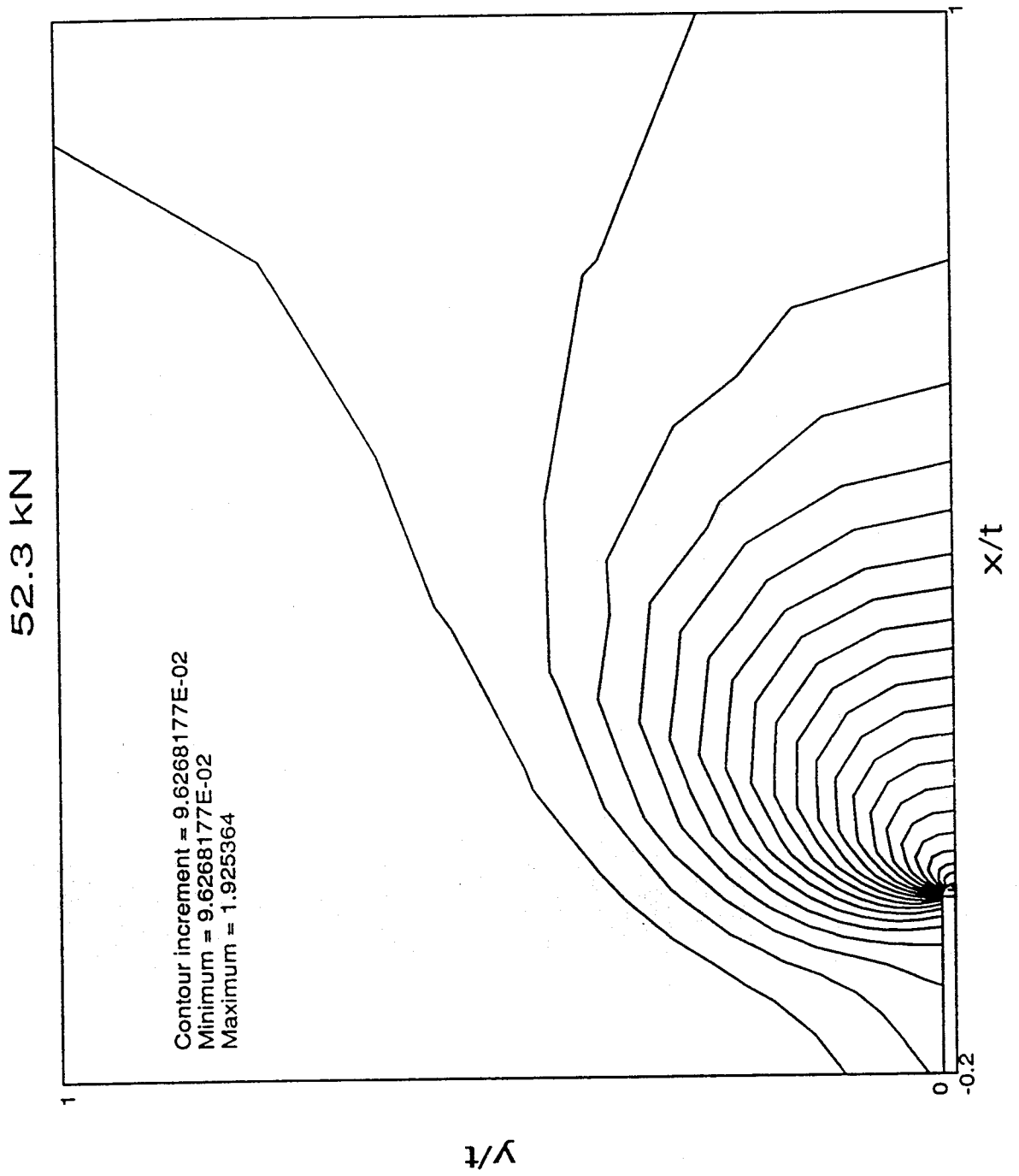

52.3 kN
Contour increment = 9.6268177E-02
Minimum = 9.62681177E-02
Maximum = 1.925364
y/t
x/t
1
0
-0.2
1

Figure (C.3)

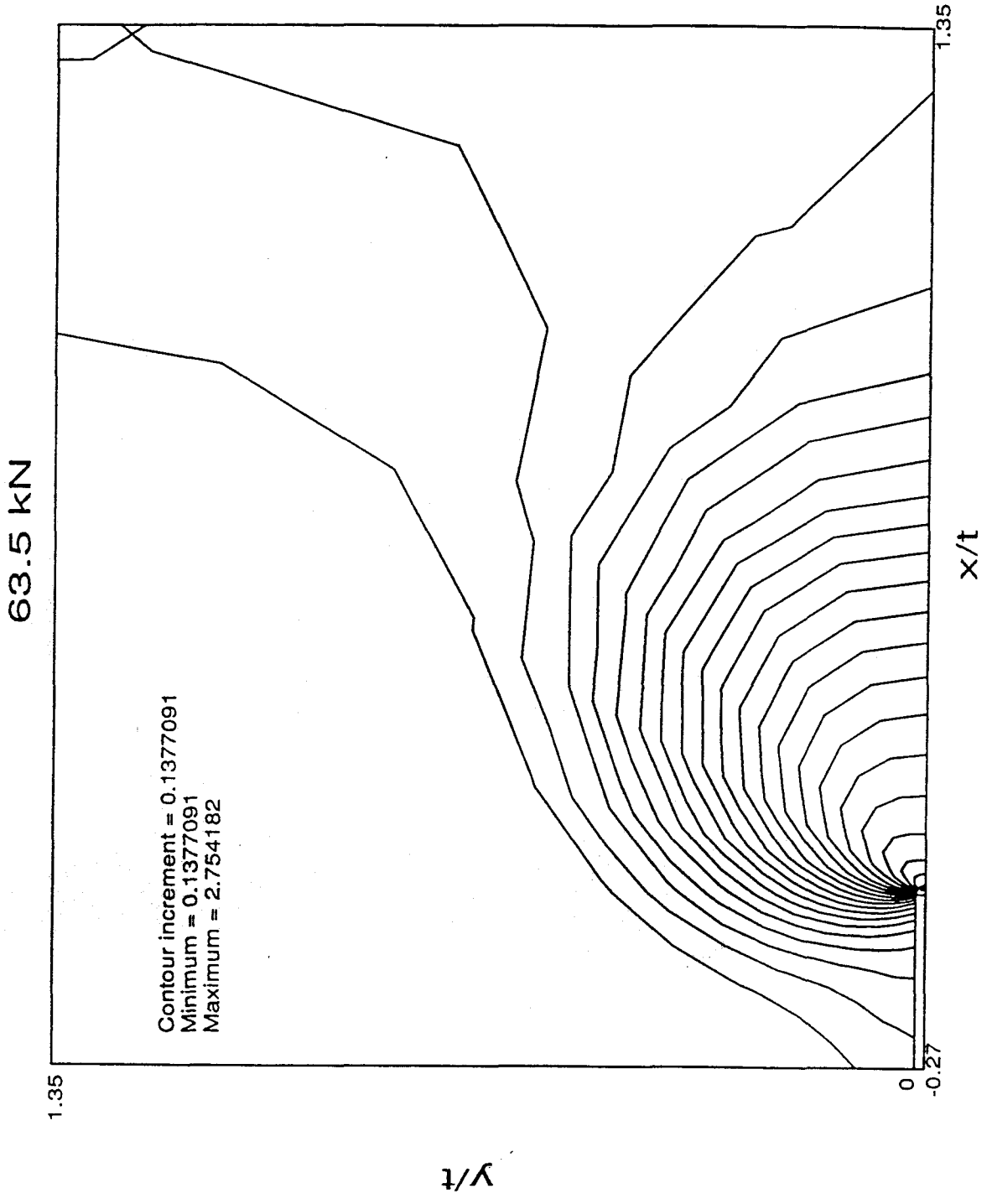

63.5 kN
Contour increment = 0.1377091
Minimum = 0.1377091
Maximum = 2.754182
1.35
1.35
0
-0.27
x/t
y/t

Figure (C.4)

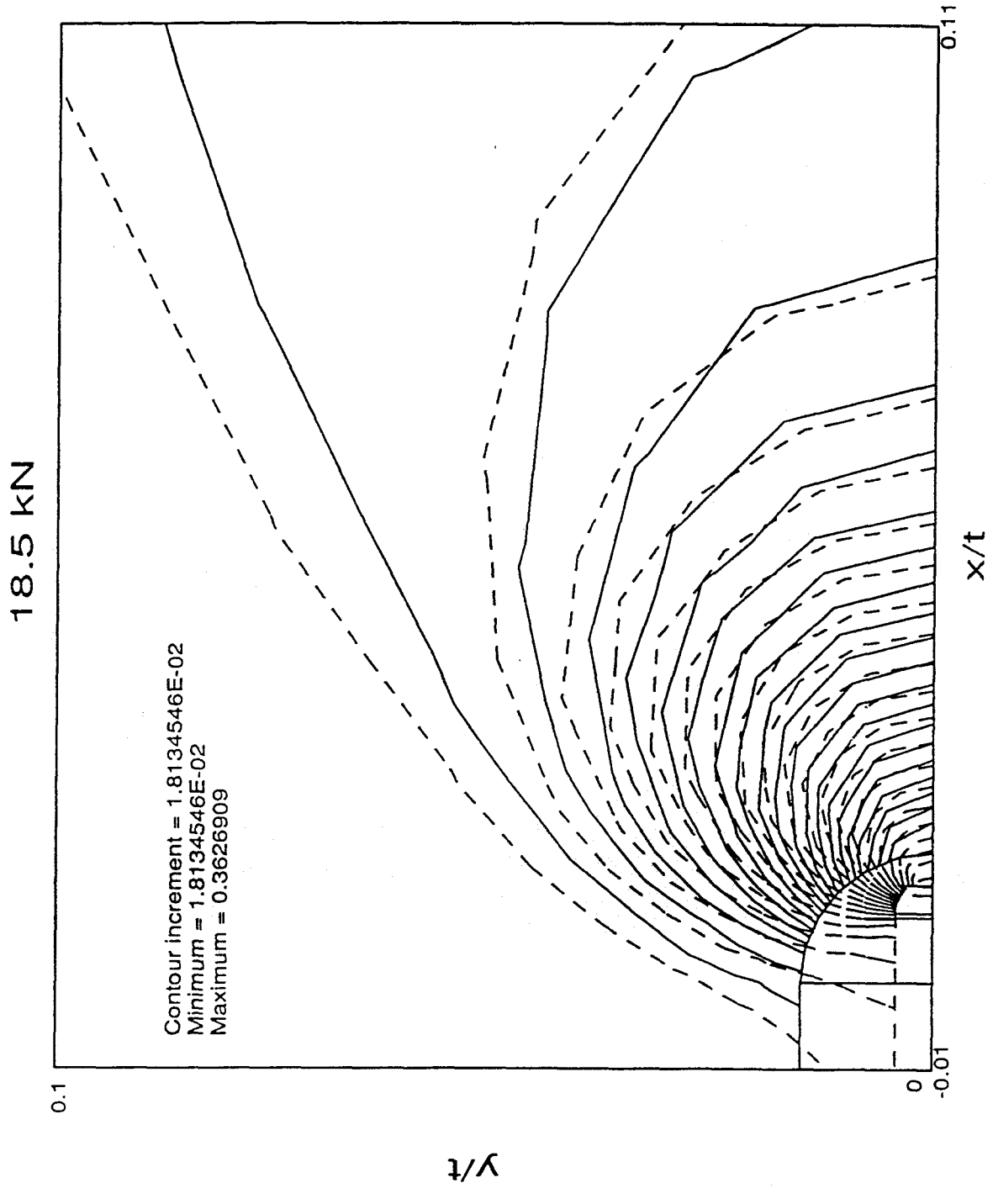

18.5 kN
Contour increment = 1.8134546E-02
Minimum = 1.8134546E-02
Maximum = 0.3626909
0.1
0
-0.01
0.11
x/t
y/t

Figure (C.5)

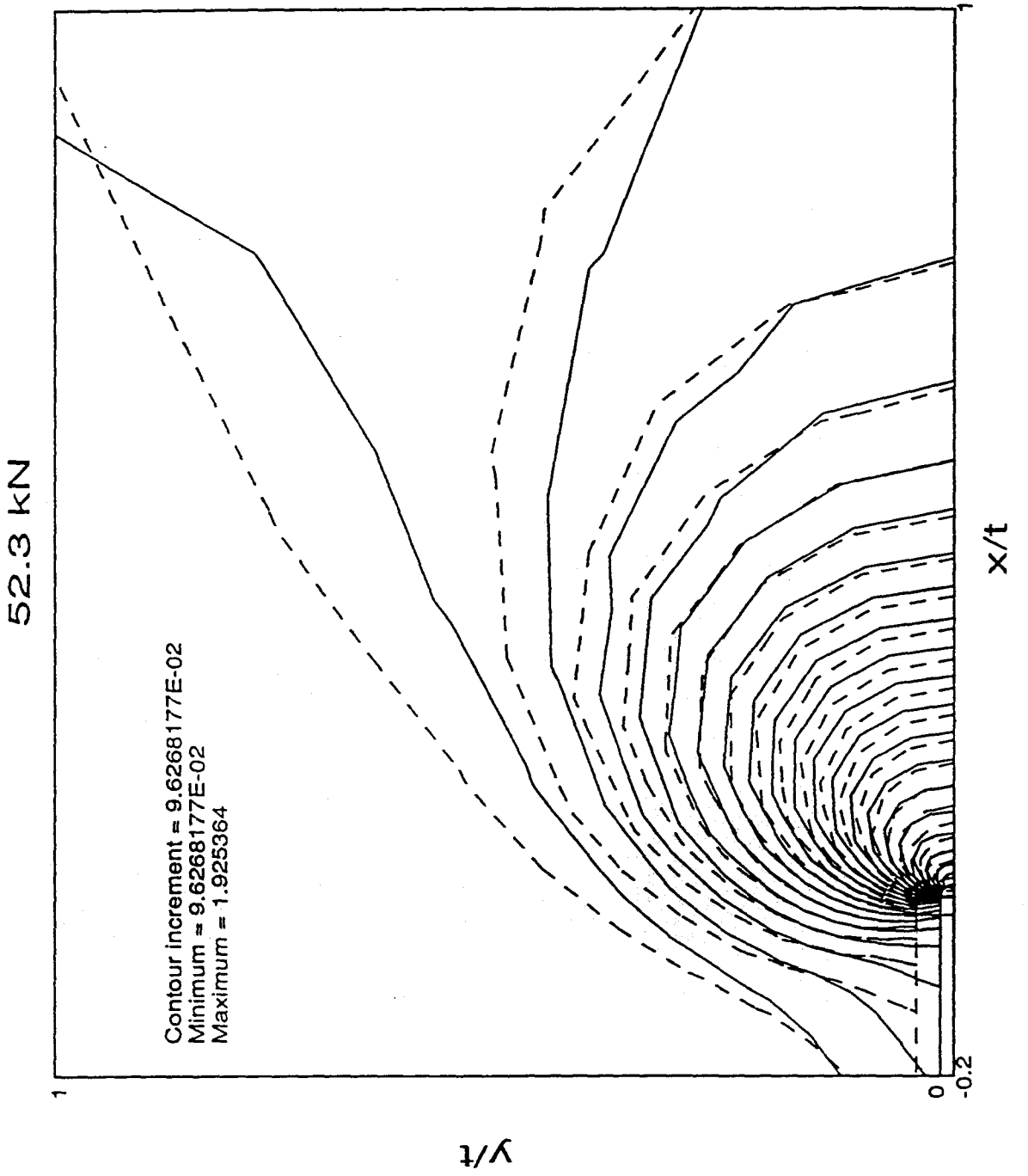

52.3 kN
Contour increment = 9.6268177E-02
Minimum = 9.6268177E-02
Maximum = 1.925364
y/t
x/t
1
0
-0.2
1

Figure (C.6)

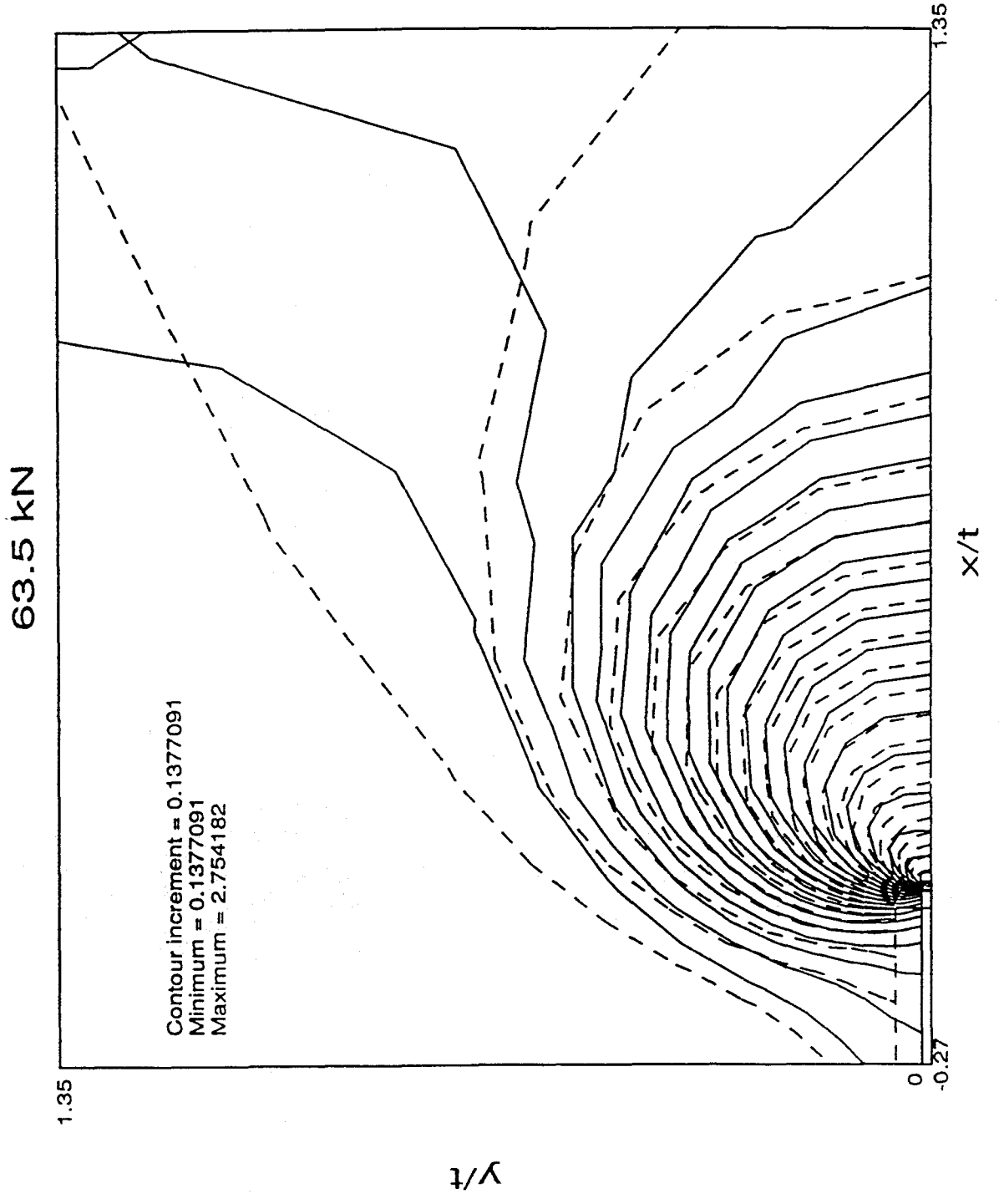

63.5 kN
Contour increment = 0.1377091
Minimum = 0.1377091
Maximum = 2.754182
1.35
1.35
0
-0.27
x/t
y/t

Figure (C.7)