Comparison of Experimental and Computational Crack–Tip Deformations Using Moire Interferometry and Finite Elements

Thesis by
Carl R. Schultheisz

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1991
(Defended March 14, 1991)

to my parents.

Acknowledgements

I would like to thank the Air Force Office of Scientific Research for extensive support for this research under Grant AFOSR–84–0254. Major David Glasgow, Dr. Nicolas Pagano and Dr. George Haritos were project monitors; their understanding and encouragement is greatly appreciated. Dr. George Haritos supported the change of the experimental material from aluminum alloy to 4340 steel because computer programing initiated by Dr. G. Ravichandran, Dr. Ares Rosakis, Dr. R. Narasimhan and Dr. Brian Moran promised financial savings, as did the availability of steel polishing equipment purchased by Dr. Rosakis. Thanks are due to these gentlemen for their assistance.

In addition, certain numerical aspects during the later phases of this work were supported by the Office of Naval Research under contract N00014–84–K–0424, monitored by Dr. Larry Peebles; this support is also much appreciated.

I would like to thank Dr. Alan T. Zehnder for his assistance in describing the heat treatment and uniaxial test procedures he employed in his experiments, and Dr. Rosakis for the use of his lapping machine.

I am very grateful to the National Science Foundation and the San Diego Supercomputer Center for providing access to the CRAY X-MP and Y-MP computers on which the numerical calculations were performed. I must also thank a large number of people for their help with the finite–element procedure and the use of the CRAY, especially Dr. R. Narasimhan, Dr. Ares J. Rosakis and Dr. Brian Moran for allowing me the use of their finite–element model and providing assistance in adapting the model to this particular study. Dr. Rosakis also graciously allowed me to use some of his allotment of time on the CRAY at the San Diego Supercomputer Center. Others who helped me to implement the finite–element model include Dr. Sridhar Krishnaswamy, Dr. Giancarlo Losi and Dr. Sy Shimabukuro. Dr. John Hall and Dr. G. Ravichandran were quite patient and very helpful in consultations about the numerical model.

I would also like to thank my *colleagues* Dr. Peter Washabaugh, Dr. Richard

Pfaff and Guillermo Pulos for their friendship and useful input on this project.

I must offer special thanks to the Newport Corporation for their generous donations of optical equipment and for providing a class in holography.

Finally, no list of acknowledgements would be complete without thanks to my advisor, Dr. Wolfgang G. Knauss.

Abstract

The large plastic deformations at the tip of a crack in a ductile heat treatment of 4340 steel are studied experimentally and numerically to investigate the details of the deformation in a tough material. The specimen is loaded in a three–point–bend arrangement. The finite–element model of the experiment uses a small–strain, incremental plasticity law, with a power–law hardening behavior. Both the in-plane and out-of-plane deformations were measured on the same specimen at the same time.

The experimental technique of moire interferometry is used to measure the in-plane displacements. This technique is described in detail, including an analysis of the effect of out-of-plane rotations on the use of the technique. A four–beam interferometer for measuring orthogonal displacment components is described, and its performance analyzed.

The three-dimensional, finite–element model has 11913 degrees of freedom, and provides data for comparison with the experiment between 4000 N (linear behavior) up to 73.5 kN (continuous fracture of the steel specimen). The model material properties are determined from a uniaxial test on specimens taken from the same bar as the fracture specimens and with identical heat treatment. This model characterizes the crack as a rounded notch to match the notch in the steel fracture specimen. The effects of tunneling of the crack are introduced through the release of nodes along the crack plane corresponding to measured crack profiles.

Results indicate that the numerical model matches the experiment quite well up to a load of 52.3 kN; mismatch at higher loads may be caused by a lack of finite–strain formulation in the code. The finite notch tip negates the singularity in either the stress or strain fields; the HRR field seems to have no region of dominance. However, the function of the J–integral appropriate to the HRR field does normalize the stresses and strains well, indicating that the J–integral is still a good fracture criterion. The effects of the added tunnel indicate that failure of the material depends on both the plastic strain and the hydrostatic stress.

Table of Contents

CHAPTER 1

Introduction

1.1 Motivation

Traditionally, the study of fracture mechanics has concentrated on providing for design purposes a measure of the resistance to fracture of a material. This resistance to fracture, or fracture toughness, is assumed to be a quantifiable material property that is independent of the specimen geometry, like Young's modulus or Poisson's ratio. However, the fracture toughness is not independent of the specimen geometry; for ductile materials, the fracture toughness in geometries approximating plane strain is much lower than the fracture toughness in geometries approximating plane stress [14].

Quantifying a fracture toughness of a material is obviously necessary in the design of structures, but reduction of fracture mechanics to a single parameter may neglect important details of the deformation at the tip of a crack. One common toughness parameter, the stress–intensity factor, K, is derived from the analytical series solution of Williams [90], called the K field. The stress–intensity factor K represents the strength of the singular term in the series, which is assumed to dominate the deformation near the tip of the crack. However, Williams's series is derived for two–dimensional deformations, plane strain or plane stress, which are often thought of as the limiting cases of infinitely thick and infinitely thin, planar bodies. For a body of finite thickness, there will be some added contribution to the deformation that is due to the three–dimensionality of the geometry. In addition, Williams's series is restricted to linearly elastic materials, whereas the stress singularity predicted by the series solution will certainly induce some plastic deformation in ductile materials. One approach to treating the effects of plasticity was offered

by Hutchinson [31] and by Rice and Rosengren [69], using a constitutive model in which the stress and strain are related through a power law. However, their solution (called the HRR field) is also restricted to two–dimensional geometries, and the stresses and strains are still predicted to be singular at the crack tip. The strength of of the HRR singularity is found using a second common fracture toughness parameter, the J–integral. This path–independent integral has an advantage over the stress intensity factor as a fracture toughness measure in that it was derived for any material in which the stress may be calculated from a potential, so that J is not restricted to linearly elastic materials.

The details of the deformation near the tip of a crack or notch are important, especially in the cases of materials that exhibit high resistance to fracture. If a better understanding can be reached regarding the mechanisms that lead to high fracture toughness, it may be possible to adapt those mechanisms to materials that lack natural resistance to fracture. For example, high fracture toughness in metals is usually due to large plastic deformations, whereas composites that exhibit high fracture toughness usually employ some kind of crack–stopping mechanism such as the Cook–Gordon mechanism, in which relatively weak interfacial bonds between matrix and fiber serve to redirect a crack along the interface parallel to the fibers instead of crossing the fibers.

In particular, it is important to be able to characterize the deformation near the crack or notch tip out to a scale where the nonlinear effects are not a factor. The state of the deformation may be much more confidently modeled in regions where the deformation is well described using linear elasticity. It is necessary, therefore, to gain a better understanding of the continuum behavior in the nonlinear region that lies within the region of linear elasticity. When this nonlinear continuum behavior is well understood, the next step in understanding the fracture process is to investigate material behavior at the size scale at which the breaking of molecular or atomic bonds occurs. A further goal of this work is the detection of effects that indicate behavior that cannot be described using continuum mechanics; detection of such effects may help in the estimation of the size of the region in which the discrete nature of the material must be included in any analysis.

This characterization of the deformation near the notch tip is attacked in two ways. The first line of attack is through an experimental investigation of the deformation surrounding the notch tip. The experimental technique should have high sensitivity and be capable of capturing the displacement over a wide area as well as over the large range of displacement caused by nonlinear deformation. The technique chosen was moire interferometry, a full–field technique, which generates contour maps of in-plane displacement. This experimental investigation is the major part of this study.

The second complementary line of attack is through a finite–element model of the experiment. This finite–element model employs a small–strain incremental plasticity theory, and the constitutive relation follows a power law in order to permit comparison of the results with the HRR field. An effort was made to match as closely as possible the material properties of the finite–element model with the measured properties of the material used in the experiment. As such, the measurements of the material properties that were made for use in the finite–element code represent a significant part of this study.

The experiment should serve to either validate or invalidate the results from the finite–element model, which is why so much effort has been put into matching the material properties of the numerical model to the physical material. As computers become faster and the size of the problems which they can solve becomes larger, it becomes easier to make predictions on the basis of numerical models rather than on experimental observations, even in cases in which there is substantial nonlinearity. If the numerical model agrees well with experiments over a wide range of specimen geometries and materials, there is more confidence in using the numerical model to predict the behavior of specimens that have not been, and perhaps cannot be, modeled in the laboratory. In that sense, this study serves as a single datum in the comparison between code and experiment.

It is very easy to manipulate the parameters entered into the numerical model. If it should be accepted that the numerical model is a good one, and that the time and expense of running a numerical simulation is less than the equivalent

experiment, the results of many calculations that include such a manipulation of parameters might be used as the basis for a better analytical approximation to the behavior of tough materials in the region surrounding a crack. Such an analytical model should include the effects of material nonlinearity and three–dimensionality, and perhaps be formulated in a manner that includes large strains.

1.2 Experimental Measurement

The main focus of this study is an experimental investigation into the deformation near the tip of a crack in a material that exhibits high resistance to fracture, in this case a ductile heat treatment of 4340 steel. 4340 steel is a common high–quality engineering material that may be heat–treated to give a wide range of material properties. In this case, the high fracture toughness of this material is brought about through considerable plastic deformation at the notch tip. The specimen is loaded in a three–point–bend geometry. The experimental method employed in this investigation is moire interferometry, a technique developed by Post [61,25,62,63] as an extension to geometric moire techniques. This technique measures in-plane displacement fields through the interference pattern developed between two laser beams diffracted from the surface of the specimen. The interference pattern consists of fringes, alternating dark and light lines, which represent a contour map of in-plane displacement; each fringe maps a locus of constant displacement. The sensitivity of the technique, defined as the displacement increment between adjacent dark fringes, is on the order of the wavelength of visible light. At the same time the in-plane displacements were measured on one side of the specimen; the out-of-plane displacements were measured on the opposite side using Twyman–Green interferometry. See Pfaff [56]. This technique also yields contour maps, in this case contours of constant out-of-plane displacement, and the sensitivity of this technique is one-half of the wavelength of light. The out-of-plane displacement measurements are not a topic of this thesis, so reference to the out-of-plane measurements will be brief.

A sketch of a moire interferometer is shown in Figure 1. This technique relies

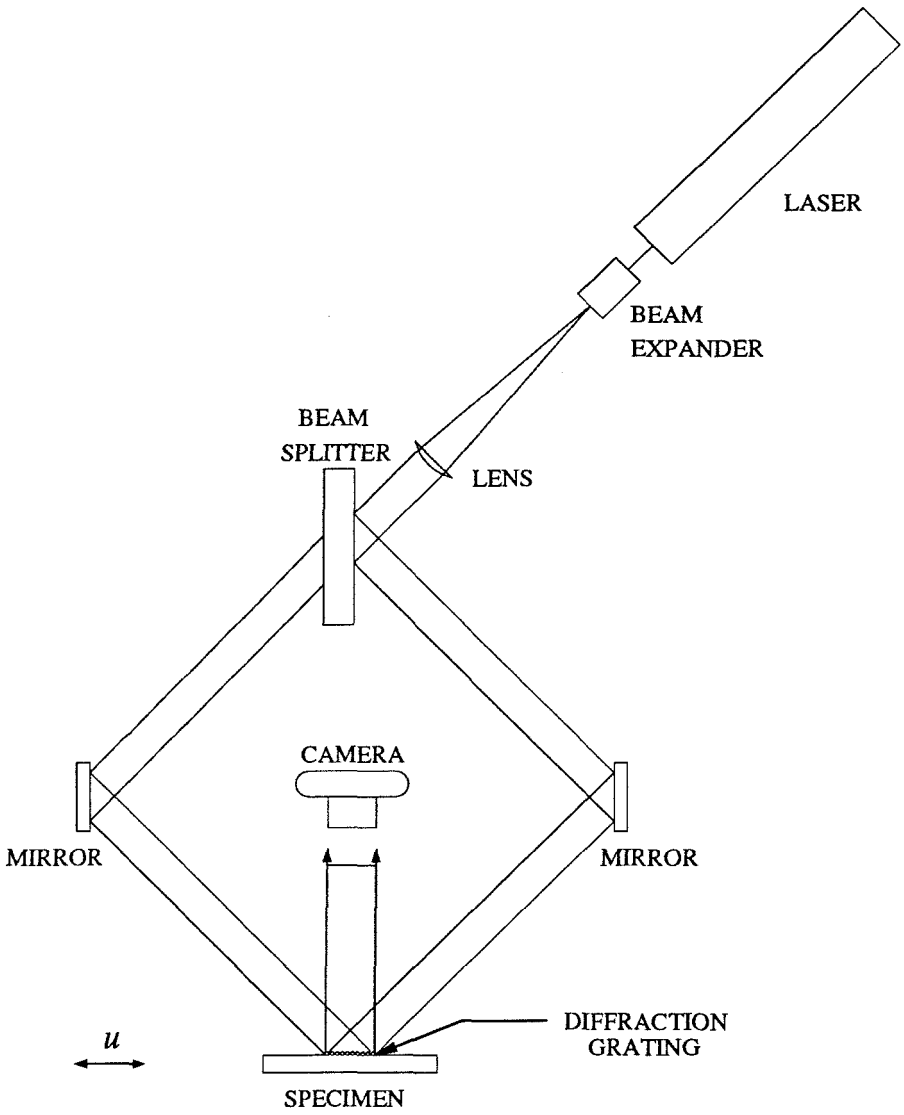


FIGURE 1. A sketch of the configuration of a moire interferometer. The displacement measured is in the plane of the paper, parallel to the specimen surface, as indicated by the arrow u.

on the diffraction of laser light from a grating attached to the specimen surface. The diffraction grating is a reflective, periodic height variation with a wavelength on the order of the wavelength of light. The angle of diffraction of each of the two laser beams depends on the wavelength of the grating and the wavelength of the light from the laser; if the wavelength of the grating changes because of a deformation of the specimen, the angle of diffraction changes as well. The symmetric

arrangement of the two incoming laser beams results in diffracted light that leaves the specimen at equal but opposite angles, which causes an interference pattern. This interference pattern is viewed through the camera as a pattern of light and dark fringes overlaid on the specimen surface; this fringe pattern forms a contour map of the deformation of the specimen surface. The laser beams must originate at the same source, travel paths of similar lengths and have nearly equal intensities in order to produce fringe patterns of high quality. Also, since the diffracted laser light consists of nearly plane waves, the camera aperture must be as large as the field of view over which the displacements are to be measured. In Figure 1, the measured displacement component lies in the plane of the paper and parallel to the surface of the specimen, as shown by the arrow labelled u. A second interferometer is needed, arranged in a plane perpendicular to the paper, in order to measure the displacement component perpendicular to the plane of the paper and parallel to the surface of the specimen.

A more thorough review of the technique of moire interferometry may be found in the *Handbook on Experimental Mechanics* [25]. Advantages of this technique include the fact that it is a full–field technique, so that it yields displacement information over an entire field of view; the technique is not limited to small displacements; the fringe pattern is coincident with the surface of the specimen; the fringe patterns are insensitive to rigid translations; and the fringes are formed in real time, making the technique suitable for dynamic measurements. Already mentioned is the fact that the sensitivity is on the order of the wavelength of light; in addition, the sensitivity may be adjusted somewhat if the magnitude of the expected deformation can be estimated in advance. In this study, the displacement increment between adjacent black fringes is 0.833 microns; Post [25] suggests that resolution of one-fifth of a fringe may be obtained by manual interpolation. Also, the technique does not require any assumption about the form of the deformation, unlike the method of caustics, nor is a knowledge of the effect of stress on the optical behavior of the material needed as in photoelasticity. In the *Handbook* [25], Post also suggests a method for determining the out-of-plane displacements at the same time as the in-plane displacements.

Some of the disadvantages of moire interferometry include the high fringe densities generated by large deformations; somewhat noisy fringe patterns, at least compared to fringe patterns generated using a Twyman-Green interferometer; and a light efficiency of only about 10% because of the crossed gratings necessary to measure orthogonal in-plane displacement components. Perhaps the most important problem for this study is a sensitivity to out-of-plane displacement gradients, as the high in-plane stresses will cause large out-of-plane displacement gradients near the notch tip. The out-of-plane displacement gradients have two effects on the moire interferometer. The first effect is that sufficiently large out-of-plane displacement gradients will introduce an error in the fringe patterns. This error might be corrected if the out-of-plane displacement field were known from some independent source, such as on out-of-plane displacement measurement or estimate from a numerical model. The second effect is that the out-of-plane displacement gradients will direct some of the laser light so that it falls outside the aperture of the camera used to record the fringes, causing a shadow region in which no fringe data may be recorded. This problem did occur near the tip of the notch in the experiment described in this work, and the performance of the optical system may be improved by moving the aperture closer to the specimen, with the constraint that the laser beams that are directed at the specimen may not be blocked. As it happens, in the experimental configuration used for this investigation, the shadow region corresponded quite well to the region where the fringe patterns were expected to be affected by the out-of-plane displacement gradients, so that the effect of the out-of-plane gradients could be neglected. Even if the data had been available from the shadow region, a correction to the fringe patterns should probably be ignored as a first approximation, as the effect of the out-of-plane gradients is lessened when the in-plane deformation is large, as it is in the near-tip region. The effects of out-of-plane displacement gradients are discussed in Chapter 2.

In the chapter on moire interferometry in the *Handbook on Experimental Mechanics* [25], Post discusses briefly the effects of out-of-plane displacement gradients on the technique. He concludes that moire interferometry is affected only by out-of-plane gradients that are equivalent to rotations about an axis parallel to the

lines of the grating, and suggests that the effect of such a rotation is to introduce fringes corresponding to a compressive displacement; these fringes have a frequency F given by

$$F = f\theta^2/2, \tag{1.2.1}$$

where f is the frequency of the reference grating, and θ is the angle of rotation. In Chapter 2, however, it will be shown that rotations about an axis perpendicular to the grating will also have an effect when coupled to the rotation about the axis parallel to the lines of the grating. Post's analysis of the effects of the tilting of the grating is based on his analogy between moire interferometry and geometric moire. The analogy between geometric moire and moire interferometry is exact if there is no tilting of the grating plane. The result shown in equation (1.2.1) follows from the ideas discussed in Section 2.3.14, as in equation (2.3.69). However, the analysis in section 2.3 indicates that there are some additional effects peculiar to moire interferometry that must be included for completeness. Post's analysis does indicate the sensitivity of the moire interferometer; he suggests that rigid rotations as small as 0.001 radian can be observed in an experiment, and removed.

1.3 Finite–Element Analysis

The finite–element model is adapted from the model used by Narasimhan, Rosakis and Moran [52], and their assistance with the model is gratefully acknowledged. This is a three–dimensional model employing a small–strain incremental plasticity theory following a Huber–Von Mises yield condition. The model consists of 3200 eight–noded block elements in ten layers through the half–thickness of the specimen, resulting in 11913 degrees of freedom. The constitutive relation used in the finite–element model relates the stress and strain through a power law for comparison with the analytical HRR field solution. The dimensions of the finite–element model were matched to the experimental specimen, which is 30 cm long by 7.6 cm high by 1 cm thick, with a single notch 3 cm long cut into the center of one of the longest sides. The notch is cut with an electric discharge; it is 0.03 cm across. The finite–element model approximates the notch tip as circular, with a radius of

0.015 cm.

The material properties of the numerical model, such as Young's modulus and the yield stress, were determined from measurements of specimens cut from the same piece of steel as the specimen used in the experiment. Young's modulus, the yield stress, and the strain at yield were determined in a uniaxial test. The power law used for comparison with the HRR field was also fit to the measured uniaxial stress–strain behavior of the steel; the important parameter in the power law is the hardening exponent, which was taken to be 20. This high value for the hardening exponent implies that the material is not much different from an elastic/perfectly plastic material. Additionally, the value of Poisson's ratio was measured for this steel using the moire interferometer. Assuming plane stress, the out-of-plane displacements are linearly dependent on the value of Poisson's ratio; the dependence of the in-plane displacements on Poisson's ratio may be estimated using the K field solution; that dependence is in the same range.

In addition to matching the specimen geometry and material properties, the tunneling behavior of the material was also included in the numerical model, through a node–release procedure. The shape of the internal tunnel was measured in a series of experiments as a function of load, and the appropriate nodes corresponding to the shape of the tunnel were released. A numerical model that did not include tunneling was also run to the same load levels for comparison. (The simulation of Narasimhan, Rosakis and Moran [52] included the Gurson material model for void growth; however, the Gurson model was not included in the present simulation, because it led to convergence problems at failure initiation.)

The careful characterization of the material properties, even including the measured tunneling behavior, was an effort to remove as many discrepancies between the experiment and the finite–element model as possible. The numerical model still has three major areas in which its simulation of the experiment could be improved. The first of these areas is the use of the power–law constitutive formulation. The choice of a power law to approximate the constitutive behavior of the material allows a direct comparison between the HRR field and the results from the finite–element

model. The hardening exponent was chosen as 20 so as to compare directly with the tabulated HRR functions of Shih [77], whereas the best fit for the hardening exponent was 23.7. However, the power–law formulation does not follow the measured uniaxial stress–strain relation very well, as can be seen in Section 3.3. In order to ensure that the finite–element model matches the experiment exactly, the measured uniaxial stress–strain behavior of the steel could be input as a table for the code to use in calculating the constitutive response of the numerical model. The use of the power law in the numerical model was the result of the fact that the power–law formulation was already implemented in the numerical code.

The second major area in which the code may fall short is that it is restricted to a linearized (small) strain formulation. At the highest load levels, the strains predicted by this numerical model fall outside the range in which the linearized–strain theory is valid, at least in a small region surrounding the notch tip. Some effort has been made to characterize the uniaxial stress–strain response of the steel at finite strains, after necking has occurred, through the assumption of plastic incompressibility and the ratio of the minimum area of the necked region to the initial area of the specimen. This finite–strain response of the steel may be included in a model that is formulated to account for strains outside the range of the linearized theory.

The third possible major source of discrepancy between the experiment and the numerical model is due to the discretization necessary for the finite–element method. It is assumed that if the elements are made smaller, the solution obtained from the calculation is closer to the true solution. As the elements become smaller, however, the required computing time increases. The choice for the discretization must make a compromise between the need for accuracy and the cost involved. Fracture problems require very small elements near the tip in order to model the high stress gradients in that region adequately. The crack in this study is actually a rounded notch, so it is expected to be somewhat more benign than a mathematically sharp crack. The in-plane dimensions of the elements surrounding the notch tip were chosen to be on the order of one fifth of the notch tip radius. As the plastic zone becomes large, however, it may be that more elements are needed away from the notch tip to characterize the elastic–plastic boundary correctly. With access

to the new CRAY Y-MP at the San Diego Supercomputer Center, it was possible to increase the number of layers through the half–thickness from six (as used by Narasimhan, Rosakis and Moran [52]) to ten.

1.4 Discussion

In many ways, this specimen geometry and material should provide a difficult test of the analytical approximations, and of the ability of the numerical model to simulate the experiment. The finite thickness of the specimen makes it incompatible with the two–dimensional assumptions of both the K field and the HRR field. The finite notch tip requires the addition of a more singular term to provide the traction–free boundary condition on the surface of the notch; such a correction is available from another term in Williams's K field series solution, but not for the HRR field. Also, the effects of the bending present in this specimen geometry cannot be captured by the singular terms employed by the K field or the HRR field. Again, higher–order terms may be added to the K field from Williams's series, but the HRR field has no higher–order terms. Shih [78] believes that the HRR field actually works better for bending across the uncracked ligament than for simple tension across the uncracked ligament, but this depends on the criteria used to make such a decision. The singular term of the HRR field is closer to the case with bending very near the crack tip, but the bending solution soon drops off very quickly. The case of simple tension does not match as well at the crack tip, but it is closer to the singular term of the HRR field over a wider range away from the crack tip.

The large plastic zone at the higher loads clearly lies outside the linear elasticity assumption of the K field. However, the plasticity formulation may also be incompatible with the HRR field. The HRR field and the J–integral are both derived within the framework of deformation plasticity theory, equivalent to nonlinear elasticity. The code employs an incremental plasticity theory, which is believed to better represent the actual constitutive behavior of real materials. Incremental theory and deformation theory coincide only in the case of proportional loading,

although the results of the two theories should grow closer if the load history approaches proportional loading. The two theories differ the most when unloading occurs, such as when the crack propagates. With deformation theory, unloading occurs along the same nonlinear path followed during loading, while with incremental theory, unloading occurs along a linearly elastic path.

It should be possible to use any discrepancies between the experimentally measured displacements and the numerically calculated displacements to indicate where the numerical model needs improvement. It is expected that the finite–element model will closely match the physical experiment at low loads, before any plasticity develops. If this is not the case, it is likely that the discretization of the model needs to be improved further. Discrepancies that occur at higher loads, but before the strains pass beyond the linearized theory, are likely due to the constitutive formulation. This is particularly true if these discrepancies occur at the boundary between elastic and plastic response. The steel used in the experiment displays the effect known as Lüder's bands in its uniaxial stress–strain response, which is caused by the release of dislocations pinned by trapped nitrogen and carbon atoms in the interstitial regions between iron atoms. The Lüder's bands are seen near the yield point as a zone of increasing strain without appreciable increase in stress, and this effect is poorly modeled by the power law used in the code. It may be that the Lüder's bands are observed in the uniaxial test only because of the nearly uniform state of stress and strain in the uniaxial specimen test section, so that the effect of the Lüder's bands is observed as they propagated through the entire test section. Since the stress and strain in the experiment and the numerical model are not uniform, the release of any pinned dislocations may be localized, so that the local constitutive behavior is more like the power–law approximation. It would be interesting to study the Lüder's bands phenomenon in a uniaxial test using the moire interferometer. At higher loads, when the plastic zone has become relatively large, discrepancies between the experiment and the numerical model at the elastic/plastic boundary would point to a need for mesh refinement away from the crack tip. Discrepancies deep within the plastic zone would indicate a need for a finite–strain formulation of the constitutive model.

The consequences derived from Williams's K field indicate that investigation into the three–dimensional effects that are due to the finite thickness of the specimen is probably better undertaken through measurements of the out-of-plane displacement than through the in-plane displacements. Assuming that the two–dimensional cases of plane strain and plane stress represent the limiting cases of infinitely thick and infinitesimally thin three–dimensional bodies, the K field solution predicts that the out-of-plane displacement at the crack tip proper is zero in plane strain and infinite in plane stress. On the other hand, the in-plane displacements are bounded in both cases, with the difference between the two cases being a function of the Poisson's ratio of the material. For a Poisson's ratio of 0.3, the difference in the in-plane displacements between plane stress and plane strain varies from about 10% to 30% around the crack. Thus, the out-of-plane displacement field should be much more sensitive to the finite thickness effects than the in-plane displacement fields.

The onset of plasticity changes things somewhat. The HRR field also predicts zero out-of-plane displacement for plane strain and infinite out-of-plane displacement for plane stress, but the HRR in-plane displacement fields are also quite different, depending on whether plane strain or plane stress is assumed. It is clear from the experiments that the fringes corresponding to the in-plane displacement parallel to the crack change their shape considerably after plastic deformation has begun, unlike the fringes corresponding to the other two displacement components. It is expected that the in-plane deformation of a plate such as the specimen in this study is still nearly two–dimensional. The in-plane displacement fields at the specimen surface in the presence of plasticity should indicate the nature of the deformation fairly well, while the out-of-plane displacement at the surface is effectively a thickness average. In addition, the fringe densities of the in-plane displacement fields are still smaller than the fringe density of the out-of-plane displacement field, so that collecting data from the in-plane displacements is easier than from the out-of-plane displacements.

One initial goal of this study was to determine where continuum mechanics fails in modeling the behavior of the material at the tip of a crack. However, for this specimen geometry and material, the crack begins to propagate in the center of the

specimen, away from the surface where the measurements are made. This is clear from the tests made to characterize the tunneling behavior of the crack for input to the finite–element model. However, it may be that the effects of crack tunneling are visible on the specimen surface. The fringes of in-plane displacement perpendicular to the crack show an interesting feature that indicates unloading at the surface. The finite–element code with the tunneling included using the nodal release procedure also predicts a region of unloading ahead of the crack at the surface of the specimen. At the highest load achieved in the experiment, just before constant, slow creeping failure began, this unloading zone lies about 4 or 5 mm ahead of the original notch tip, while the extent of the tunnel is less than 1 mm ahead of the original notch tip. Detection and measurement of such internal crack propagation might be best attacked through x–ray methods.

There are some interesting results from comparison of the tunneling tests to the results of the void growth simulation of Narasimhan, Rosakis and Moran [52]. The first of these is that the estimate of the critical value of the J–integral from the void growth model [52] agrees fairly well with the critical value of J calculated from the tunneling tests. These results are discussed in more detail in Chapter 4. Another interesting result from [52] is the difference in the location of the first element to fail from the case of plane strain to the three–dimensional model, also evident in the plot showing the contours of equivalent stress (the current yield stress). In the plane–strain case, the first element to fail lies on the crack line, while in the three–dimensional case the first element to fail lies slightly above or below the crack line. The three–dimensional case mimics the results from the tunneling experiments, in which small cracks initiated above or below the crack line along the thickness direction and then grew together to meet at the crack plane to form the tunnel. The measured tunnel profiles more closely resemble the contours of constant hydrostatic stress rather than the contours of constant equivalent stress from Reference [52]; this result is not surprising because the hydrostatic stress is mostly due to the tensile stress acting to open the crack, while the equivalent stress is related to the deviatoric stress.

As already mentioned, the extent of the tunnel was less than 1 mm at the

highest load achieved in the experiment. The ultimate failure of this ductile steel occurs in a shear mode, with shear cracks initiating at the free surfaces at about $\pm 45°$ above and below the crack line after the tunnel has achieved that 1 mm in the center of the specimen. The shear cracking then occurs in tandem with the tensile crack that proceeds along the original crack plane, until the shear cracks from opposite sides of the specimen join up and the tensile crack is eliminated. If the specimen is loaded continuously until it breaks into two pieces, a tensile crack of 5 to 7 mm in length may be observed. However, most of this tensile crack surface has been formed after the shear cracking has begun. Before the tests were performed, it was expected that the tunnel might progress this far into the center of the specimen, and then the shear cracking would initiate and cause failure. This is the generally accepted notion of the progression of crack growth. For this ductile material, however, the experiments indicate that the shear cracks begin soon after the initiation of tunneling, proceeding with the tensile cracking and eventually the shear failure mode overwhelms the tensile failure mode.

CHAPTER 2

Moire Interferometry

2.1 Introduction

Moire interferometry is an experimental technique for measuring in-plane displacements at the surface of a body. This technique relies on the interference pattern formed between two diffracted laser beams to provide interference fringes, which form a map of contours of constant in-plane displacement. The technique of moire interferometry is comparable to two other experimental techniques that may be more well known. The first of these techniques is what will be called geometric moire for this discussion, in which in-plane displacements are determined by directly comparing two gratings made up of closely spaced lines, one grating drawn on a specimen and one drawn on a transparent reference medium. Moire interferometry grew out of earlier work involving geometric moire, and the analogy that will be drawn between moire interferometry and geometric moire is quite useful. The second technique that shares some of the attributes of moire interferometry is Twyman-Green interferometry, in which out-of-plane displacements are measured by comparing a laser beam reflected from the specimen with a laser beam reflected from a flat reference mirror. Moire interferometry shares with Twyman-Green interferometry the advantage of high sensitivity derived from the use of the wavelength of light as the measurement scale, as well as the problems of measuring large deformations. All three techniques, moire interferometry, geometric moire and Twyman-Green interferometry, employ fringe patterns representing contours of constant displacement to infer the deformation of the specimen.

Moire interferometry was introduced by Post [61] as an extension of earlier geometric moire techniques in an effort to make more sensitive measurements, and the analysis that follows investigates the relationship between moire interferometry and geometric moire. Although the fringes observed in moire interferometry and in geometric moire are formed by different mechanisms, Post drew upon an analogy to relate the two techniques. In Section 2.2, geometric moire will be discussed by way of providing some background for moire interferometry. The correspondence between the two techniques will be demonstrated by comparing the equations that govern the fringe patterns observed in both geometric moire and moire interferometry. From geometric considerations, it can be seen that the fringes observed using geometric moire represent contours of constant in-plane displacement. A match between the equations governing the fringes in moire interferometry and the equations governing the fringes in geometric moire would indicate that the fringes observed in moire interferometry also represent contours of constant in-plane displacement. The equations describing the fringes in moire interferometry will be shown to be identical to those for geometric moire, indicating that the fringes in moire interferometry also represent exactly contours of displacement, as long as the specimen surface does not tilt, either rigidly or as a result of out-of-plane displacement gradients.

The fringes formed in moire interferometry are strongly dependent on the angular deflections of laser beams diffracted from the specimen surface. Tilting of the specimen surface, either through a rigid rotation or a slope that is due to out-of-plane deformation, will have some effect on the angular deflection of the laser beams. This is particularly important in studying fracture, as there are strong out-of-plane displacement gradients near the tip of the crack. In Section 2.3, the effect of a rigid tilt of the specimen surface on the fringes will be studied; the effect of out-of-plane displacement gradients will be similar as long as the rotation of the grating surface can be taken to be locally homogeneous. The equations describing the fringes for a two–beam interferometer will be derived for the general problem of interference between two diffracted plane waves. This derivation will then be made more specific to the case of moire interferometry, which includes a particular initial alignment of the laser beams with the undeformed specimen. A special

case of deformation without tilting will be described. For this special case, the correspondence between the equations describing the fringes in moire interferometry and the equations describing fringes observed in geometric moire is exact, leading to the conclusion that the fringes observed in moire interferometry can be taken as contours of constant relative in-plane displacement for planar deformation. The effects of nonplanar deformation will be investigated under the assumption of small rotations of the surface away from the original plane of the grating, in order to investigate the dominant sources of error if the fringes are assumed to be contours of displacement.

The four–beam interferometer used in the experiments will then be described in Section 2.4, along with the procedure for aligning the interferometer in an experiment. In order to measure two independent components of the in-plane displacement, two laser beams are needed to interfere for each displacement component. The four–beam interferometer uses two separate two–beam interferometers arranged at 90° to one another to measure orthogonal, in-plane displacement components. The requirements for coherence, polarization and matching of beam intensities will be discussed. The optical collection apparatus will be described, and the theoretical limits of the current measurement system will be estimated. Some improvements for further tests will be suggested.

A discussion of the sources of uncertainty in the measurement then follows in Section 2.5. Particular attention is paid to the sources of systematic error that may arise in the measurement technique. These are errors that may propagate through the entire measurement, although if the sources of systematic error may be identified and quantified, their effects on the measurement may be corrected, in some cases after the fact.

The moire interferometer used in this study was initially configured as a three–beam interferometer, in which one of the laser beams is used twice in the measurement of orthogonal, in-plane displacement components. The three–beam interferometer has some advantages and some disadvantages when compared to the four–beam interferometer finally adopted for use in the experiments. In Appendix

A, the performance of the three–beam interferometer is compared with that of the four–beam instrument. The equations describing the fringes for the three–beam interferometer will be derived for the case of planar deformation, showing that the fringes are governed by the same equations as are found in geometric moire. Thus, the fringes observed using the three–beam interferometer also represent a contour map of in-plane displacement.

2.2 Geometric Moire and Moire Interferometry

Geometric moire is a well–established experimental method used to measure in-plane deformations on the surface of a specimen by generating fringe patterns that can be related to the displacement field. This method has been used as a technique to measure displacements and strains since the late 1940's and early 1950's. [17,18,74,85] † The fringes observed in geometric moire are generated from the superposition of two overlapping line gratings. It is clear from geometrical considerations that each of the fringes generated in geometric moire represents a contour of constant in-plane displacement. Moire interferometry also generates fringe patterns; these fringes are the result of interference between two crossed laser beams. Moire interferometry was initially proposed as an extension of geometric moire in an effort to improve the sensitivity of the measurement. However, the two techniques rely on very different mechanisms to generate fringe patterns, and it is not obvious that moire interferometry can be viewed simply as a form of geometric moire.

The first concern in using moire interferometry is whether it is exactly analogous to the established techniques of geometric moire, in the sense that the fringes represent a contour map of the displacement field. A brief discussion of each technique follows, and by comparing the equations relating the fringe spacing and orientation to the displacement fields, it will be shown that moire interferometric fringes do indeed represent contours of in-plane displacement when it is assumed that the specimen surface remains flat and parallel to its undeformed position.

A second concern in using moire interferometry relates to the situation when the specimen surface does *not* remain flat or parallel to its initial position. This concern is particularly important in the case of fracture mechanics, as the deformation field that is due to a crack in a body is strongly three–dimensional, with large out-of-plane deformation gradients near the tip of the crack. The effect of these

---

† A recent review of moire techniques can be found in a paper by Sciammarella [74], including projection and reflection moire techniques used to measure out-of-plane displacements. Much more complete treatment of geometric moire can be found in the books by Durelli and Parks [17] and Theocaris [85].

out-of-plane deformation gradients must be examined before assuming that they have no effect on the fringe patterns and that the fringes accurately map the in-plane displacements. The equations governing the formation of the fringes while allowing for the out-of-plane rotations will be derived in Section 2.3 to determine the influence of such rotations on the fringe patterns. The resulting equations will be written explicitly for the special case of small rotations in order to determine the strongest influences on the fringe patterns.

2.2.1 Geometric Moire

Geometric moire is a technique that measures displacements by tracing the movement of features on the specimen surface. The technique relies on features that dare applied to the surface in a regular pattern, called a grating, † but the features themselves are not actually followed. Instead, their displacement is inferred from a pattern of interference fringes seen when two similar gratings are superposed.

A regularly spaced series of lines is inked or scribed onto the surface of the specimen, these lines form the *specimen* grating, and it is assumed that the specimen grating exactly follows the deformation of the specimen surface. A similar series of lines inked onto a transparent medium is called the *reference* grating. The reference grating retains the original shape of the specimen grating throughout the deformation. By superposing the reference grating over the deformed specimen grating, the change in position of the lines of the specimen grating from their initial alignment may be observed. In particular, if the spacing between lines of the gratings is small enough, the superposed pair of gratings forms a pattern of fringes called *moire* fringes, after a French word for a silk fabric that demonstrates these patterns. Durelli and Parks [18] have shown that moire fringes represent parametric curves of constant in-plane displacement relative to the lines of the reference grating. This can be seen to be a result of the geometric relation of the overlap between the lines of the reference grating and the lines of the specimen grating. The sensitivity of the

† The most commonly used gratings are equally spaced parallel lines, but gratings of concentric circles or radial lines are also used.

measurement, which will be defined here as the displacement increment between adjacent dark fringes, is equal to the distance between adjacent lines on the reference grating. (In fact, a coarse, more easily replicated specimen grating can be used with a finer reference grating to improve the sensitivity of the technique. This phenomenon is called fringe multiplication [59], and it will be described in more detail below.) The direction of the displacement measured is perpendicular to the lines of the reference grating, indicating that two sets of linearly independent reference and specimen gratings are needed to determine the in-plane deformations completely. For straight–line gratings the fringe pattern is independent of translation of the reference grating with respect to the specimen grating, which is an indication that the measurement is relative, and cannot be made absolute unless the fringe pattern extends to a point where the absolute displacement is known by other means.

The spacing and inclination of the fringes are determined through geometry from the line spacing, or *pitch*, p_0, of the reference grating, the pitch p of the deformed specimen grating, and the angle θ between the lines of the two gratings, as seen in Figure 2. Initially, before deformation, the pitch of the specimen grating is also equal to p_0. It is assumed that the deformation is homogeneous over the field of view. (For inhomogeneous deformations, it is assumed that the pitches of the gratings are small enough so that the deformation is *locally* homogeneous over several lines of the gratings.) The spacing between adjacent dark fringes, δ, is found as

$$\delta = \frac{pp_0}{[p^2 + p_0^2 - 2pp_0 \cos\theta]^{\frac{1}{2}}} \tag{2.2.1}$$

The inclination of the fringes is described by the angle ϕ, which is the angle between the fringes and the lines of the reference grating. The angle ϕ is related to the pitches of the gratings p and p_0 and to the angle θ between the lines of the gratings by

$$\tan\phi = \frac{p_0 \sin\theta}{p_0 \cos\theta - p} \tag{2.2.2}$$

The increment of displacement between adjacent fringes is p_0, and the direction of the displacement measured is perpendicular to the lines of the reference grating.

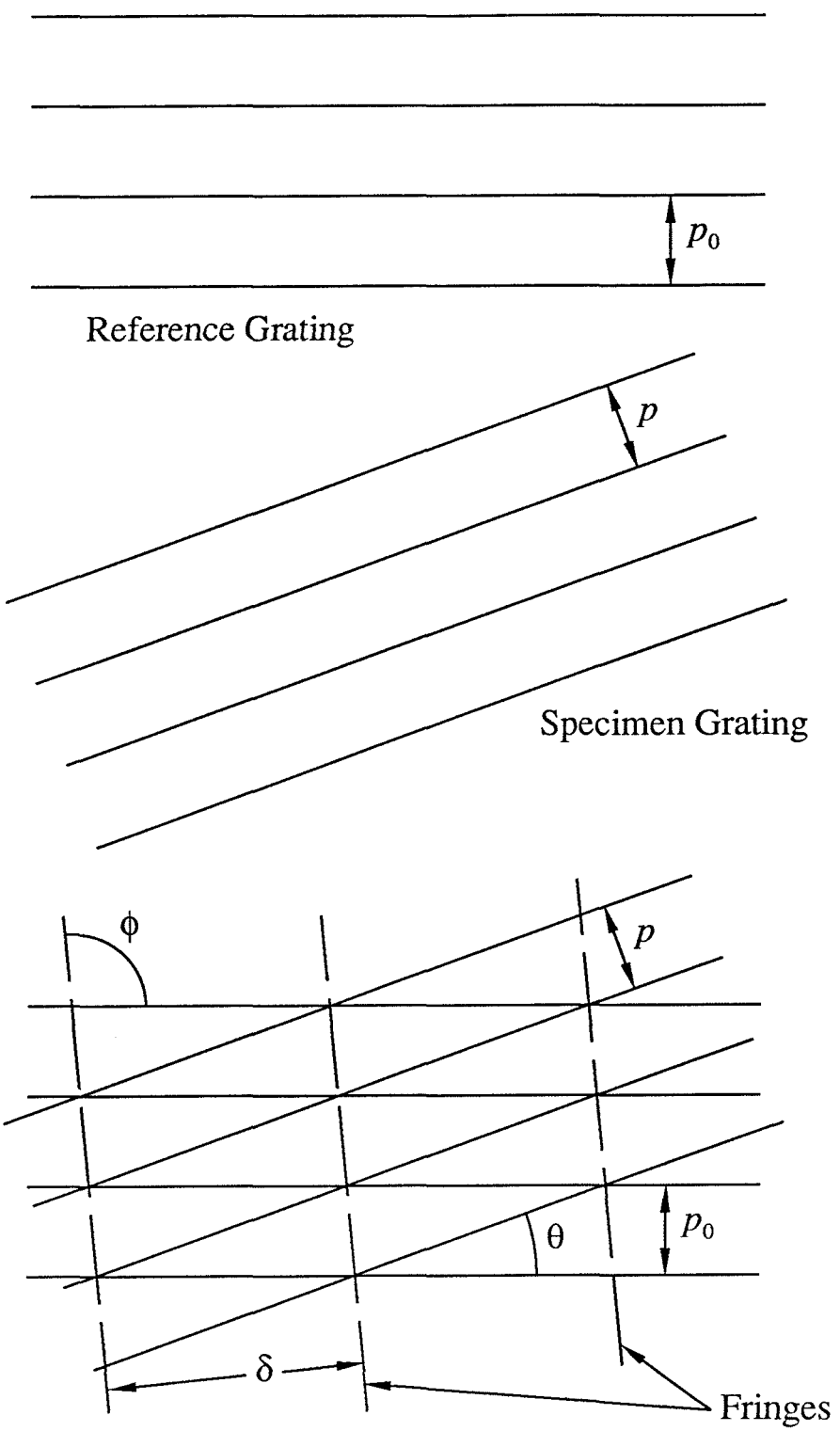


FIGURE 2. Fringe formation in geometric moire. The reference grating is shown at the top, and the specimen grating below it. The formation of the fringes is shown at the bottom, when the two gratings are superposed.

By measuring ϕ and δ, one could in principle determine values of p and θ, but the relations (2.2.1) and (2.2.2) do not provide unique values of p and θ for a given ϕ and δ. Additionally, if the fringe pattern is complicated because of large displacement gradients, the calculation of the fringe spacing and inclination at a point is difficult, whereas the location of each fringe can be made quite accurately. The actual values of p and θ are related to the strain field, so determination of p and θ is equivalent to differentiation of the displacement field found from the fringe pattern. As always, the differentiation of quantities measured at discrete locations introduces additional difficulties, since any lack of smoothness in the measurements will be magnified by the process of differentiation.

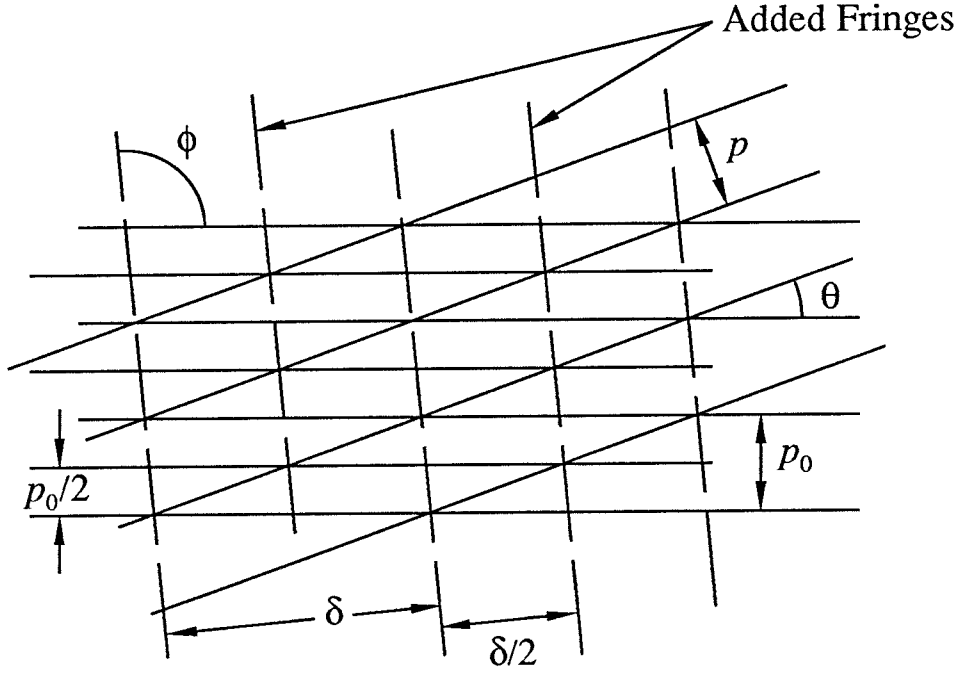


FIGURE 3. Fringe multiplication. When the reference grating spacing is halved from p_0 to $p_0/2$, additional fringes are seen halfway between each pair of fringes that were observed when the reference grating spacing was p_0.

If the reference grating is replaced by a reference grating having a pitch of $p_0/2$, the phenomenon of fringe multiplication [59] becomes apparent. Again, the specimen grating has pitch p, and the angle between the lines of the two gratings is θ. The initial pitch of the specimen grating is still p_0, or twice the pitch of

the reference grating. This arrangement can be thought of as using two reference gratings of pitch p_0 offset with respect to one another by $p_0/2$, leading to two sets of fringes, one corresponding to each reference grating. Thus, another dark fringe is observed exactly halfway between any two dark fringes that were seen when the reference grating had a pitch of p_0. In this case the spacing between adjacent fringes, δ, is half of that in equation (2.2.1), so that δ is found as

$$\delta = \frac{pp_0}{2\,[p^2 + p_0^2 - 2pp_0\cos\theta]^{\frac{1}{2}}} \tag{2.2.3}$$

The inclination of the fringes, ϕ, remains unchanged, so that ϕ is determined by

$$\tan\phi = \frac{p_0\sin\theta}{p_0\cos\theta - p} \tag{2.2.4}$$

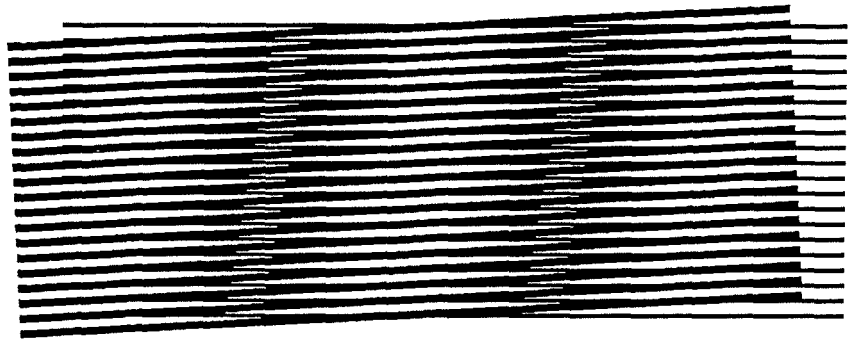
The sensitivity of this measurement (the displacement increment between adjacent fringes) is now $p_0/2$, in the direction perpendicular to the lines of the reference grating. By replacing the reference grating with a grating that has a pitch half that of the original, but using the same specimen grating, the sensitivity of the measurement has also been halved.

In using the reference grating with a pitch of $p_0/2$, another set of fringes will be formed at a spacing and inclination found by replacing p_0 by $p_0/2$ in equations (2.2.1) and (2.2.2). However, these fringes are difficult to see. The most easily resolved fringes are fringes formed between gratings of nearly equal pitch at small angular differences. Thus, the dominant fringe pattern is that pattern described above, corresponding to two reference gratings of pitch p_0 offset from one another by $p_0/2$. The dominant fringe pattern is described by equations (2.2.3) and (2.2.4).

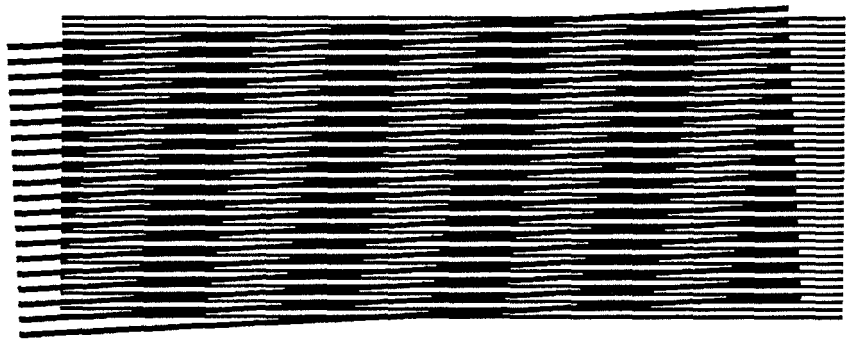
The sensitivity of the measurement is important, because it is desirable to have a fringe pattern that contains enough fringes so that the confidence of the measurement is high over the entire field. The sensitivity of geometric moire techniques is limited by the minimum spacing between lines on the reference grating. Many efforts were made to improve the sensitivity of geometric moire, including the fringe multiplication technique introduced by Post in 1967 [59] and the idea of fringe interpolation introduced by Sciammarella, also in 1967 [76]. In 1971, Post [60] suggested

replacing the reference grating with a diffraction grating, making use of the work of Guild [23,24]. Walker and Mckelvie [87] introduced the use of crossed laser beams to form a virtual reference grating in 1978; however, their fringes were formed from gratings that were copied from the deformed specimen. In 1980, Post [61] suggested the technique of *moire interferometry*, in which a diffraction grating is produced on the specimen surface and then interrogated by two crossed laser beams to produce interference fringes. All of these techniques were introduced as simple extensions to geometric moire, but the result is a technique far removed from the simple crossed line gratings. In the next section, moire interferometry and its relation to geometric moire will be discussed.

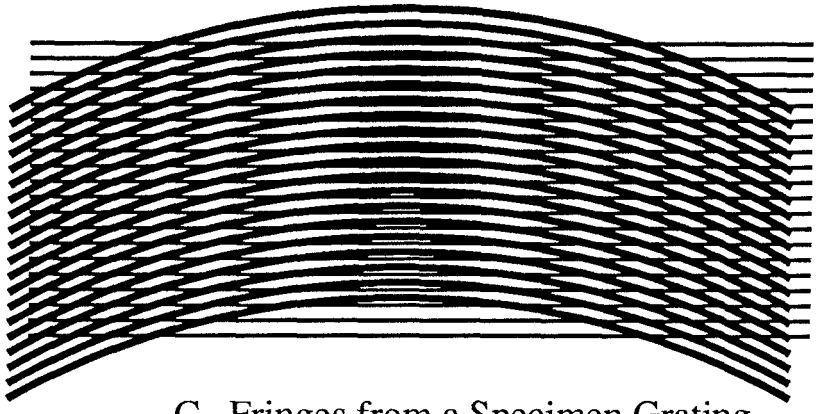
Some sample moire fringes are shown in Figure 4. In Figure 4 (A), are shown fringes formed by line gratings having a similar spacing at a slight relative rotation. Figure 4 (B) shows the phenomenon of fringe multiplication, as the reference grating of (A) is replaced by a reference grating of one-half the pitch, showing the resulting fringe between each pair of fringes visible in (A). Figure 4 (C) shows fringes formed with a specimen grating of circular arcs and a reference grating of straight lines. Note that in passing from one fringe to the next, the circular arcs travel one reference grating spacing in the direction normal to lines of the reference grating.



A. Geometric Moire Fringes



B. Fringe Doubling



C. Fringes from a Specimen Grating of Circular Arcs

FIGURE 4. **A.** An example of fringes formed using geometric moire.
B. Fringe doubling. The specimen grating is the same as in (**A**), but the reference grating has twice as many lines.
C. An example of fringes formed when the specimen grating is curved lines. The fringes still represent contours of displacement.

2.2.2 Moire Interferometry

Moire interferometry substitutes a diffraction grating for the specimen grating. In principle, a diffraction grating is identical to the gratings used in geometric moire. However, the spacing between the lines of a diffraction grating is on the order of the wavelength of light, which means that the wave interactions of the light with the grating become important. Optical effects on the scale of the wavelength of light are called diffraction effects, and diffraction gratings are useful because they have a regular structure. The most common application of diffraction gratings is in spectrum analysis; since the angle of diffraction varies with the wavelength of light, a grating can be used to separate the spectral components of a light source. The diffraction by gratings can be caused by two different mechanisms. The grating can cause local changes in the *amplitude* of the light waves, or it can cause local changes in the *phase* of the light waves. An amplitude grating varies the transmissivity or reflectivity of the diffraction grating periodically, such as by printing lines on a glass plate or on a mirror. A phase grating keeps the transmissivity or reflectivity constant, but varies the phase of the light waves periodically. In transmission this can be accomplished by varying the thickness of the transparent medium, while in reflection it can be done by varying the height of the reflective surface. The diffraction grating used in moire interferometry is a reflective phase grating, so the *lines* of the grating are in fact periodically alternating ridges and valleys, and the spacing of the lines is more naturally described in terms of the wavelength of the grating profile.

The fringes observed in moire interferometry are the result of interference between two laser beams diffracted from the specimen grating. Initially, before any deformation, the two laser beams are directed at the specimen so that they intersect at the specimen surface. The laser beams are arranged symmetrically so that each first–order diffracted beam exits along the normal to the specimen surface. In this initial position, and assuming perfect plane light waves and a perfect grating. the two diffracted beams overlap exactly so that the resulting interference pattern has a uniform intensity over the field of view. As the specimen and the attached diffraction grating deform, the laser beams are diffracted away at slightly different

angles, causing an interference pattern of light and dark fringes. The interference pattern obviously depends on how the diffraction grating has changed, but it is *not* obvious that the dependence of the fringe pattern on the deformation of the diffraction grating in moire interferometry is the same as the dependence of the fringe pattern on the deformation of the specimen grating in geometric moire.

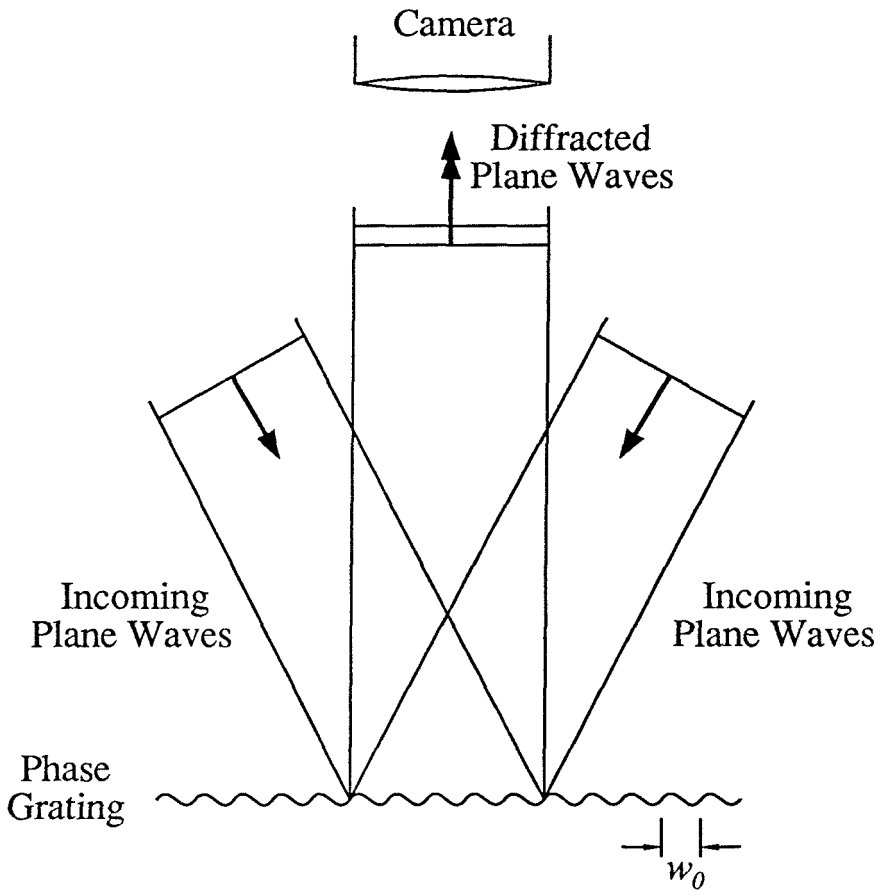


FIGURE 5. The initial arrangement of the laser beams for moire interferometry. The initial grating wavelength is w_0, and the diffracted plane waves travel to the camera in the same direction. In this initial arrangement, the diffracted waves have a constant phase difference over the field of view, leading to a *null field*, a field of constant intensity.

Two books by Guild [23,24] describe the use of interference patterns of light passing through consecutive diffraction gratings as measuring scales, and spell out the nature of the diffraction of plane light waves in three dimensions. That work

was carried out before the widespread availability of lasers; Guild's analysis involved the diffraction of light through two consecutive diffraction gratings, and the mutual interference between the different diffraction orders. These equations can also be employed to calculate the effect of the deformed grating geometry on the diffraction of the laser beams in moire interferometry, and to determine the relation between the deformation and the resulting interference fringe pattern.

Again, the first question in using moire interferometry is whether or not the resulting fringes are indeed exactly equivalent to the fringes observed in geometric moire techniques. The fringe patterns in geometric moire result from the overlap of the lines of the specimen and reference gratings, and the fact that the fringes represent contours of displacement is obvious from the geometry of the overlap. The fringe patterns in moire interferometry are the result of interference between two diffracted laser beams. Since the mechanism for the formation of the fringes in moire interferometry is very different from the mechanism in geometric moire, it is not obvious that the resulting fringe patterns correspond to parametric curves in moire interferometry. By using Guild's equations for the diffraction grating for each of the two intersecting laser beams, it can be shown that the equations relating the fringes to the displacement field are exactly the same in moire interferometry and geometric moire, as long as the plane of the grating remains parallel to the plane it occupied initially. This result was shown by the author in a GALCIT Solid Mechanics report in 1984 [72], and by Livnat and Post in 1985 in a slightly different way [41]. Pirodda [57] also showed the dependence of the fringes on the state of the grating using Fourier optics.

In the case of the moire interferometer, the equations describing the fringes are written in terms of the initial, undeformed orientation and wavelength of the grating (knowledge of which is retained by the fixed arrangement of the incident laser beams) and in terms of the the current or deformed orientation and wavelength of the grating. The grating is assumed to be a perfectly periodic surface height variation that has an initial wavelength of w_0 and a current (deformed) wavelength of w. The angle between the lines described by the current grating ridges relative to the lines along the initial grating ridges is γ. As with the geometric moire

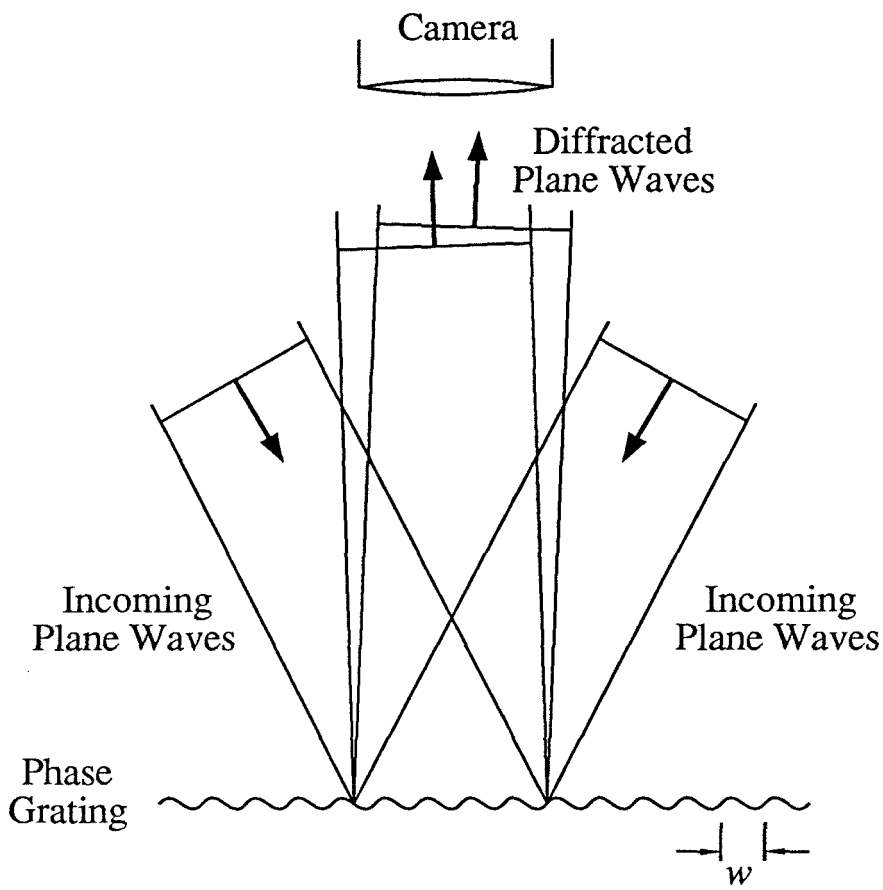


FIGURE 6. The grating has stretched to a new wavelength w, causing the diffracted planes waves to leave the surface at equal but opposite angles. The interference between the diffracted plane waves will vary across the field of view, forming light and dark fringes.

measurement, the deformation is assumed to be homogeneous. The fringe spacing δ is found to be

$$\delta = \frac{w w_0}{2\,[w^2 + w_0^2 - 2 w w_0 \cos\gamma]^{\frac{1}{2}}} \tag{2.2.5}$$

Note that there is a fringe multiplication factor of two in the denominator of equation (2.2.5), making equation (2.2.5) identical in form to equation (2.2.3), implying that the sensitivity of the measurement is equal to one–half the initial grating wavelength, $w_0/2$. The angle ϕ between the fringes and the ridges of the grating in its initial position is described by an equation identical in form to equations (2.2.2)

and (2.2.4), so that ϕ is is related to the initial and current grating parameters by

$$\tan\phi = \frac{w_0 \sin\gamma}{w_0 \cos\gamma - w} \tag{2.2.6}$$

The identity between equations (2.2.3) and (2.2.5) and between equations (2.2.4) and (2.2.6) indicates that the fringes generated in moire interferometry do represent contours of in-plane displacement. The displacement increment between adjacent fringes is $w_0/2$, and the direction of the displacement measured is perpendicular to the ridges of the diffraction grating in its initial position.

As mentioned before, the initial wavelength and orientation of the grating are retained in the alignment of the laser beams that are diffracted to produce the interference fringes observed in moire interferometry. These laser beams are aligned to the initial grating position and held fixed while the grating deforms. In an effort to make a strong, direct comparison between moire interferometry and geometric moire, Post suggests that the crossed laser beams in moire interferometry are equivalent to replacing the reference grating of geometric moire by a *virtual* reference grating formed by the interference pattern between the incoming laser beams [61,62,63]. This interference pattern (which should not be confused with the fringes seen that are due to the interference between the outgoing *diffracted* laser beams) consists of parallel light and dark bands. These bands have a spacing that depends on the wavelength of the laser light used and on the angle of intersection of the laser beams. † In aligning the laser beams so that the diffracted laser beams exit from the surface along the surface normal, a virtual grating is formed with a wavelength of $w_0/2$. Thus the analogy between moire interferometry and geometric moire is complete, and the fringe patterns of moire interferometry do represent contour maps of the in-plane displacement.

It is desirable to be able to view the fringes observed using the moire interferometer as contours of displacement, because then similar data reduction methods

† The absolute limit of the space between lines in the virtual reference grating is $\lambda/2$, where λ is the wavelength of the laser light used to form the virtual grating. This limit suggests that moire interferometry can have a theoretical maximum sensitivity of $\lambda/2$; that is, the displacement increment between adjacent dark fringes has a lower bound of $\lambda/2$. Note that techniques for measuring the out-of-plane displacement of a surface, such as holographic interferometry and Twyman-Green interferometry, also have a sensitivity of $\lambda/2$.

can be used on data generated from either the Twyman-Green interferometer or the moire interferometer. The fringes observed in a Twyman-Green interferometer represent contours of constant relative out-of-plane displacement, if the slope of the surface is sufficiently small. In the Twyman-Green interferometer, a path–length difference caused by out-of-plane deformation causes interference between two laser beams, one reflected from the specimen and one reflected from a flat reference mirror. Along a given fringe, the phase difference of the interfering light waves is the same, indicating that the path–length difference is also the same, indicating that the out-of-plane displacement is constant along that fringe.

2.2.3 Out–of–Plane Rotations

An implicit assumption that underlies the previously mentioned derivations [41,57,72] of the equations describing the fringes in moire interferometry is that the plane of the grating does not change during the deformation. While translation in any direction does not affect the measurement (unless it is so large that the grating is no longer in the path of the intersecting laser beams), rotations of the plane of the grating so that the grating is no longer parallel to its initial position will have an effect. The fringes observed in moire interferometry are generated by interfering light waves that intersect at a small angle. This angle varies with the in-plane displacement through the deformation of the diffraction grating on the surface of the specimen. However, the angle may also vary because of rotations of the surface out of the plane, as is the case in fracture problems, which may have sizable out-of-plane displacement gradients near the tip of the crack. These rotations will introduce additional terms in the equations describing the fringes, which, if significant, will mean that the fringes can no longer be interpreted as displacement contours. A similar effect is also seen in geometric moire if the two gratings are not parallel, creating fringes that are due to the wedge angle between them.

In the next section, the equations describing the fringes will be derived for cases that allow rotations that change the plane of the grating in an attempt to

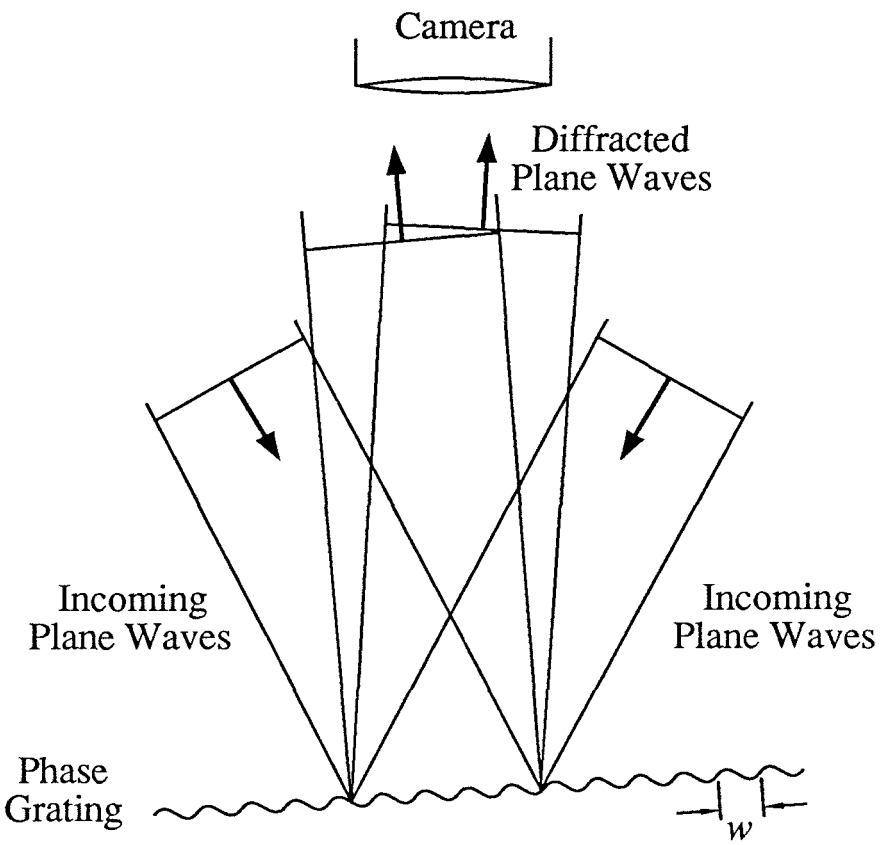


FIGURE 7. The grating has stretched to a new wavelength w, and has also tilted. The tilting may cause some additional deviation of the diffracted plane waves, which is not related to the in-plane deformation.

estimate the magnitude of the error introduced if the fringes are taken as contours of displacement. Alternatively, this investigation will help to determine how corrections could be made if the out-of-plane displacement gradients are known through some other method, as, for example, from measurements using the Twyman-Green interferometer.

2.3 The Governing Equations of Moire Interferometry Allowing for Arbitrary Changes in the Orientation of the Grating

In order to determine the effect of the out-of-plane deformation on the fringe pattern, the equations that describe the fringe spacing and orientation will be derived while allowing for an arbitrary tilting of the grating surface away from its original position. Since the rotation of the grating about a normal to the surface is a part of the in-plane deformation and has already been accounted for in previous investigations [41,57,72]; only rotations that cause the specimen surface to tilt in relation to its initial position need to be added to the equations.

This investigation is motivated by the desire to use moire interferometry to make sensitive measurements of the deformations at the tip of a crack. Moire interferometry relies on the rotation of two laser beams to cause interference fringes. The rotation of the laser beams can be caused by a tilting of the specimen surface as well as by deformation of the diffraction grating. Rotations of the specimen surface may occur either through a global rigid motion or through local out-of-plane displacement gradients. Rigid body rotations can be eliminated, but the deformation at the tip of a crack has a strong out-of-plane displacement gradient caused by the Poisson contraction resulting from the stress concentration at the crack tip. In some cases the angular shift in the diffracted laser beams may be dominated by the out-of-plane deformation, in which case it would be incorrect to assume that the fringes observed represent in-plane displacement contours.

2.3.1 Assumptions Underlying the Governing Equations

It is first assumed that the fringes in moire interferometry result from the interference between plane, coherent, monochromatic light waves, where a fringe is defined as a locus where the phase difference between the interfering light waves is constant. Since the light waves are assumed to be planes, each fringe is itself a plane, and all the fringe planes are parallel. The *fringe spacing* is the distance between planes where the phase difference between the light waves has changed by 2π. The fringe planes are arbitrarily oriented in space, but they are uniquely located in any coordinate frame if the normal vector to the fringe planes and the fringe spacing are determined.

The light waves that interfere are assumed to be monochromatic and to be coherent. It is assumed that the wave trains have a constant wavelength λ, although the physical situation is slightly more complicated. For a physical light source such as a laser, the wavelength changes slightly in time. If two sets of waves are coherent, any change in the wavelength of one set of waves is matched by a similar change in the wavelength of the second set of waves. In an experiment, coherence is assured by causing the two incoming laser beams to originate at the same source and to travel paths of equal length to the specimen. This is an example of the temporal coherence of two interfering waves. Another form of coherence is spatial coherence, a measure of the similarity of the light waves across the wave fronts. A laser source exhibits good coherence in both space and time. It will be assumed for this derivation that the light waves are exactly monochromatic and that they are identical across the wave fronts. The notion of coherence is ordinarily discussed in the framework of an experiment. In that context, the coherence between two wave trains of light is measured by the contrast of the interference between the two wave trains. All electromagnetic waves interfere, but wave trains having little coherence interfere on such a small time scale that the result is not usually observable. Coherence will be discussed further in Section 2.4.1, which deals with experimental apparatus.

The assumption that the diffracted light waves are plane waves is actually the result of assuming that the incoming light waves are perfect plane waves, that the

mean grating surface remains a plane, and that the grating surface profile remains perfectly periodic. Stroke [81] has shown that these conditions result in diffracted waves that are also plane waves. The actual grating profile is unrestricted, since the effect of different profiles is seen only in the distribution of the intensity among the various diffraction orders, as investigated by Petit *et al.*[55].The grating surface and the light waves are assumed to be infinite in extent so that there are no edge effects.

In addition to assuming that the light consists of plane waves, it is also assumed that the light is linearly polarized. This means that the electric–field vector corresponding to the plane wave can be written as a constant amplitude vector multiplied by a harmonic oscillation in space and time. Most lasers are linearly polarized. The most general case of elliptical polarization may reduce the contrast of the fringes, but the spatial distribution of the interference pattern is unchanged.

The grating plane is defined as the mean grating surface. The assumption that the grating plane remains plane and the grating profile remains periodic is equivalent to assuming that the deformation at the surface is homogeneous, since a homogeneous deformation maps parallel lines to parallel lines. The restriction of the deformation to homogeneous deformations will have to be relaxed through a localization argument for general deformations.

The equations that relate the directions of the diffracted beams to the direction of the incoming beam have been given by Guild [23,24]. These equations are used to determine the direction of the diffracted plane waves of any diffraction order from an incident plane wave that is directed at the grating from an arbitrary angle. The use of these equations requires the resolution of the directions of the beams into an orthogonal coordinate frame defined by the ridges of the grating profile and the plane of the grating.

The parameters of interest in the derivation are the wavelength of the deformed grating profile, the orientation of the grating plane after deformation, and the direction of the ridges of the grating after deformation. The grating will be assumed to be stretched from its initial wavelength of w_0 to a new wavelength of w, after

which it is rigidly rotated to a new orientation. The orientation of the grating plane after deformation will be related to the grating plane before deformation, through Euler angles. These angles represent a series of rigid rotations about the axes of a coordinate frame that rotates with the deformed grating.

The observer views the specimen along the normal to the undeformed specimen surface. This is a fixed observation position, which is part of the experimental arrangement. The fringe patterns are recorded by a camera focused on the initial position of the grating, and the camera maps the initial grating plane to the film plane. Since the fringe planes are arbitrarily oriented in space, they may not intersect the initial grating plane at a right angle. If this is true, the observed fringe spacing will not coincide with the actual fringe spacing. To appreciate this statement, think of placing a screen in the volume occupied by the fringes. The intensity distribution on the screen will depend on the orientation of the screen with respect to the fringe planes. If the screen is parallel to the fringe planes, the observed intensity will be constant, so that the observed fringe spacing is infinite, while the true fringe spacing is finite. This effect must also be taken into consideration.

The absolute phase information of the light waves may be left arbitrary, since the parameters of interest are the fringe spacing and orientation, which are unaffected by a change in the phase offset between the two interfering light waves. (Such a change in the phase offset might be accomplished by changing the length of the path taken by one of the incoming laser beams from the source to the specimen.) In other words, whether the fringe is light or dark is not important as long as the fringe spacing and inclination remain the same. This statement emphasizes one of the important facts about interferometry in general, namely, that the displacements measured at a point are known only in relation to the displacements at other points in the field of view. In order to make absolute measurements, the absolute displacement of one point in the field must be known from some other method.

The incoming laser beams are arranged to intersect the grating in its initial position in such a way that the diffracted beam of order one (or minus one) leaves the specimen surface along the normal to the grating plane. Since the phase dif-

ference between the diffracted plane waves is constant over the field of view, this initial configuration produces a *null field*, a field of uniform intensity. This initial condition determines the angle at which the incoming laser beams must intersect the specimen surface; the angle is a function of the initial grating wavelength w_0 and the wavelength of light λ.

This derivation of the equations describing the fringes in moire interferometry is similar to that by Livnat and Post [41] in the use of the vector normal to describe the fringe planes, although they restrict their derivation to planar deformation. The present approach was derived independently by the author, preceding the publication of Reference [41]. The work by Livnat and Post offers some interesting comparisons between moire interferometry and geometric moire methods. The equations for the case of planar deformation and the correspondence between moire interferometry and geometric moire were also derived earlier in a slightly different way by the author [72] and independently in a very different manner by Pirodda [57]. The paper by Pirodda contains interesting descriptions of several interferometric measurement techniques.

2.3.2 Plane Harmonic Monochromatic Light Waves

The diffracted laser beams are assumed to consist of plane harmonic waves having a single wavelength λ. For plane harmonic electromagnetic waves, the electric and magnetic fields are closely related to each other, so that the plane waves can be described by either the electric–field vector $\boldsymbol{\mathcal{E}}$ or the magnetic–field vector $\mathbf{H}$. In this case, we will use the electric–field vector $\boldsymbol{\mathcal{E}}$.

In free space, the electric–field vector $\boldsymbol{\mathcal{E}}$ satisfies the wave equation

$$\nabla^2 \boldsymbol{\mathcal{E}} = \frac{1}{c^2} \frac{\partial^2 \boldsymbol{\mathcal{E}}}{\partial t^2}, \tag{2.3.1}$$

where c is the speed of light. The general solution for plane waves having unit normal $\mathbf{u}$ is

$$\boldsymbol{\mathcal{E}} = \mathbf{f}_1(\mathbf{r} \cdot \mathbf{u} - ct) + \mathbf{f}_2(\mathbf{r} \cdot \mathbf{u} + ct), \tag{2.3.2}$$

where $\mathbf{f}_1$ and $\mathbf{f}_2$ are any twice continuously differentiable vector functions, and $\mathbf{r}$ represents the position vector of any point. The first term in 2.3.2 represents waves traveling in the direction of $\mathbf{u}$, while the second term represents waves traveling in the direction of $-\mathbf{u}$.

For waves that are harmonic in space and time, with a wavelength λ, and that are propagating in the direction of $\mathbf{u}$, the solution can be written explicitly as

$$\mathcal{E}^{(\mathrm{u})} = Re\left\{ \mathbf{A}^{(\mathrm{u})} \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{u} - ct) + i\phi^{(\mathrm{u})} \right] \right\}, \tag{2.3.3}$$

where $\mathbf{A}^{(\mathrm{u})}$ is a constant vector, possibly complex, and $\phi^{(\mathrm{u})}$ is a real, constant phase offset. The superscript (u) is used to indicate that the electric–field vector and the amplitude and phase offset are associated with plane waves traveling in the direction of $\mathbf{u}$.

The *intensity*, I, of the light is a measure of the time average of the energy carried by the electromagnetic waves. Following Born and Wolf [11] (or Fowles [19]), it is found that for plane waves the intensity is proportional to the square of the magnitude of the electric–field vector. Since we are interested only in the spatial distribution of the intensity, the proportionality constant can be taken as unity, so that the intensity is found as

$$I = \frac{1}{2T}\int_{-T}^{T} [\mathcal{E}\cdot\mathcal{E}]\, dt \tag{2.3.4}$$

where T is very long compared with the time period of the oscillation of the waves. We can incorporate all of the complex exponentials except for the time dependence into the amplitude vector so that $\mathbf{E}$ is given by

$$\begin{aligned} \mathcal{E} = Re\left\{ \mathcal{E}_0 \exp\left[\frac{-i2\pi}{\lambda} ct \right] \right\} &= Re\left\{ \mathcal{E}_0 \exp(-i2\pi\nu t) \right\} \\ &= \frac{1}{2}\left[\mathcal{E}_0 \exp(-i2\pi\nu t) + \overline{\mathcal{E}_0} \exp(i2\pi\nu t) \right] \end{aligned} \tag{2.3.5}$$

Here $\overline{\mathcal{E}_0}$ indicates the complex conjugate of $\mathcal{E}_0$. The quantity ν is defined by $c = \lambda\nu$. The integral of equation (2.3.4) can then be evaluated as

$$I = \frac{1}{8T}\int_{-T}^{T} \left[(\mathcal{E}_0\cdot\mathcal{E}_0)\exp(-4i\pi\nu t) + 2(\mathcal{E}_0\cdot\overline{\mathcal{E}_0}) + (\overline{\mathcal{E}_0}\cdot\overline{\mathcal{E}_0})\exp(4i\pi\nu t) \right] dt \tag{2.3.6}$$

The second term in equation (2.3.6), after integration and division by $8T$, yields a constant of $(\mathcal{E}_0 \cdot \overline{\mathcal{E}_0})/2$. The complex vector quantities $\mathcal{E}_0$ and $\overline{\mathcal{E}_0}$ are independent of time, so they may be taken out of the integrals, and the only remaining terms to integrate are the $\exp(-4i\pi\nu t)$ and $\exp(4i\pi\nu t)$ terms. Now

$$\frac{1}{8T}\int_{-T}^{T} \exp(-4i\pi\nu t)dt = \frac{-1}{32i\pi\nu T}\Big[\exp(-4i\pi\nu t)\Big]_{-T}^{T} = \frac{1}{16\pi\nu T}\sin 4\pi\nu T \quad (2.3.7)$$

Similarly,

$$\frac{1}{8T}\int_{-T}^{T} \exp(4i\pi\nu t)dt = \frac{1}{32i\pi\nu T}\Big[\exp(4i\pi\nu t)\Big]_{-T}^{T} = \frac{1}{16\pi\nu T}\sin 4\pi\nu T \quad (2.3.8)$$

The period of oscillation of the waves, $\widehat{T}$, is just $1/\nu$, so $(1/16\pi\nu T) = (1/16\pi)(\widehat{T}/T)$. Since T is assumed to be very long compared to $\widehat{T}$, the first and third terms of equation (2.3.6) do not contribute to the average, and

$$I = \frac{1}{2}\Big[\mathcal{E}_0 \cdot \overline{\mathcal{E}_0}\Big] \quad (2.3.9)$$

Returning to the electric–field vector of equation (2.3.3), which describes harmonic plane waves traveling in the direction of unit vector $\mathbf{u}$, the electric field is given by

$$\mathcal{E}^{(\mathrm{u})} = Re\left\{\mathbf{A}^{(\mathrm{u})}\exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{u} - ct) + i\phi^{(\mathrm{u})}\right]\right\} \quad (2.3.3)$$

The intensity of these plane waves is found from equation (2.3.9), which implies that all of the exponential terms drop out, so that the intensity of the plane waves in the direction of $\mathbf{u}$ is found as

$$I^{(\mathrm{u})} = \frac{1}{2}\big[\mathbf{A}^{(\mathrm{u})} \cdot \overline{\mathbf{A}}^{(\mathrm{u})}\big] \quad (2.3.10)$$

If the light described by equation (2.3.3) is linearly polarized, the vector amplitude $\mathbf{A}^{(\mathrm{u})}$ is a constant, real–valued vector. The light from the laser used in the experiments is strongly linearly polarized, and we will assume now that both diffracted laser beams are linearly polarized for simplicity. The general case of elliptically polarized beams will be discussed briefly in Appendix B for completeness. For the electric–field vector of equation (2.3.3), assuming that $\mathbf{A}^{(\mathrm{u})}$ is a constant, real–valued vector, the intensity is given by

$$I^{(\mathrm{u})} = \frac{1}{2}\big[\mathbf{A}^{(\mathrm{u})} \cdot \mathbf{A}^{(\mathrm{u})}\big] \quad (2.3.11)$$

2.3.3 Interference of Plane Harmonic Monochromatic Light Waves

We are interested in the interference pattern between two sets of plane waves which cross each other at an angle. These two sets of waves are diffracted from the specimen surface. The direction of the diffracted waves, which are represented by the unit normals to the waves, can be determined from the direction of the incoming light waves and the diffraction–grating wavelength and orientation, as will be shown in the next section. In this section we will determine the dependence of interference pattern on the diffracted beams.

The electric–field vector corresponding to the first set of waves, which have unit normal $\mathbf{u}$, is given by equation (2.3.3). The electric–field vector corresponding to the second set of waves, which have unit normal $\mathbf{v}$, is given by a similar expression. The electric–field vectors for both sets of waves are thus

$$\begin{aligned} \mathcal{E}^{(\mathrm{u})} &= Re\left\{ \mathbf{A}^{(\mathrm{u})} \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{u} - ct) + i\phi^{(\mathrm{u})} \right] \right\} \\ \mathcal{E}^{(\mathrm{v})} &= Re\left\{ \mathbf{A}^{(\mathrm{v})} \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{v} - ct) + i\phi^{(\mathrm{v})} \right] \right\} \end{aligned} \tag{2.3.12}$$

The superscript (u) corresponds to the waves having unit normal $\mathbf{u}$, and the superscript (v) corresponds to the waves having unit normal $\mathbf{v}$.

The plane waves described in equation (2.3.12) are assumed to be linearly polarized, so that the amplitude vectors $\mathbf{A}^{(\mathrm{u})}$ and $\mathbf{A}^{(\mathrm{v})}$ are real–valued constant vectors. The constant phase offsets $\phi^{(\mathrm{u})}$ and $\phi^{(\mathrm{v})}$ are a function of the path length taken by the laser beam from the source. The amplitude vectors and the phase offsets will be left as arbitrary constants.

The total electric–field vector for both sets of waves is given by the summation of the two terms $\mathcal{E}^{(\mathrm{u})}$ and $\mathcal{E}^{(\mathrm{v})}$.

$$\mathcal{E}^{(total)} = Re\left\{ \mathbf{A}^{(\mathrm{u})} \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{u} - ct) + i\phi^{(\mathrm{u})} \right] + \mathbf{A}^{(\mathrm{v})} \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{v} - ct) + i\phi^{(\mathrm{v})} \right] \right\} \tag{2.3.13}$$

The total intensity that is due to both sets of waves is then found using equation

(2.3.9) as

$$2I^{(total)} = |\mathbf{A}^{(u)}|^2 + |\mathbf{A}^{(v)}|^2 + (\mathbf{A}^{(u)} \cdot \mathbf{A}^{(v)}) \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r} \cdot \mathbf{u} - ct) + i\phi^{(u)}\right] \exp\left[\frac{-i2\pi}{\lambda}(\mathbf{r} \cdot \mathbf{v} - ct) - i\phi^{(v)}\right] + (\mathbf{A}^{(u)} \cdot \mathbf{A}^{(v)}) \exp\left[\frac{-i2\pi}{\lambda}(\mathbf{r} \cdot \mathbf{u} - ct) - i\phi^{(u)}\right] \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r} \cdot \mathbf{v} - ct) + i\phi^{(v)}\right] \tag{2.3.14}$$

The time–dependent contribution cancels, and the exponentials can be rewritten as a real–valued cosine function. The total intensity is thus given by

$$2I^{(total)} = |\mathbf{A}^{(u)}|^2 + |\mathbf{A}^{(v)}|^2 + 2(\mathbf{A}^{(u)} \cdot \mathbf{A}^{(v)}) \cos\left[\frac{2\pi}{\lambda}(\mathbf{r} \cdot (\mathbf{u} - \mathbf{v})) + (\phi^{(u)} - \phi^{(v)})\right] \tag{2.3.15}$$

The first term in equation (2.3.15) represents the intensity that is due only to the waves traveling in the direction defined by $\mathbf{u}$, the second term represents the intensity that is due only to the waves in the direction defined by $\mathbf{v}$, and the third term represents the interference between the two waves.

The intensity distribution of equation (2.3.15) represents a standing wave pattern. This standing wave has planes of constant phase perpendicular to the vector $\mathbf{u} - \mathbf{v}$. The vector $\mathbf{z}$ will be defined so that

$$\mathbf{z} = \mathbf{u} - \mathbf{v} \tag{2.3.16}$$

The fringe spacing is δ, which is equal to the wavelength of the standing wave pattern, so that δ is found to be

$$\delta = \frac{\lambda}{|\mathbf{u} - \mathbf{v}|} = \frac{\lambda}{|\mathbf{z}|} \tag{2.3.17}$$

The interference is shown schematically in Figure 8. Since $\mathbf{u}$ and $\mathbf{v}$ are unit vectors, the fringe spacing can vary from infinity, when $\mathbf{u} = \mathbf{v}$, to $\lambda/2$, when $\mathbf{u} = -\mathbf{v}$. The fringe spacing can also be related to the angle between the two unit normals, $\mathbf{u}$ and $\mathbf{v}$. Let 2β be the angle between the unit normals; then the fringe spacing is given by

$$\delta = \frac{\lambda}{2|\sin\beta|} \tag{2.3.18}$$

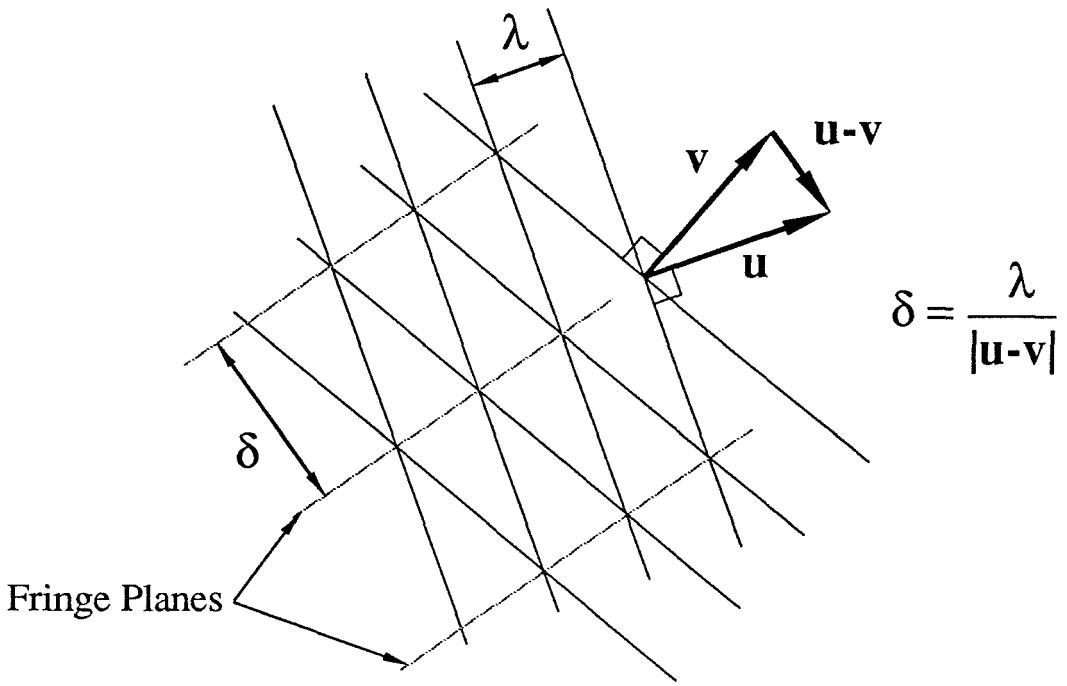


FIGURE 8. The standing interference pattern formed between two sets of plane waves crossing at an angle. The first set of waves has unit normal $\mathbf{u}$, and the second set has unit normal $\mathbf{v}$. The light has a wavelength of λ. The interference pattern has a wavelength $\delta = \lambda/|\mathbf{u} - \mathbf{v}|$.

This arrangement is shown in Figure 9.

The contrast of the fringes is defined by

$$\text{contrast} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{2.3.19}$$

The contrast depends on the interference term in equation (2.3.15), and is found to be

$$\text{contrast} = \frac{2|(\mathbf{A}^{(u)} \cdot \mathbf{A}^{(v)})|}{|\mathbf{A}^{(u)}|^2 + |\mathbf{A}^{(v)}|^2} \tag{2.3.20}$$

Equation (2.3.20) shows the importance of matching the polarization and intensity of the two diffracted laser beams. If the two sets of waves are polarized at 90° to each other, the inner product of the third term in equation (2.3.15) is zero, and no interference is seen. Similarly, if the intensity of one beam is very much greater than that of the other, the interference term will be much less than the

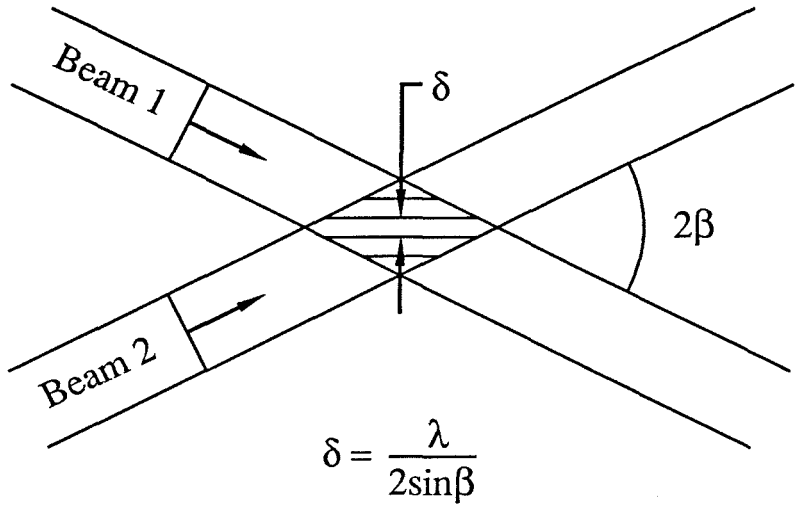


FIGURE 9. Standing interference pattern formed between collimated laser beams crossing at angle 2β. The light has a wavelength of λ. The interference pattern has a wavelength δ.

intensity of the brighter beam, and the interference will be washed out. In either case, the contrast of the fringes is reduced.

If the light is elliptically polarized, which is the most general state of polarization, the fringe contrast may be reduced, depending on the relative direction of the rotating electric–field vector. However, the spatial distribution of the fringes is unchanged, so that the fringes will still lie in planes perpendicular to the vector difference $\mathbf{z} = \mathbf{u} - \mathbf{v}$, and the fringe spacing will still be $\delta = \lambda/|\mathbf{z}|$ See Appendix B for more detail regarding interference when the light is elliptically polarized.

2.3.4 Equations of Diffraction

The effect of a diffraction grating is to break up an incoming plane wave into a series of periodic sources, where the period is determined by the grating wavelength, w. Although the sources emit spherical waves, their sum consists of plane waves. This can be seen in Figure 10, which shows the construction of two plane waves from five point sources in two dimensions. The direction of the plane waves depends on the distance w between sources, the wavelength λ of light, and the phase difference between the sources. The phase difference is determined from the angle of the incoming plane wave, which has a constant phase across its surface.

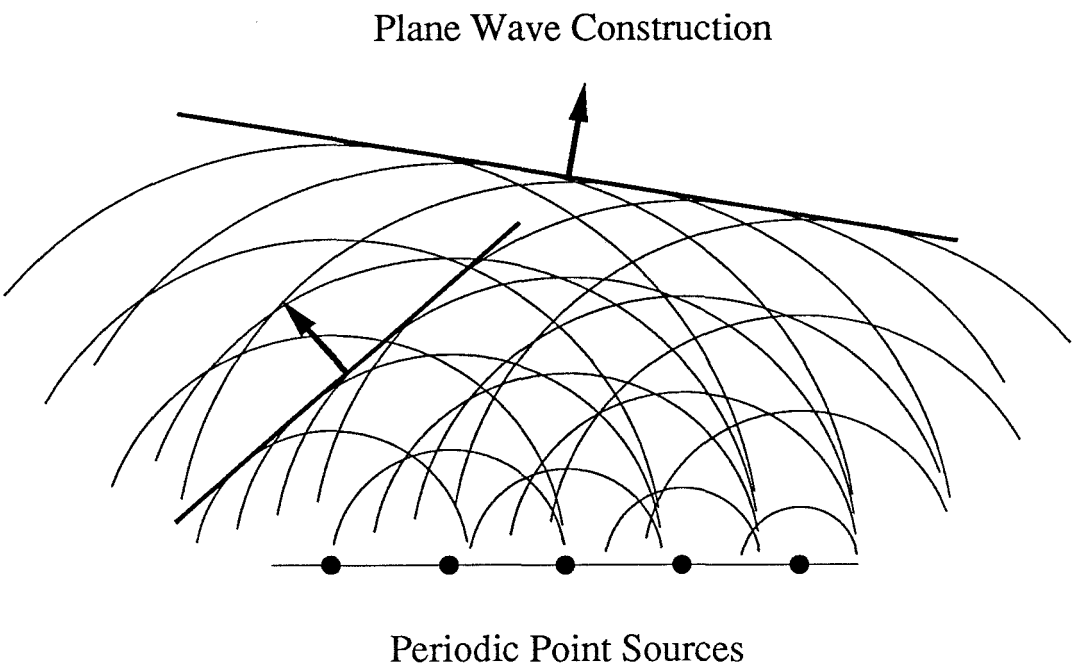


FIGURE 10. The formation of plane waves from periodic point sources radiating spherical waves. The phase of the point sources varies linearly. A diffraction grating can be thought of as a device to break up incoming plane waves into a series of point sources, leading to diffracted plane waves at angles determined by the period of the emitters, the phase variation of the emitters, and the wavelength of light.

A sketch of the effect of the diffraction grating is shown in Figure 11. In the figure, α is positive as shown, and β is negative. The requirement that the outgoing waves be in phase is equivalent to specifying that the length along path A differs

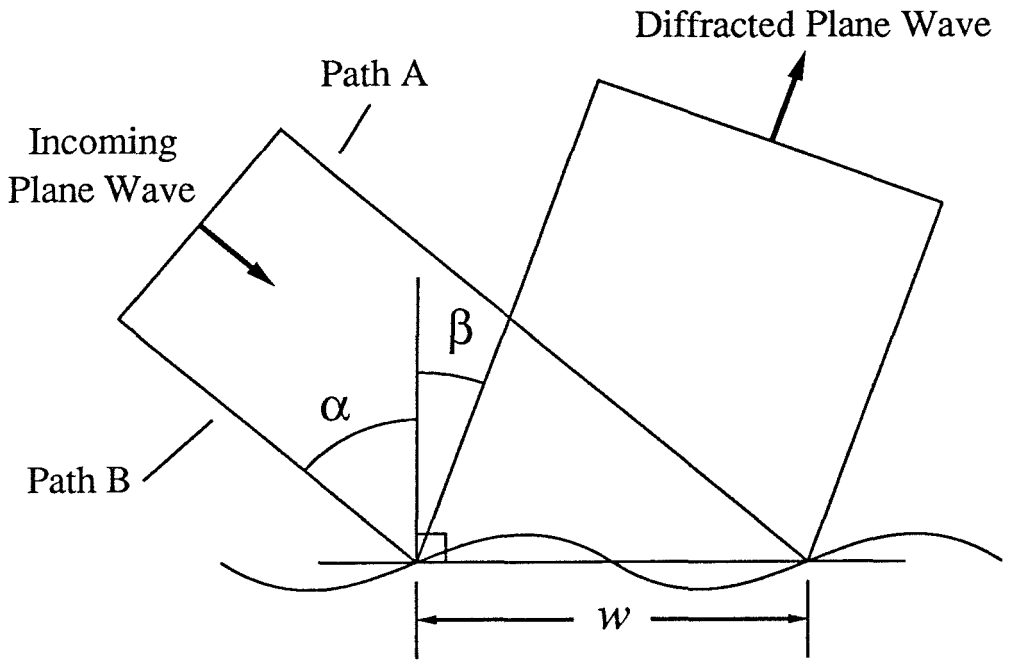


FIGURE 11. Plane waves incident on a grating result in diffracted plane waves at various angles. These angles correspond to directions where the length along path A differs from the length along path B by an integral number of wavelengths, say $n\lambda$. The angle α as shown is positive; the angle β is negative. Angles α and β are related by $\sin\alpha + \sin\beta = n\lambda/w$.

from the length along path B by an integral number of wavelengths, say $n\lambda$, where n is an integer. The number n is also called the *diffraction order*, so that in the first–order diffracted beam, the paths differ by one wavelength. The angle between the grating normal and the direction of the incoming light waves is α, and the angle between the grating normal and the direction of the diffracted light waves is β. The angle β is related to the angle α by the equation

$$w\sin\alpha + w\sin\beta = n\lambda \tag{2.3.21}$$

which is more commonly written as

$$\sin\alpha + \sin\beta = \frac{n\lambda}{w} \tag{2.3.22}$$

The case of reflection is recovered when $n = 0$, or when w tends to infinity.

The diffraction grating used in moire interferometry is a reflective diffraction grating, called a phase grating. A phase grating is simply a periodic height variation, arranged in peaks and valleys. Ordinarily, the height variation is much smaller than the distance between peaks, while the spacing between peaks, the grating wavelength, is on the order of the wavelength of light. The details of the diffraction are more complicated for the phase grating than for the series of point sources. The phase grating creates a continuous surface of point sources, with a phase variation that depends in detail on the grating profile, but the overall periodicity of the grating is sufficient to ensure plane wave solutions at the angle predicted by the simple point source model. The diffraction–grating surface introduces a periodic boundary condition to the wave equation describing the electric or magnetic field [81]. The way the energy of the incoming light is distributed into the various diffraction orders depends on the details of the shape of the diffraction–grating profile.

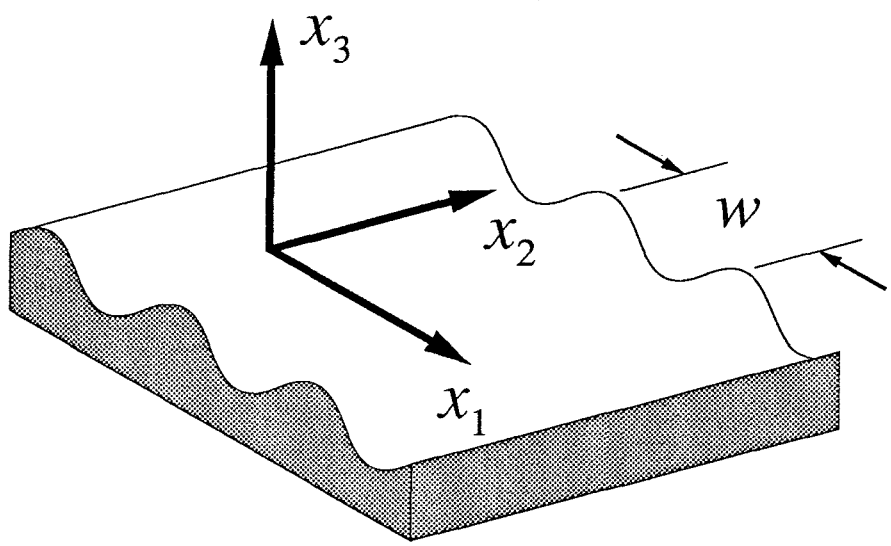


FIGURE 12. The Grating Coordinate Frame. The x_2 coordinate axis remains parallel to the ridges of the grating during deformation. The x_3 axis remains normal to the mean grating surface. The current grating wavelength is w.

The diffraction of plane light waves will be determined using the equations

of Guild [23,24]. These equations generalize the relation between the direction of the incoming light waves and the direction of the diffracted light waves to three dimensions. The directions of the various diffraction orders are determined in an orthogonal coordinate system that is defined by the ridges and valleys of the grating, as shown in Figure 12. This coordinate frame is arranged so that the x_2 axis is always parallel to the ridges of the grating, the x_1 axis is always perpendicular to the ridges of the grating, and the x_3 axis always remains normal to the mean grating surface, which has been assumed to remain a plane throughout any deformation. The current or deformed grating has a wavelength of w. The initial wavelength will be denoted w_0.

To calculate the direction of propagation of diffracted plane waves of any diffraction order from incident plane light waves which are directed at the grating from an arbitrary angle, the unit normal to the incoming light waves must be resolved into components with respect to the coordinate frame described above. The components of the unit normal to the diffracted plane wave can then be calculated in the grating coordinate frame.

For instance, if unit vector $\mathbf{j}$ represents the normal of the incident plane waves and unit vector $\mathbf{u}^{(n)}$ represents the normal to the diffracted light waves of order n, the equations relating the components of $\mathbf{j}$ to the components of $\mathbf{u}^{(n)}$ are:

$$u_1^{(n)} = \frac{n\lambda}{w} + j_1 \tag{2.3.23}$$

$$u_2^{(n)} = j_2 \tag{2.3.24}$$

where j_α and $u_\alpha^{(n)}$, $(\alpha = 1, 2)$ are the components of $\mathbf{j}$ and $\mathbf{u}^{(n)}$ in the grating coordinate frame, and λ is the wavelength of light used.

The component of the unit vector $\mathbf{u}^{(n)}$ that is normal to the grating plane, namely, $u_3^{(n)}$, is nonnegative, and since $\mathbf{u}^{(n)}$ is a unit vector, equations (2.3.23) and (2.3.24) uniquely determine $\mathbf{u}^{(n)}$ from any given $\mathbf{j}$. Since the waves defined by $\mathbf{j}$ are directed toward the grating, the component j_3 is negative.

2.3.5 Geometry Relating the Deformed and Undeformed Gratings

For moire interferometry, it is necessary to calculate the directions of diffracted beams originating from two different incident beams. This is necessary so that we may determine the fringe pattern resulting from interference between the two diffracted beams. The incident beams are fixed in space, while the grating is allowed to undergo an arbitrary change in orientation and ridge spacing.

The deformed grating itself defines a coordinate frame, as shown in Figure 12. The grating x_1 axis is in the grating plane, perpendicular to the ridges of the grating. The x_2 axis is also in the grating plane, but parallel to the ridges of the grating. The x_3 axis is normal to the grating plane.

As has been mentioned, moire interferometry essentially compares the current deformed grating to its initial, undeformed configuration. The configuration of the undeformed grating is retained in the alignment of the laser beams, which remain fixed throughout the deformation. A second coordinate frame, which is fixed in space, will be defined by the initial position of the grating. The position of the coordinate frame of the deformed grating will be described with respect to the fixed coordinate frame through a series of rotations using Euler angles.

The fixed frame will be denoted by capital letters, so that the axes are given by X_1, X_2, X_3. The grating coordinate system will be given using small letters, x_1, x_2, x_3 as shown in Figure 12, and the grating system is aligned with the fixed system before deformation.

As the grating deforms, the ridges of the grating shift, so that the grating coordinate frame no longer coincides with the fixed frame. Additionally, the plane of the grating is allowed to rotate with respect to the fixed axes. The deformation of the grating is homogeneous, and it can be described by first letting the grating stretch or contract from its initial wavelength of w_0 to the deformed wavelength w. The change in the orientation of the grating can then be described through the use of Euler angles. First, let the grating rotate by an angle γ_3 about the x_3 (or X_3) axis. Next, let the grating rotate an angle γ_2 about the new position of the grating

x_2 axis. Finally, let the grating rotate an angle γ_1 about the new position of the x_1 axis to the final position of the grating.

The relation between the components of any vector **a** resolved in the $X_1X_2X_3$ coordinate frame and the components of **a** resolved in the final position of the $x_1x_2x_3$ (grating) coordinate system are determined through the successive matrix operation:

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma_1 & \sin\gamma_1 \\ 0 & -\sin\gamma_1 & \cos\gamma_1 \end{pmatrix} \begin{pmatrix} \cos\gamma_2 & 0 & -\sin\gamma_2 \\ 0 & 1 & 0 \\ \sin\gamma_2 & 0 & \cos\gamma_2 \end{pmatrix} \begin{pmatrix} \cos\gamma_3 & \sin\gamma_3 & 0 \\ -\sin\gamma_3 & \cos\gamma_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \qquad (2.3.25)$$

where a_1, a_2, a_3 are the components of **a** in the $x_1x_2x_3$ frame and A_1, A_2, A_3 are the components of **a** in the $X_1X_2X_3$ frame.

This transformation is necessary, because the diffraction equations are written in terms of components in the deformed grating frame, while the incoming laser beams are specified in components in the fixed frame.

2.3.6 Moire Interferometry

So far, the discussion of the interference patterns has been quite general. The key point in moire interferometry is the alignment of the incident laser beams to the initial position of the grating. We now need to describe the arrangement of the fixed paths of the incident, plane light waves in moire interferometry, and resolve the components of their unit normals into the deformed grating frame using equation (2.3.25). The directions of the diffracted plane waves can be determined using equations (2.3.23) and (2.3.24). The resulting interference pattern can then be found from equation (2.3.17).

The plane light waves traveling in the directions specified by unit vectors **u** and **v** are diffracted from two *fixed*, incident light beams. The two fixed, incident beams of plane waves have unit normals denoted by **j** and **k**. These vectors are

directed toward the initial grating surface, and are arranged symmetrically with respect to the normal to the undeformed grating plane. These two unit vectors have components J_i and K_i, $i = 1, 2, 3$, respectively, in the globally fixed $X_1X_2X_3$ frame, and components j_i and k_i, respectively, in the $x_1x_2x_3$ frame defined by the deformed grating.

Let the unit vector **u** represent the diffracted ray of order 1 from the incident ray represented by **j**, and let unit vector **v** represent the diffracted ray of order -1 from the incident ray represented by **k**. These vectors have components u_i and v_i, $i = 1, 2, 3$, respectively, in the grating fixed $x_1x_2x_3$ frame. Then from equations (2.3.23) and (2.3.24), and using the fact that **u** and **v** are unit vectors, the components of **u** and **v** in the grating coordinate frame are found as

$$\begin{aligned} u_1 &= \frac{\lambda}{w} + j_1 \\ u_2 &= j_2 \\ u_3 &= [1 - (u_1)^2 - (u_2)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.26}$$

$$\begin{aligned} v_1 &= -\frac{\lambda}{w} + k_1 \\ v_2 &= k_2 \\ v_3 &= [1 - (v_1)^2 - (v_2)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.27}$$

In the experimental arrangement used in this study, the two fixed, incident rays represented by **j** and **k** lie in the X_1–X_3 plane and are symmetric with respect to the X_2–X_3 plane. In the global $X_1X_2X_3$ coordinate frame, these vectors have a simple representation:

$$\begin{Bmatrix} J_1 \\ J_2 \\ J_3 \end{Bmatrix} = \begin{Bmatrix} -\sin\alpha \\ 0 \\ -\cos\alpha \end{Bmatrix}, \qquad \begin{Bmatrix} K_1 \\ K_2 \\ K_3 \end{Bmatrix} = \begin{Bmatrix} \sin\alpha \\ 0 \\ -\cos\alpha \end{Bmatrix} \tag{2.3.28}$$

This representation indicates that the unit normals to the incident plane waves form angles of $\pm\alpha$ to the X_3 axis.

The components of **j** and **k** in the $x_1x_2x_3$ frame are then found, using equation (2.3.25), as

$$\begin{Bmatrix} j_1 \\ j_2 \\ j_3 \end{Bmatrix} = \begin{Bmatrix} -\sin\alpha\cos\gamma_3\cos\gamma_2 + \cos\alpha\sin\gamma_2 \\ \sin\alpha\sin\gamma_3\cos\gamma_1 - \sin\alpha\cos\gamma_3\sin\gamma_2\sin\gamma_1 - \cos\alpha\cos\gamma_2\sin\gamma_1 \\ -\sin\alpha\sin\gamma_3\sin\gamma_1 - \sin\alpha\cos\gamma_3\sin\gamma_2\cos\gamma_1 - \cos\alpha\cos\gamma_2\cos\gamma_1 \end{Bmatrix} \tag{2.3.29}$$

$$\begin{Bmatrix} k_1 \\ k_2 \\ k_3 \end{Bmatrix} = \begin{Bmatrix} \sin\alpha\cos\gamma_3\cos\gamma_2 + \cos\alpha\sin\gamma_2 \\ -\sin\alpha\sin\gamma_3\cos\gamma_1 + \sin\alpha\cos\gamma_3\sin\gamma_2\sin\gamma_1 - \cos\alpha\cos\gamma_2\sin\gamma_1 \\ \sin\alpha\sin\gamma_3\sin\gamma_1 + \sin\alpha\cos\gamma_3\sin\gamma_2\cos\gamma_1 - \cos\alpha\cos\gamma_2\cos\gamma_1 \end{Bmatrix} \tag{2.3.30}$$

By using equations (2.3.26) and (2.3.27), the components of the unit vectors **u** (which represents the diffracted ray of order 1 from the incident ray represented by **j**) and **v** (which represents the diffracted ray of order -1 from the incident ray represented by **k**) in the grating coordinate frame are found as

$$\begin{aligned} u_1 &= \frac{\lambda}{w} - \sin\alpha\cos\gamma_3\cos\gamma_2 + \cos\alpha\sin\gamma_2 \\ u_2 &= \sin\alpha\sin\gamma_3\cos\gamma_1 - \sin\alpha\cos\gamma_3\sin\gamma_2\sin\gamma_1 - \cos\alpha\cos\gamma_2\sin\gamma_1 \\ u_3 &= [1-(u_1)^2-(u_2)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.31}$$

$$\begin{aligned} v_1 &= -\frac{\lambda}{w} + \sin\alpha\cos\gamma_3\cos\gamma_2 + \cos\alpha\sin\gamma_2 \\ v_2 &= -\sin\alpha\sin\gamma_3\cos\gamma_1 + \sin\alpha\cos\gamma_3\sin\gamma_2\sin\gamma_1 - \cos\alpha\cos\gamma_2\sin\gamma_1 \\ v_3 &= [1-(v_1)^2-(v_2)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.32}$$

The components of the difference vector $\mathbf{z} = \mathbf{u} - \mathbf{v}$, which characterizes the fringe pattern in the grating coordinate frame are then found as

$$\begin{aligned} z_1 &= 2\left(\frac{\lambda}{w} - \sin\alpha\cos\gamma_3\cos\gamma_2\right) \\ z_2 &= 2(\sin\alpha\sin\gamma_3\cos\gamma_1 - \sin\alpha\cos\gamma_3\sin\gamma_2\sin\gamma_1) \\ z_3 &= [1-(u_1)^2-(u_2)^2]^{\frac{1}{2}} - [1-(v_1)^2-(v_2)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.33}$$

As was shown earlier, the fringe planes are perpendicular to the vector $\mathbf{z}$ with a spacing δ given by

$$\delta = \frac{\lambda}{|\mathbf{z}|} \tag{2.3.34}$$

2.3.7 Initial Alignment of the Laser Beams

The initial configuration of the grating is retained by the fixed, incident laser beams, which must be aligned to that configuration. Before any deformation occurs, while the grating axes and the global axes coincide, the angle between the normal to the grating and the incoming laser beams, $\pm\alpha$, is adjusted so that the fringe spacing δ is infinite. This angle α remains constant throughout the deformation. The angle α depends only on the initial wavelength of the grating, and in effect, fixing α acts to retain the initial grating wavelength to compare it with the current, deformed grating wavelength, in the form of a virtual reference grating as described in Section 2.2.2.

Recall that the initial grating wavelength is w_0; this length is known in advance. Also, γ_1, γ_2 and γ_3 are initially zero. The two incident laser beams are arranged so that the diffracted beams lie along the x_3 axis, meaning $u_1 = u_2 = v_1 = v_2 = 0$ and $u_3 = v_3 = 1$.

It follows from equations (2.3.31) through (2.3.32), making the indicated substitutions, that

$$\begin{aligned} u_1 = 0 \quad &= \quad \frac{\lambda}{w_0} - \sin\alpha \\ v_1 = 0 \quad &= \quad -\frac{\lambda}{w_0} + \sin\alpha \end{aligned} \tag{2.3.35}$$

which requires that

$$\sin\alpha = \frac{\lambda}{w_0} \tag{2.3.36}$$

This relation does not change during the deformation.

There is no similar simple representation for $\cos\alpha$, so the terms $\sin\alpha$ and $\cos\alpha$ will be retained unless the substitution helps to simplify an equation. It should be understood, however, that α is a known quantity that is related to the undeformed grating wavelength w_0 and to the wavelength of light used λ.

2.3.8 Observed Fringe Spacing and Inclination

In general, the fringe surfaces determined by the vector $\mathbf{z}$ are not directly observed, because the fringe planes are not aligned with the optical axis of the recording camera. The optical system used to record the fringes is aligned along the fixed X_3 axis and focuses on a plane parallel to the X_1–X_2 plane; it will thus record the fringe planes where they intersect a plane parallel to the X_1–X_2 plane. Thus, the optical system will not in general record the true fringe wavelength δ, but instead an apparent fringe wavelength δ^*. This apparent fringe wavelength corresponds to the spacing between the intersections of the fringe planes with planes parallel to the X_1–X_2 plane.

Let $\mathbf{E}_1, \mathbf{E}_2$ and $\mathbf{E}_3$ be unit vectors aligned with the globally fixed X_1, X_2 and X_3 axes, respectively. The observed fringes are determined by the vector that is the projection of $\mathbf{z}$ on the X_1–X_2 plane. Thus the *observed* fringe spacing, δ^*, is given by

$$\delta^* = \frac{\lambda}{|\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)\mathbf{E}_3|} \tag{2.3.37}$$

which can also be written as

$$\delta^* = \frac{\lambda}{[\mathbf{z}\cdot\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)^2]^{\frac{1}{2}}} \tag{2.3.38}$$

We also need to determine the angular orientation of the fringes in the fixed frame. As discussed earlier, we want to determine the angle between the fringes and the lines of the virtual reference grating formed by the incoming laser beams. This is the analog to the angle between the fringes and the lines of the reference grating in geometric moire techniques. The incoming laser beams are aligned so that the lines of the virtual reference grating are parallel to the fixed X_2 axis. Let ϕ^* be

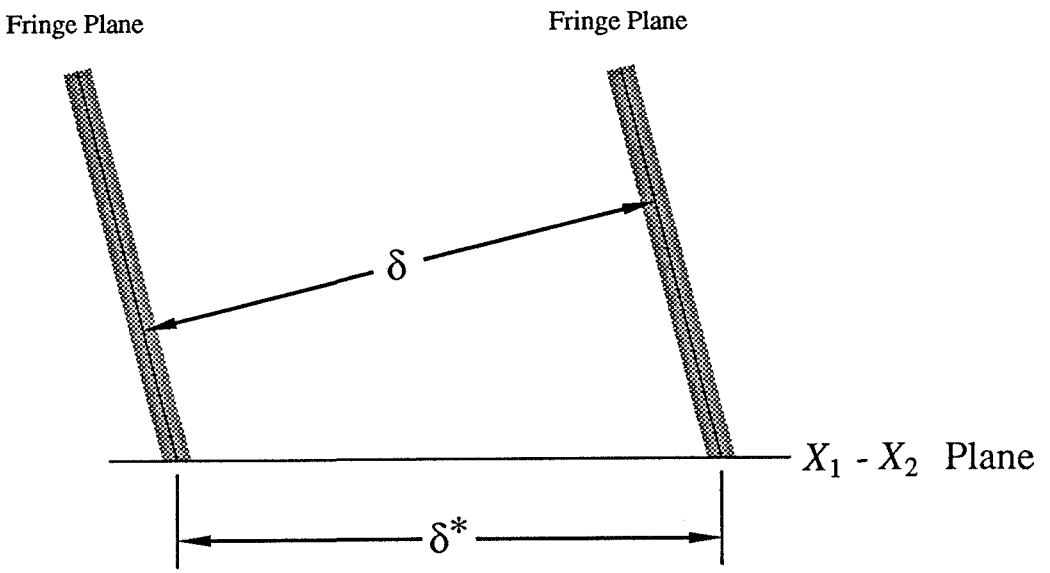


FIGURE 13. The relation between the apparent fringe spacing δ^* and the true fringe spacing $\mathbf{d}$. The apparent fringe spacing δ^* is the distance between the intersections of the fringe planes with the $X_1 - X_2$ plane. The camera maps the $X_1 - X_2$ plane to the film plane, so the apparent fringe spacing δ^* is measured instead of the true fringe spacing δ.

the angle between the observed fringes and the X_2 axis. Since the projection of the vector $\mathbf{z}$ on the X_1–X_2 plane, namely, $\mathbf{z} - (\mathbf{z} \cdot \mathbf{E}_3)\mathbf{E}_3$, is normal to the intersection of the fringe planes with the the X_1–X_2 plane, ϕ^* is also the angle between the vector $\mathbf{z} - (\mathbf{z} \cdot \mathbf{E}_3)\mathbf{E}_3$ and the X_1 axis, as shown in Figure 14, so that

$$\tan \phi^* = \frac{[\mathbf{z} - (\mathbf{z} \cdot \mathbf{E}_3)\mathbf{E}_3] \cdot \mathbf{E}_2}{[\mathbf{z} - (\mathbf{z} \cdot \mathbf{E}_3)\mathbf{E}_3] \cdot \mathbf{E}_1} = \frac{\mathbf{z} \cdot \mathbf{E}_2}{\mathbf{z} \cdot \mathbf{E}_1} \tag{2.3.39}$$

Again, $\mathbf{z}$ is the difference vector between the unit normals of the the diffracted plane waves, $\mathbf{z} = \mathbf{u} - \mathbf{v}$, and $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$ are unit vectors aligned with the globally fixed coordinate frame.

The components of $\mathbf{z}$ are known in the grating coordinate frame. In order to use equations (2.3.38) and (2.3.39), one must find the components of $\mathbf{z}$ in the globally fixed frame, or find the components of the vectors $\mathbf{E}_1, \mathbf{E}_2$ and $\mathbf{E}_3$ in the grating system. The latter course will be followed here. Let $\mathbf{e}_1, \mathbf{e}_2$ and $\mathbf{e}_3$ be unit vectors aligned with the grating x_1, x_2 and x_3 axes, respectively. $\mathbf{E}_1, \mathbf{E}_2$ and $\mathbf{E}_3$ are

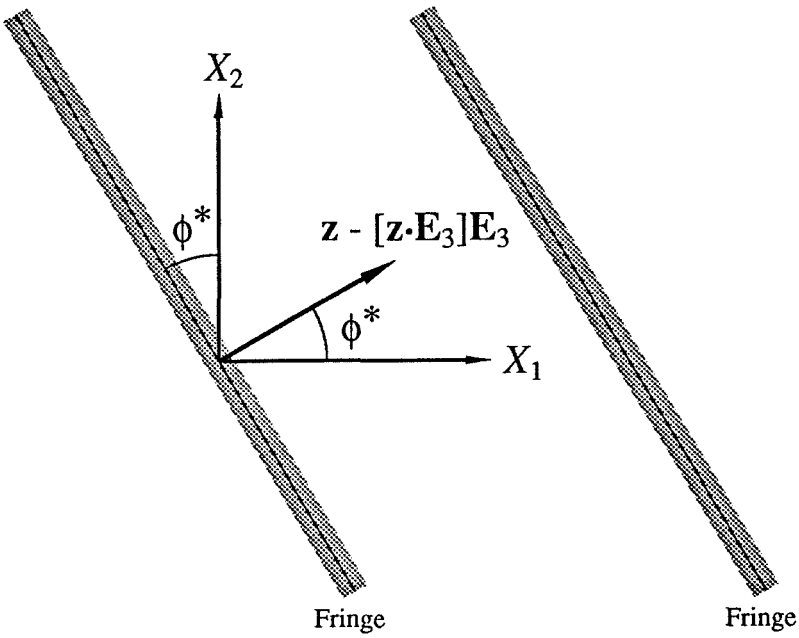


FIGURE 14. The angle between the observed fringes and the X_2 axis is ϕ^*. The X_2 axis is aligned with the lines of the virtual reference grating formed by the interference between the incoming laser beams. The angle ϕ^* is also the angle between the X_1 axis and the vector $\mathbf{z} - (\mathbf{z} \cdot \mathbf{E}_3)\mathbf{E}_3$, the projection of the difference vector $\mathbf{z}$ $(= \mathbf{u} - \mathbf{v})$ on the X_1–X_2 plane.

the unit vectors aligned with the globally fixed X_1, X_2 and X_3 axes. Then, again using equation (2.3.25),

$$\begin{aligned}
\mathbf{E}_1 &= (\cos\gamma_2 \cos\gamma_3)\mathbf{e}_1 + (\sin\gamma_1 \sin\gamma_2 \cos\gamma_3 - \cos\gamma_1 \sin\gamma_3)\mathbf{e}_2 \\
&\quad + (\cos\gamma_1 \sin\gamma_2 \cos\gamma_3 + \sin\gamma_1 \sin\gamma_3)\mathbf{e}_3 \\
\mathbf{E}_2 &= (\cos\gamma_2 \sin\gamma_3)\mathbf{e}_1 + (\sin\gamma_1 \sin\gamma_2 \sin\gamma_3 + \cos\gamma_1 \cos\gamma_3)\mathbf{e}_2 \\
&\quad + (\cos\gamma_1 \sin\gamma_2 \sin\gamma_3 - \sin\gamma_1 \cos\gamma_3)\mathbf{e}_3 \\
\mathbf{E}_3 &= (-\sin\gamma_2)\mathbf{e}_1 + (\sin\gamma_1 \cos\gamma_2)\mathbf{e}_2 + (\cos\gamma_1 \cos\gamma_2)\mathbf{e}_3
\end{aligned} \tag{2.3.40}$$

These vectors are used in the evaluation of the equations (2.3.38) and (2.3.39).

2.3.9 Special Case: Planar Deformation (No Tilting)

We will first examine the case of deformation in which the mean grating surface remains parallel to the globally fixed X_1–X_2 plane. This is the ideal case in which the moire interferometric fringes correspond exactly to the fringes of geometric moire as contours of constant in-plane displacement, as shown in previous investigations [41,57,72]. The rotations about the axes in the plane of the grating, γ_1 and γ_2, are identically zero. The components of the unit normals to the diffracted plane waves, **u** and **v**, in the deformed grating frame are given by:

$$\begin{aligned} u_1 &= \frac{\lambda}{w} - \sin\alpha\cos\gamma_3 \\ u_2 &= \sin\alpha\sin\gamma_3 \\ u_3 &= [1-(u_1)^2-(u_2)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.41}$$

$$\begin{aligned} v_1 &= -\frac{\lambda}{w} + \sin\alpha\cos\gamma_3 \\ v_2 &= -\sin\alpha\sin\gamma_3 \\ v_3 &= [1-(v_1)^2-(v_2)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.42}$$

Since $u_1 = -v_1$ and $u_2 = -v_2$, it follows that $u_3 = v_3$. The vector $\mathbf{z}$ $(= \mathbf{u} - \mathbf{v})$, which describes the fringe spacing and orientation, then has components in the grating coordinate frame given by

$$\begin{aligned} z_1 &= 2\left(\frac{\lambda}{w} - \sin\alpha\cos\gamma_3\right) \\ z_2 &= 2(\sin\alpha\sin\gamma_3) \\ z_3 &= 0 \end{aligned} \tag{2.3.43}$$

The true fringe spacing δ is found from equation (2.3.17), as

$$\delta = \frac{\lambda}{|\mathbf{z}|} = \frac{\lambda}{2\left[\left(\frac{\lambda}{w}\right)^2 + \sin^2\alpha - 2\frac{\lambda}{w}\sin\alpha\cos\gamma_3\right]^{\frac{1}{2}}} \tag{2.3.44}$$

The equations describing the unit vectors aligned with the axes of the globally fixed coordinate frame, equation (2.3.40), reduce to

$$\begin{aligned} \mathbf{E}_1 &= \cos\gamma_3 \mathbf{e}_1 - \sin\gamma_3 \mathbf{e}_2 \\ \mathbf{E}_2 &= \sin\gamma_3 \mathbf{e}_1 + \cos\gamma_3 \mathbf{e}_2 \\ \mathbf{E}_3 &= \mathbf{e}_3 \end{aligned} \tag{2.3.45}$$

Therefore, $\mathbf{z} \cdot \mathbf{E}_3 = 0$, and so the observed fringe spacing δ^* is the same as the true fringe spacing δ in equation (2.3.44),

$$\delta^* = \frac{\lambda}{2\left[\left(\frac{\lambda}{w}\right)^2 + \sin^2\alpha - 2\frac{\lambda}{w}\sin\alpha\cos\gamma_3\right]^{\frac{1}{2}}} \tag{2.3.46}$$

The inclination of the fringe planes with respect to the X_2 axis (and the lines of the virtual reference grating) is found using equation (2.3.39), as

$$\tan\phi^* = \frac{\left(\frac{\lambda}{w}\right)\sin\gamma_3}{\left(\frac{\lambda}{w}\right)\cos\gamma_3 - \sin\alpha} \tag{2.3.47}$$

Substituting the condition required by the initial alignment of the incident laser beams, that $\sin\alpha = \lambda/w_0$, from equation (2.3.36), one finds for the observed fringe spacing and inclination

$$\delta^* = \frac{ww_0}{2[w^2 + w_0^2 - 2ww_0\cos\gamma_3]^{\frac{1}{2}}} \tag{2.3.48}$$

$$\tan\phi^* = \frac{w_0\sin\gamma_3}{w_0\cos\gamma_3 - w} \tag{2.3.49}$$

Equations (2.3.48) and (2.3.49) thus imply that in the case of purely planar deformation, the fringes can be interpreted as contours of constant displacement, exactly as in geometric moire, as shown in Section 2.2.1. Note that the fringe spacing and inclination depend only on the initial and current states of the grating and are independent of the wavelength of light used. The constant displacement increment between adjacent fringes is $w_0/2$, in the direction of the fixed X_1 axis.

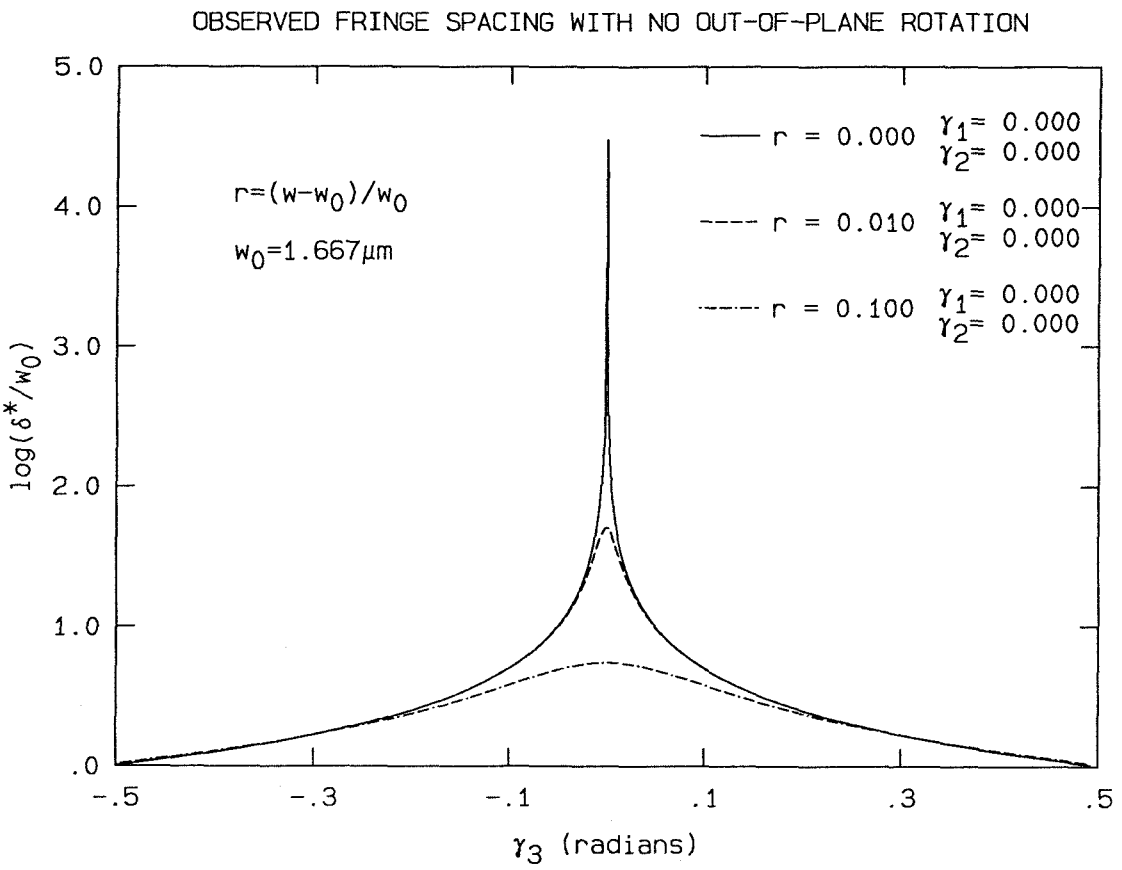


FIGURE 15. The variation of the observed fringe spacing δ^* when the out-of-plane rotations are identically zero. The angle γ_3 is the angle between the ridges of the deformed grating and the fixed X_2 axis, which is aligned with the ridges of the undeformed grating (γ_3 is the rotation of the ridges about the normal to the grating plane). The fringe spacing is always maximum at $\gamma_3 = 0$. Negative values for the stretch ratio r have a similar effect on the fringe spacing. $w_0 = 1.667\mu$m corresponds to the grating wavelength used in the experiments.

Figures 15 and 16 show the variation of the fringe spacing and inclination in the case of zero tilting of the grating plane. The choice of $w_0 = 1.667$ μm matches the grating wavelength used in the experiments.

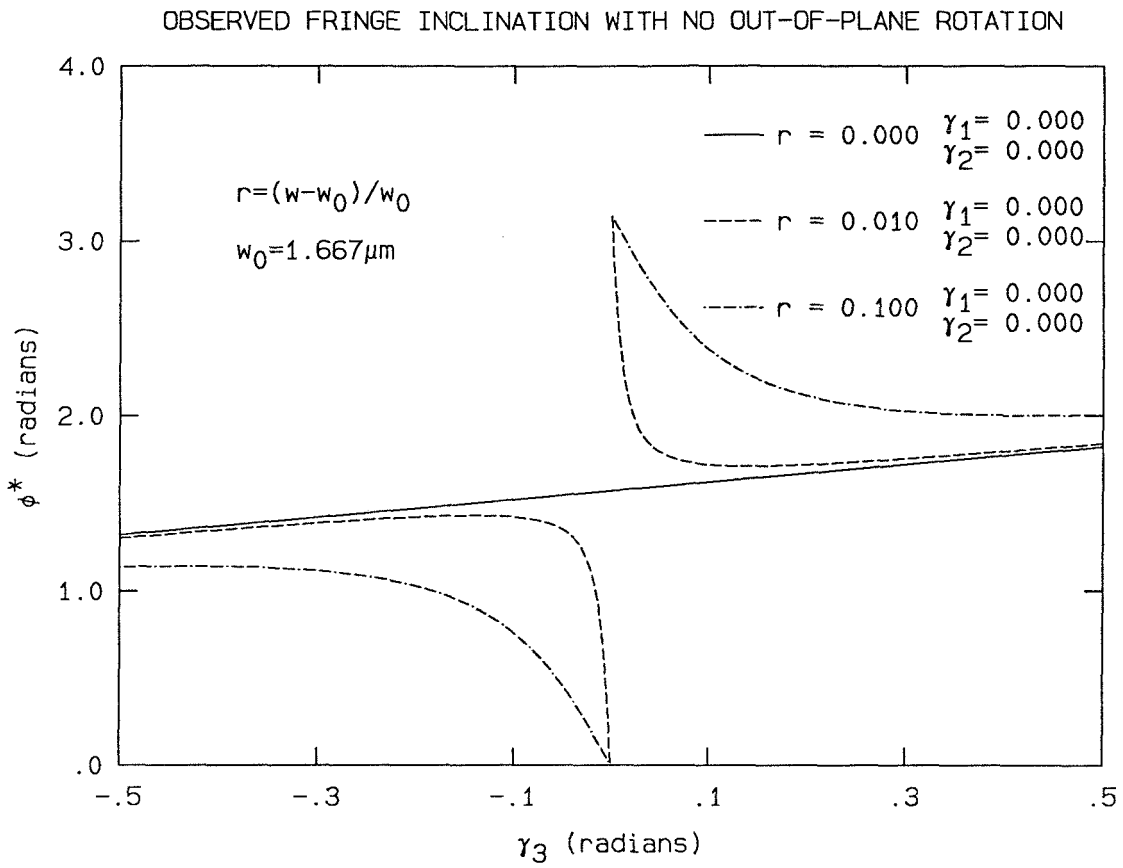


FIGURE 16. The variation of the observed fringe inclination ϕ^* when the out-of-plane rotations are identically zero. The angle γ_3 is the angle between the ridges of the deformed grating and the fixed X_2 axis, which is aligned with the ridges of the undeformed grating. (Alternatively, γ_3 is the rotation of the ridges about the normal to the grating plane.) Note that the fringe inclination is periodic in π; the discontinuities at $\gamma_3 = 0$ are transitions across branches of the arctangent function. Negative values for the stretch ratio r change the sign of the deviation from the line at r = 0. $w_0 = 1.667\mu$m corresponds to the grating wavelength used in the experiments.

2.3.10 Approximations for General Deformations

The case of planar deformation provides a starting point for approximating the equations describing the fringes in cases where the deformation does not remain strictly in a plane parallel to the original mean grating surface. These approximations will help to show the dominant sources of error caused by the out-of-plane

rotations γ_1 and γ_2. In effect, it is desirable to be able to interpret the fringes directly as contour lines of displacement, rather than to try to calculate the underlying parameters w and γ_3. The identity of equations (2.3.48) and (2.3.49) with the equations for the case of geometric moire means that the fringes do represent contours of displacement when $\gamma_1 = \gamma_2 = 0$. If the effect of out-of-plane rotations on equations (2.3.38) and (2.3.39) can be neglected, then the fringes can still be interpreted as contours of displacement. In the initial, undeformed configuration, both of the diffracted rays are parallel to the global X_3 axis, which corresponds to the grating x_3 axis, so $u_3 \equiv v_3 \equiv 1$. We assume that as the grating deforms, these components do not differ greatly from unity. † Using this assumption, u_3 and v_3 can be approximated by two–term Taylor expansions about unity:

$$u_3 = 1 - \frac{1}{2}(u_1^2 + u_2^2)$$
$$v_3 = 1 - \frac{1}{2}(v_1^2 + v_2^2) \tag{2.3.50}$$

With this assumption, the component of $\mathbf{z}$ along the x_3 axis is

$$z_3 = u_3 - v_3 = \frac{1}{2}(v_1^2 + v_2^2 - u_1^2 - u_2^2) \tag{2.3.51}$$

Using equations (2.3.31) through (2.3.32), which spell out the components u_1, u_2, v_1 and v_2, one finds

$$z_3 = 2\left[\left(\frac{\lambda}{w} - \sin\alpha\cos\gamma_3\cos\gamma_2\right)(-\sin\gamma_2\cos\alpha)\right.$$
$$\left. + (\sin\alpha\sin\gamma_3\cos\gamma_1 - \sin\alpha\cos\gamma_3\sin\gamma_2\sin\gamma_1)(\cos\alpha\cos\gamma_2\sin\gamma_1)\right] \tag{2.3.52}$$

From the expressions for z_1 and z_2 in equations (2.3.33) it is seen that

$$z_3 = z_1(-\sin\gamma_2\cos\alpha) + z_2(\sin\gamma_1\cos\gamma_2\cos\alpha) \tag{2.3.53}$$

---

† This assumption can be justified on the basis of the physical limitations of the optics used to record the fringes. Large changes in these components that would violate this assumption cause the light from the specimen to fall outside the aperture of the recording camera.

Thus, for this approximation to the components u_3 and v_3, we can write the components of $\mathbf{z}$ as

$$\begin{aligned} z_1 &= 2\left(\frac{\lambda}{w} - \sin\alpha\cos\gamma_3\cos\gamma_2\right) \\ z_2 &= 2(\sin\alpha\sin\gamma_3\cos\gamma_1 - \sin\alpha\cos\gamma_3\sin\gamma_2\sin\gamma_1) \\ z_3 &= z_1(-\sin\gamma_2\cos\alpha) + z_2(\sin\gamma_1\cos\gamma_2\cos\alpha) \end{aligned} \tag{2.3.54}$$

We can now evaluate the expressions for the observed fringe spacing, δ^*, and the fringe inclination, ϕ^*, in equations (2.3.38) and (2.3.39). The denominator in equation (2.3.38) is $[\mathbf{z}\cdot\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)^2]^{\frac{1}{2}}$. After some algebra, this is found as

$$\begin{aligned} &\left[\mathbf{z}\cdot\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)^2\right]^{\frac{1}{2}} \\ &\quad = \left[z_1^2 + z_2^2\right. \\ &\qquad \left. + (z_1\sin\gamma_2 - z_2\sin\gamma_1\cos\gamma_2)^2\left(\cos^2\alpha - (1+\cos\gamma_1\cos\gamma_2\cos\alpha)^2\right)\right]^{\frac{1}{2}} \end{aligned} \tag{2.3.55}$$

In order to determine the fringe inclination ϕ^*, we must find $\mathbf{z}\cdot\mathbf{E}_1$ and $\mathbf{z}\cdot\mathbf{E}_2$. Again, after some algebra, we find

$$\begin{aligned} \mathbf{z}\cdot\mathbf{E}_2 = {} & z_1(\cos\gamma_2\sin\gamma_3) + z_2(\sin\gamma_1\sin\gamma_2\sin\gamma_3 + \cos\gamma_1\cos\gamma_3) \\ & + z_1(-\sin\gamma_2\cos\alpha)(\cos\gamma_1\sin\gamma_2\sin\gamma_3 - \sin\gamma_1\cos\gamma_3) \\ & + z_2(\sin\gamma_1\cos\gamma_2\cos\alpha)(\cos\gamma_1\sin\gamma_2\sin\gamma_3 - \sin\gamma_1\cos\gamma_3) \end{aligned} \tag{2.3.56}$$

$$\begin{aligned} \mathbf{z}\cdot\mathbf{E}_1 = {} & z_1(\cos\gamma_2\cos\gamma_3) + z_2(\sin\gamma_1\sin\gamma_2\cos\gamma_3 - \cos\gamma_1\sin\gamma_3) \\ & + z_1(-\sin\gamma_2\cos\alpha)(\cos\gamma_1\sin\gamma_2\cos\gamma_3 + \sin\gamma_1\sin\gamma_3) \\ & + z_2(\sin\gamma_1\cos\gamma_2\cos\alpha)(\cos\gamma_1\sin\gamma_2\cos\gamma_3 + \sin\gamma_1\sin\gamma_3) \end{aligned} \tag{2.3.57}$$

2.3.11 Small–Angle Approximations

As the plane of the grating rotates away from its initial orientation, fringes are generated that are not due to the in-plane deformation alone, but have a contribution that is due to the tilt of the grating surface. The effect of this tilt can be estimated by approximating the rotations that change the orientation of the grating plane with small angles. This approximation will provide information about the leading sources of discrepancy in the equations describing the fringe spacing and orientation.

If the angles γ_1 and γ_2 (which are measures of how far the plane of the grating has varied from its initial orientation) are restricted to be small, one has $\sin\gamma_\alpha = \gamma_\alpha$, $\cos\gamma_\alpha = 1$, where $\alpha = 1, 2$. The angle γ_3 is not restricted, as it is part of the in-plane deformation, although this angle is also usually small.

With these additional approximations, the components of the vector $\mathbf{z}$ are found to be

$$\begin{aligned} z_1 &= 2\left(\frac{\lambda}{w} - \sin\alpha\cos\gamma_3\right) \\ z_2 &= 2(\sin\alpha\sin\gamma_3 - \gamma_1\gamma_2\sin\alpha\cos\gamma_3) \\ z_3 &= z_1(-\gamma_2\cos\alpha) + z_2(\gamma_1\cos\alpha) \end{aligned} \tag{2.3.58}$$

Again, to determine the observed fringe spacing and inclination, we must evaluate equations (2.3.38) and (2.3.39). The denominator in equation (2.3.38) is found as

$$\left[\mathbf{z}\cdot\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)^2\right]^{\frac{1}{2}} = \left[z_1^2 + z_2^2 - (z_1\gamma_2 - z_2\gamma_1)^2(1 + 2\cos\alpha)\right]^{\frac{1}{2}} \tag{2.3.59}$$

The terms needed to evaluate equation (2.3.39) are

$$\begin{aligned} \mathbf{z}\cdot\mathbf{E}_2 = {} & z_1(\sin\gamma_3) + z_2(\gamma_1\gamma_2\sin\gamma_3 + \cos\gamma_3) \\ & + z_1(-\gamma_2\cos\alpha)(\gamma_2\sin\gamma_3 - \gamma_1\cos\gamma_3) \\ & + z_2(\gamma_1\cos\alpha)(\gamma_2\sin\gamma_3 - \gamma_1\cos\gamma_3) \end{aligned} \tag{2.3.60}$$

$$\begin{aligned}\mathbf{z}\cdot\mathbf{E}_1 = {} & z_1(\cos\gamma_3) + z_2(\gamma_1\gamma_2\cos\gamma_3 - \sin\gamma_3) \\ & + z_1(-\gamma_2\cos\alpha)(\gamma_2\cos\gamma_3 + \gamma_1\sin\gamma_3) \\ & + z_2(\gamma_1\cos\alpha)(\gamma_2\cos\gamma_3 + \gamma_1\sin\gamma_3) \qquad (2.3.61)\end{aligned}$$

Making further use of the small-angle approximations, the third term in equation (2.3.59) can be neglected with respect to the first two terms, since it is of the order $\gamma_\alpha\gamma_\beta$ times the first two terms, and is itself identically zero when the sum of the first two terms is zero. ($x^2 + y^2 = 0 \Rightarrow x \equiv y \equiv 0$). So for this approximation

$$\begin{aligned}\left[\mathbf{z}\cdot\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)^2\right]^{\frac{1}{2}} &= \left[z_1^2 + z_2^2\right]^{\frac{1}{2}} \\ &= 2\Bigg[\left(\frac{\lambda}{w}\right)^2 + \sin^2\alpha - 2\frac{\lambda}{w}\sin\alpha\cos\gamma_3 \\ &\qquad + \gamma_1^2\gamma_2^2\sin^2\alpha\cos^2\gamma_3 - \gamma_1\gamma_2\sin^2\alpha\sin 2\gamma_3\Bigg]^{\frac{1}{2}} \\ &\qquad (2.3.62)\end{aligned}$$

Expanding the expressions in equations (2.3.60) and (2.3.61) and making use of the small-angle approximations, we find that

$$\frac{\mathbf{z}\cdot\mathbf{E}_2}{2} = \frac{\lambda}{w}\sin\gamma_3 - \gamma_1\gamma_2\sin\alpha\cos 2\gamma_3(1+\cos\alpha) + \gamma_1\gamma_2\frac{\lambda}{w}\cos\gamma_3\cos\alpha \qquad (2.3.63)$$

$$\begin{aligned}\frac{\mathbf{z}\cdot\mathbf{E}_1}{2} = {} & \frac{\lambda}{w}\cos\gamma_3 - \sin\alpha + \gamma_1\gamma_2\sin\alpha\sin 2\gamma_3(1+\cos\alpha) - \gamma_1\gamma_2\frac{\lambda}{w}\sin\gamma_3\cos\alpha \\ & - \gamma_1^2\gamma_2^2\sin\alpha\cos^2\gamma_3(1+\cos\alpha) + (\gamma_1^2 - \gamma_2^2)\sin\alpha\sin^2\gamma_3\cos\alpha \\ & \qquad (2.3.64)\end{aligned}$$

As stated earlier, it is assumed that the initial grating wavelength is w_0 and that the incoming laser beams are aligned so that the initial fringe spacing is infinite, which leads to the relation $\sin\alpha = \lambda/w_0$. Making this substitution in equation (2.3.38) for the apparent fringe spacing δ^* yields the result that

$$\delta^* = \frac{ww_0}{2\left[\{w_0^2 + w^2 - 2ww_0\cos\gamma_3\} + \gamma_1^2\gamma_2^2 w^2\cos^2\gamma_3 - \gamma_1\gamma_2 w^2\sin 2\gamma_3\right]^{\frac{1}{2}}} \tag{2.3.65}$$

Upon substitution of $\sin\alpha = \lambda/w_0$ in equations (2.3.63) and (2.3.64) in order to determine the apparent fringe inclination ϕ^*,

$$\frac{ww_0(\mathbf{z}\cdot\mathbf{E}_2)}{2\lambda} = \{w_0\sin\gamma_3\} - \gamma_1\gamma_2 w\cos 2\gamma_3(1+\cos\alpha) + \gamma_1\gamma_2 w_0\cos\gamma_3\cos\alpha \tag{2.3.66}$$

$$\begin{aligned}\frac{ww_0(\mathbf{z}\cdot\mathbf{E}_1)}{2\lambda} =& \{w_0\cos\gamma_3 - w\} + \gamma_1\gamma_2 w\sin 2\gamma_3(1+\cos\alpha) - \gamma_1\gamma_2 w_0\sin\gamma_3\cos\alpha \\ & - \gamma_1^2\gamma_2^2 w\cos^2\gamma_3(1+\cos\alpha) + (\gamma_1^2-\gamma_2^2)w\sin^2\gamma_3\cos\alpha\end{aligned} \tag{2.3.67}$$

From these two equations the apparent fringe inclination ϕ^* can be determined as

$$\tan\phi^* = \frac{ww_0(\mathbf{z}\cdot\mathbf{E}_2)/2\lambda}{ww_0(\mathbf{z}\cdot\mathbf{E}_1)/2\lambda} = \frac{(\mathbf{z}\cdot\mathbf{E}_2)}{(\mathbf{z}\cdot\mathbf{E}_1)} \tag{2.3.68}$$

2.3.12 Discussion

Equations (2.3.65) and (2.3.68) describe the observed fringe spacing δ^*, and the angle between the X_2 axis and the fringes, ϕ^*, when the angles that represent tilting of the grating plane, angles γ_1 and γ_2, are restricted to be small. The terms in braces {} in equations (2.3.65) through (2.3.67) are exactly those terms found in the special case of planar deformation, so if the angles γ_1 and γ_2 are identically zero, we recover the equations for planar deformation, as expected. The terms in braces represent the in-plane deformation, and the remaining terms are due to the out-of-plane rotation. Note that the remaining terms in equations (2.3.65) through (2.3.67) are all of second or higher order in the small angles γ_1 and γ_2. Unfortunately, even though these terms are small, it is possible for the terms that

are due to the in-plane deformation to be identically zero, so that the terms that are due to the out-of-plane rotations dominate the equations. It is interesting to see that in all the terms of equations (2.3.65) through (2.3.67), except for the last term in equation (2.3.67), the angles γ_1 and γ_2 are coupled. It is also interesting to see that $\cos\alpha$ does not appear in equation (2.3.65), which means that the observed fringe spacing is independent of the wavelength of light, λ, as it was in the case of strictly planar deformation. The angle α is related to the wavelength of light through the equation $\sin\alpha = \lambda/w_0$, so that $\cos\alpha = \sqrt{1-(\lambda^2/w_0^2)}$. Thus, the $\cos\alpha$ terms are the only terms that are wavelength–dependent.

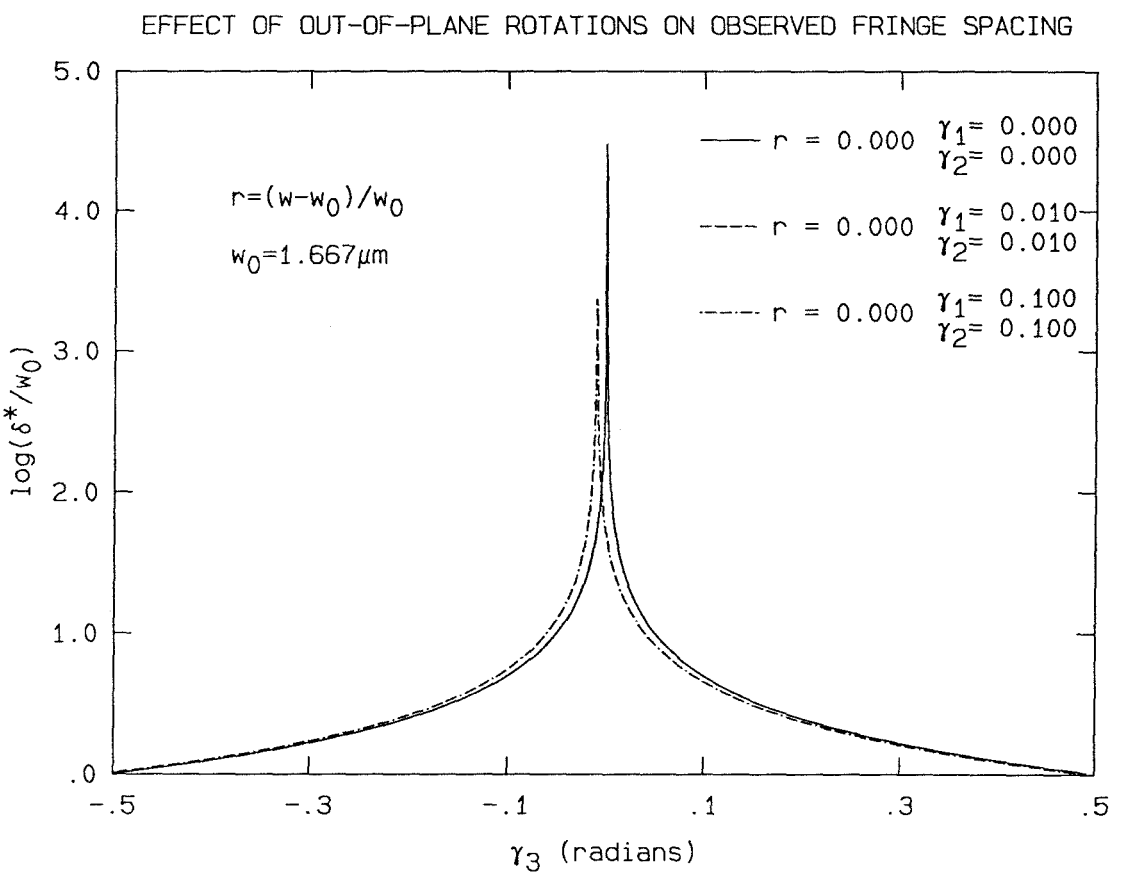


FIGURE 17. The variation for δ^* with out-of-plane rotations for the case where the stretch ratio is zero. In this case the effect of the out-of-plane rotations is most severe. The offset of the curve for larger γ_1 and γ_2 would be in the opposite direction if γ_1 or γ_2 were negative.

Recall that Figures 15 and 16 show the variation of the spacing and inclination of the observed fringes when there is no tilting of the specimen surface, so that $\gamma_1 = \gamma_2 = 0$. These plots are generated from equations (2.3.48) and (2.3.49), the equations describing the fringe spacing and inclination for the special case of planar deformation. These figures provide a baseline by which to judge the effects of the tilt introduced through angles γ_1 and γ_2, as will be seen in Figures 17 through 20.

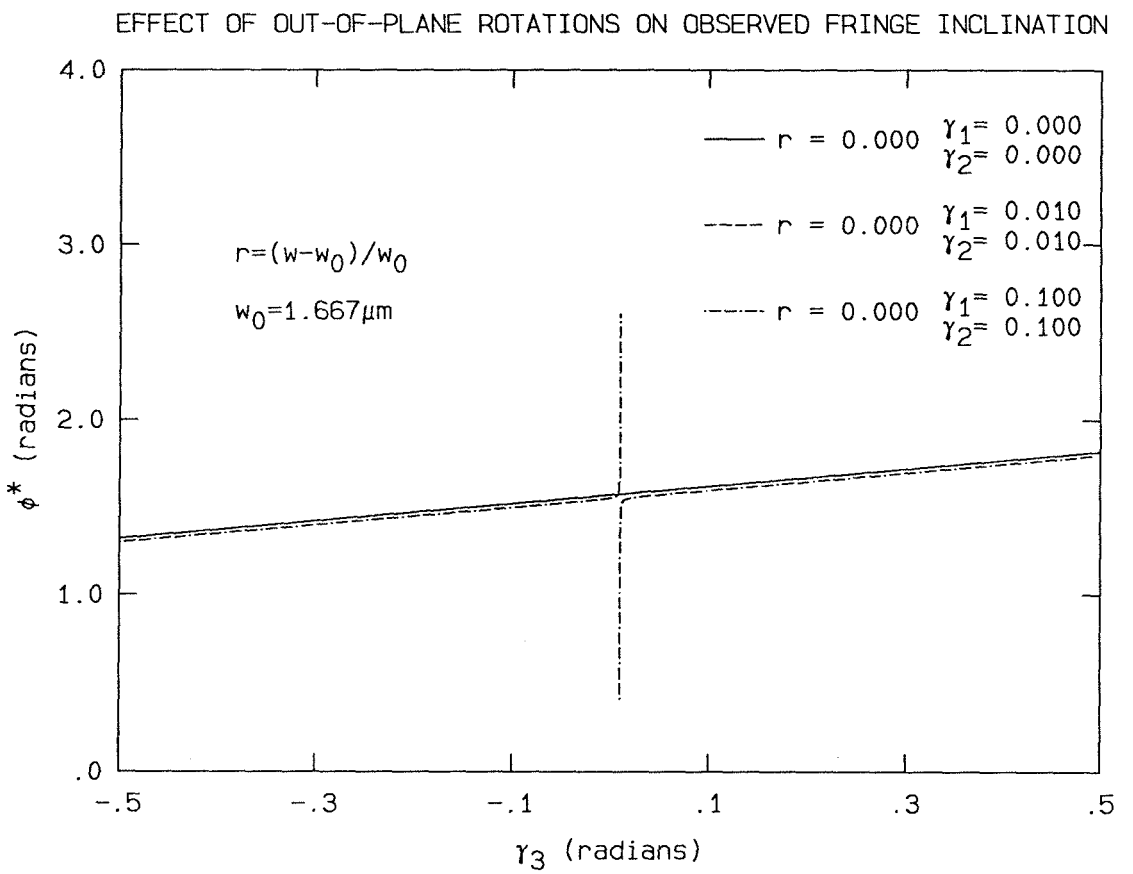


FIGURE 18. The variation for ϕ^* with out-of-plane rotations for the case where the stretch ratio is zero. Again, this is the case where the effect of the out-of-plane rotations is most severe.

Figure 15 shows the variation of fringe spacing with the stretch ratio of the grating ($r = (w - w_0)/w_0$), and the rotation of the ridges of the grating about the normal to the grating plane (γ_3). For the case of zero stretch, the fringe spacing is infinite at $\gamma_3 = 0$; however, the fringe spacing drops off quickly as γ_3 differs from

zero. For stretch ratios different from zero, the fringe spacing is still a maximum at $\gamma_3 = 0$. Only positive stretch ratios are shown in the figure; negative stretch ratios would have a similar effect on the fringe spacing. Note that at a stretch ratio of $r = 0.100$, the fringe spacing is less than $10 \times w_0$ even at $\gamma_3 = 0$, indicating that for the grating in the experiments, the fringe spacing would be less than 20μm for a stretch ratio of 0.100.

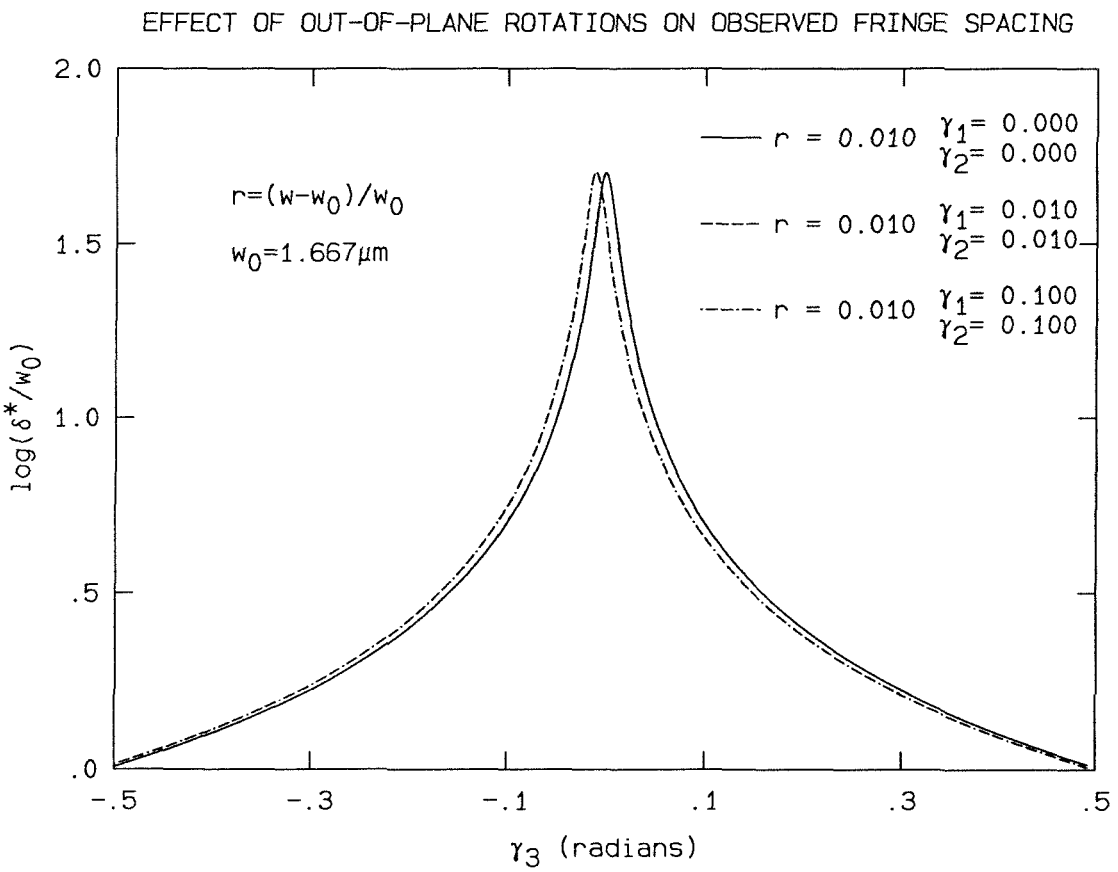


FIGURE 19. The variation for δ^* with out-of-plane rotations for the case where the stretch ratio is an intermediate value. (About 1% strain.) In this case the error that is due to the out-of-plane rotations is less severe than for the case where the stretch ratio is zero.

Figure 16 shows the variation of the observed fringe inclination with respect to the fixed X_2 axis. The apparent discontinuities at $\gamma_3 = 0$ are smooth transitions from one branch of the arctangent function to another branch. Again, only positive

stretch ratios are shown; negative stretch ratios would cause the curves to deviate from the $r = 0.000$ case in the opposite sense.

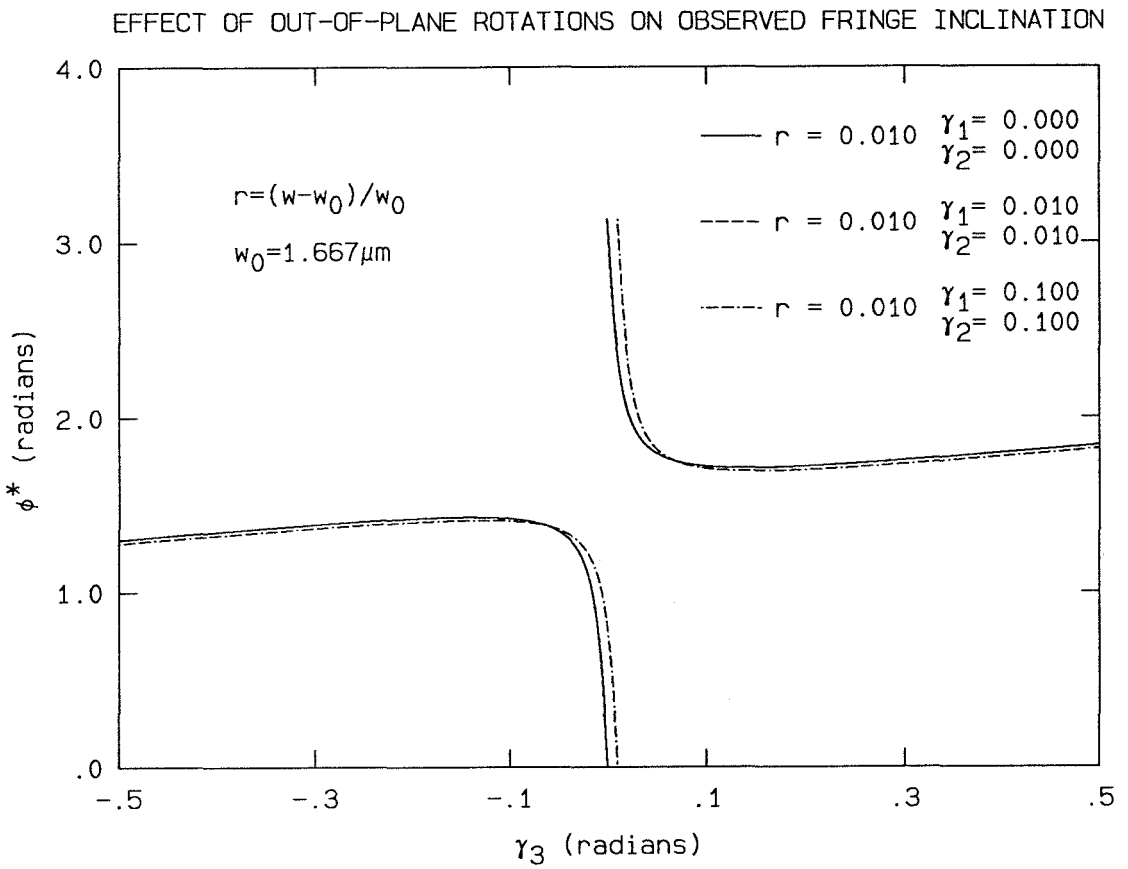


FIGURE 20. The variation for ϕ^* with out-of-plane rotations for the case where the stretch ratio is an intermediate value. (About 1% strain.) In this case the error that is due to the out-of-plane rotations is less severe than for the case where the stretch ratio is zero, except near the region where $\gamma_3 = 0$.

Figures 17 and 18 show the effect of different values for the out-of-plane rotations γ_1 and γ_2 when the stretch ratio is zero. These plots are obtained from equations (2.3.65) and (2.3.68), which determine the fringe spacing and inclination, assuming that γ_1 and γ_2 are small angles. Such is the case when the grating does not change its wavelength, but is rigidly rotated by varying amounts about the three axes defined by the ridges of the grating. The effects of the tilting of the

grating plane are most severe near $\gamma_3 = 0$; at $\gamma_3 = 0$, there is no in-plane deformation at all. It can be seen in Figure 17 that near $\gamma_3 = 0$, the fringe spacing is off by several orders of magnitude. Similarly, in Figure 18, the angle ϕ^* near $\gamma_3 = 0$ can be off by 90°. However, away from $\gamma_3 = 0$, the curves approach the baseline curves of Figures 15 and 16. It can also be seen that the curves of $\gamma_1 = \gamma_2 = 0.010$ are obscured by the baseline curves of $\gamma_1 = \gamma_2 = 0.000$. In the figures, only the case of $\gamma_1 = \gamma_2 = 0.100$ shows a deviation from the case where the tilt is zero. The condition that $\gamma_1 = \gamma_2 = 0.100$ is about the limit at which the small–angle approximation can be applied.

Figures 19 and 20 show the effect of varying amounts of tilting for a second case displayed in Figures 15 and 16, the case where the stretch ratio is $r = 0.010$. This value corresponds to about 1% strain. Again, the effects of the out-of-plane rotations are localized about $\gamma_3 = 0$, and the deviation from the case when there is no tilting is apparent only for $\gamma_1 = \gamma_2 = 0.100$, which represents a reasonable limit for applying the small–strain approximation used in generating these curves. The deviation in Figure 19 is much less severe than in Figure 17; the maximum deviation appears to be about one–half of an order of magnitude. On the other hand, in Figure 20, the deviation near $\gamma_3 = 0$ may still be 90°.

It is apparent that as the in-plane deformation increases, the errors introduced by the out-of-plane rotations are reduced. This fact indicates that for a measurement near the tip of a crack, where both the in-plane deformation and the out-of-plane rotations may be large, large–enough in-plane deformations will overwhelm the effect of the out-of-plane rotations in the fringe patterns obtained using the moire interferometer.

2.3.13 Effect on Current Study

It was hoped that the effect of the out-of-plane rotations on the technique of moire interferometry could be neglected for all cases. As shown in Figures 17 through 20, however, neglecting that effect is not possible. It is still necessary to estimate the effect of the out-of-plane deformation gradient at the tip of the crack on the measurements made for the current study. The out-of-plane deformation field from the finite–element model (described in Section 4.3) will be employed along with some details of the experimental apparatus (described in Sections 2.4.5 and 4.4).

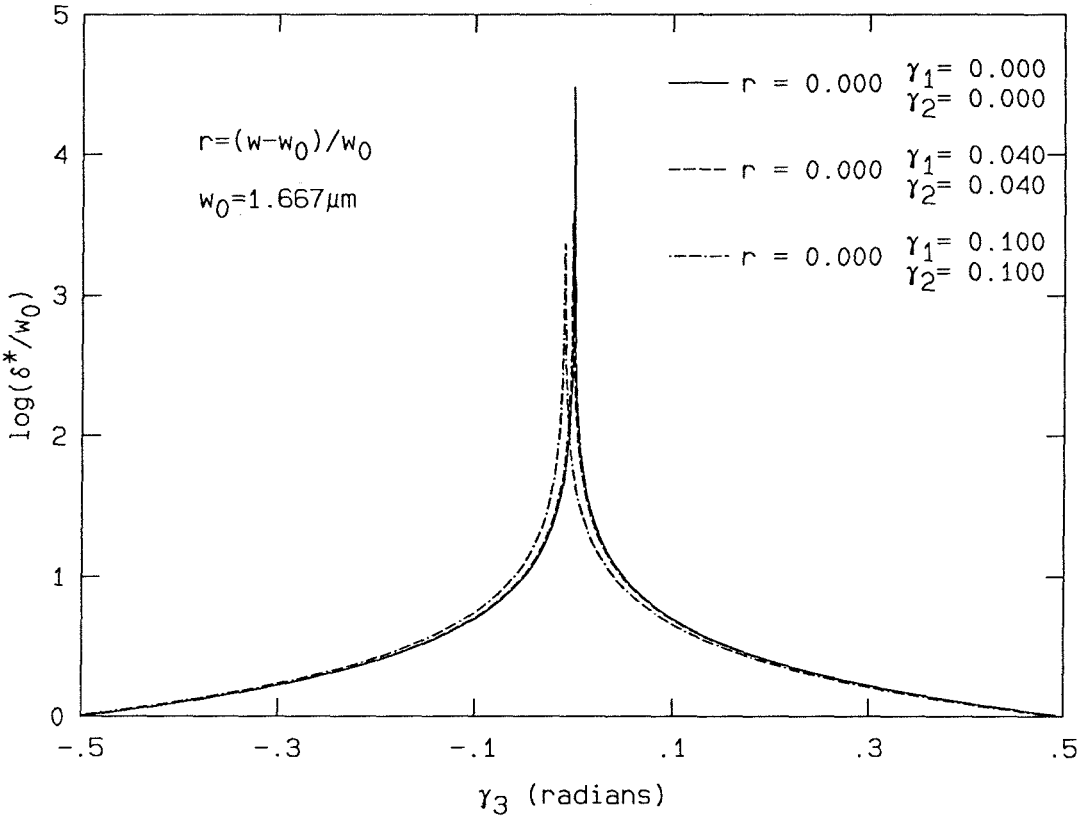


FIGURE 21. The variation for δ^* with out-of-plane rotations for the case where the stretch ratio is zero. In this case the effect of the out-of-plane rotations is most severe. The choice of $\gamma_1 = 0.040$ and $\gamma_2 = 0.040$ is motivated by the maximum rotation allowed in the experiment.

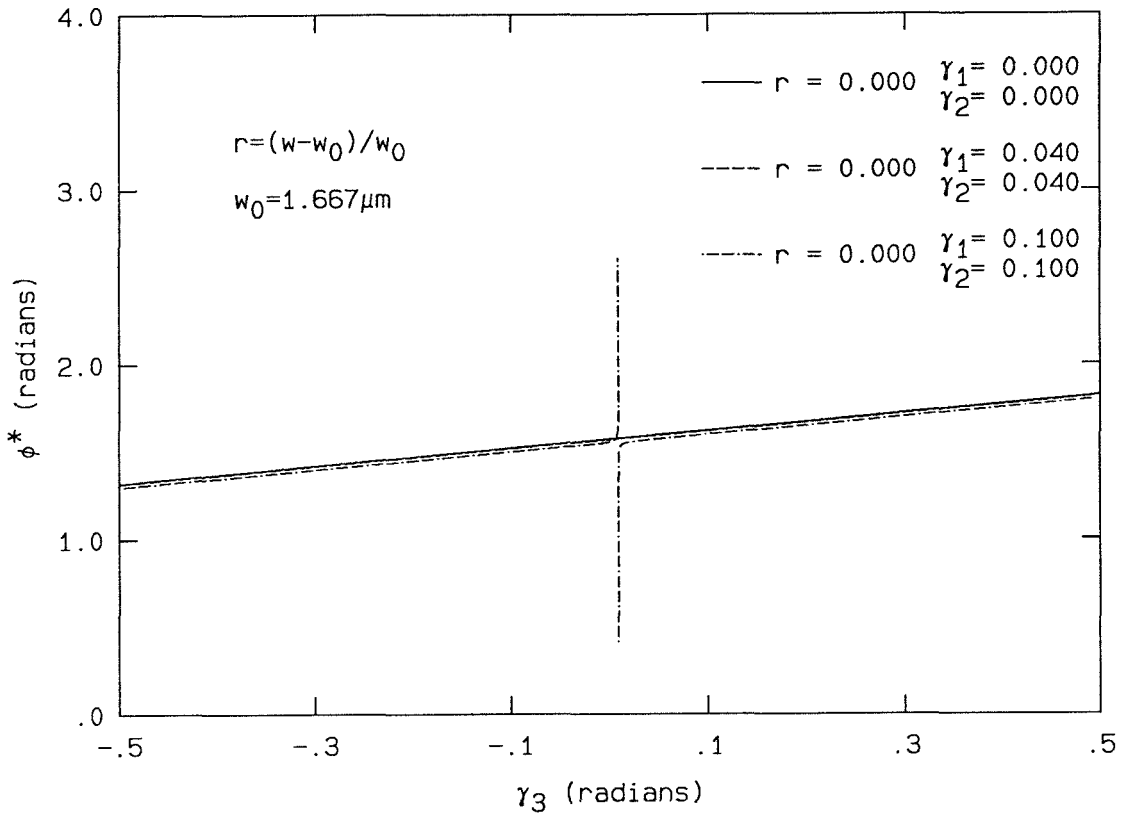


FIGURE 22. The variation for ϕ^* with out-of-plane rotations for the case where the stretch ratio is zero. Again, this is the case where the effect of the out-of-plane rotations is most severe. The choice of $\gamma_1 = 0.040$ and $\gamma_2 = 0.040$ is motivated by the maximum rotation allowed in the experiment.

The angular aperture of the camera used for the experiment is about 0.08 radian. This is a measure of the maximum allowable angular deflection from the optical axis of the camera, which may occur for a light ray originating from the center of the field of view. In effect, the angular aperture limitation results in shadows on the photograph where the light rays deviate from the camera axis by more than 0.08 radian. In reflection, a rotation of the surface of 0.04 radian leads to an angular deflection of the reflected light of 0.08 radian. The effect is somewhat less in diffraction, where a rotation of the surface of 0.04 radian leads to an angular deflection of 0.077 radian of the diffracted light. Thus, the experimental apparatus

places an upper bound on the surface slope of about 0.04, since larger slopes deflect the light beyond the aperture of the camera.

Figures 21 and 22 show the effects of a surface slope of 0.04 on the observed fringe spacing and inclination, in comparison with a surface slope of 0.1. Actually, with both $\gamma_1 = 0.040$ and $\gamma_2 = 0.040$, this is slightly more severe than a slope of 0.040. These figures represent the case where the in-plane stretch is zero, which is the case where the out-of-plane deformation has the most effect.

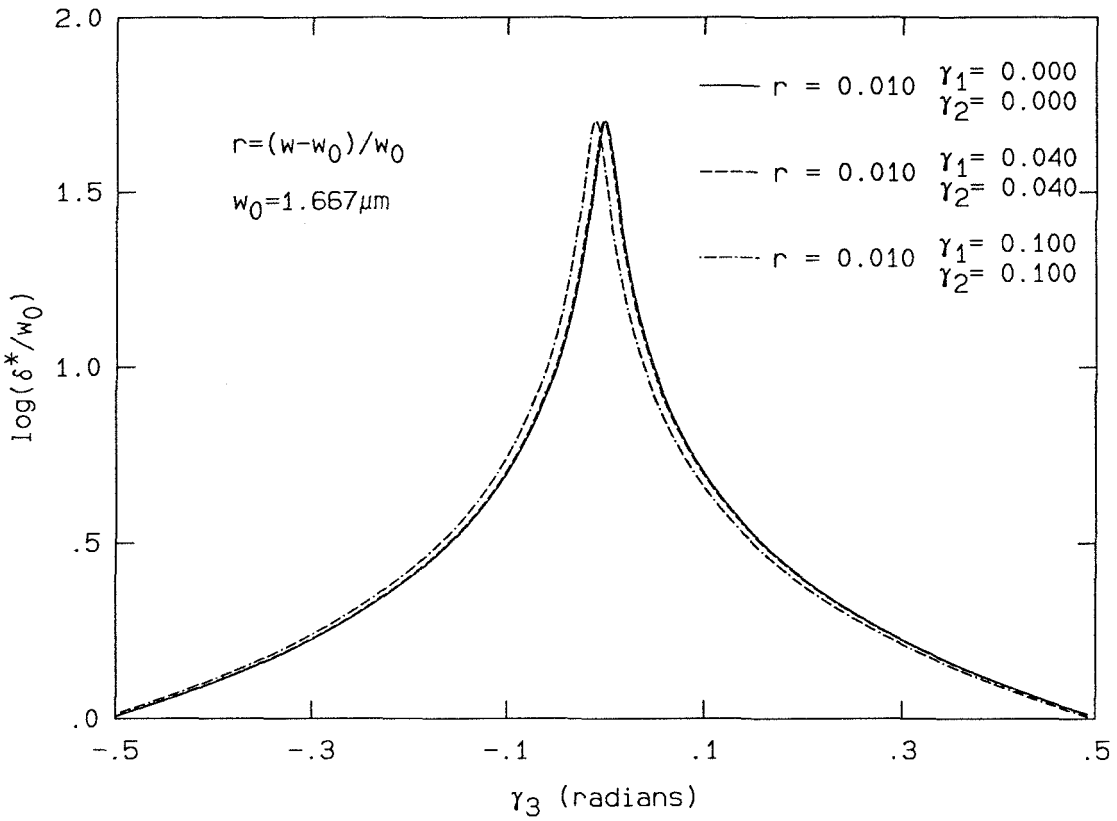


FIGURE 23. The variation for δ^* with out-of-plane rotations for the case where the stretch ratio is 0.010. This corresponds to about 1% strain. The choice of $\gamma_1 = 0.040$ and $\gamma_2 = 0.040$ is motivated by the maximum rotation allowed in the experiment.

Figures 23 and 24 show the effects of a surface slope of 0.04 on the observed fringe spacing and inclination for an intermediate level of in-plane stretch, about

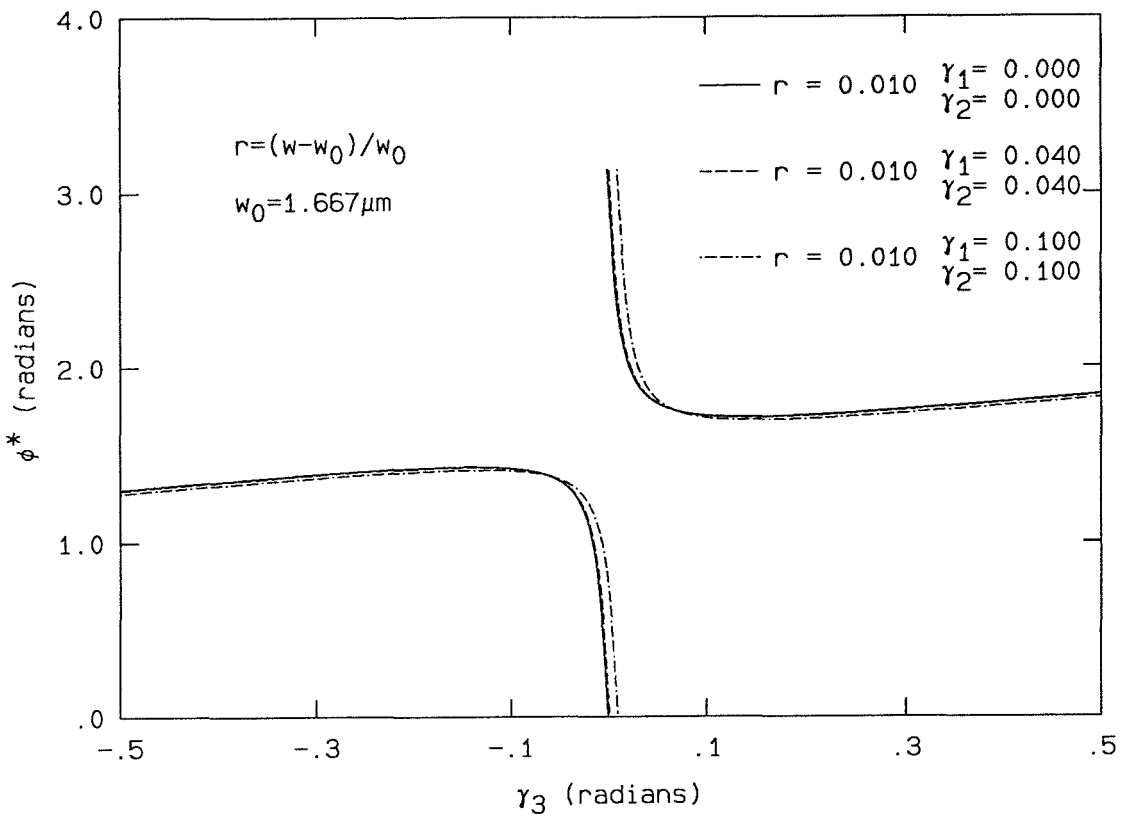


FIGURE 24. The variation for ϕ^* with out-of-plane rotations for the case where the stretch ratio is 0.010. This corresponds to about 1% strain. The choice of $\gamma_1 = 0.040$ and $\gamma_2 = 0.040$ is motivated by the maximum rotation allowed in the experiment.

1% strain. Also shown is the effect of a surface slope of 0.1. Again, with both $\gamma_1 = 0.040$ and $\gamma_2 = 0.040$, this is slightly more severe than a slope of 0.040.

It appears that the effects of the out-of-plane rotations can be neglected for slopes as large as 0.04, meaning that no correction will be necessary for the data gathered in the current study. In the future, the angular aperture of the camera will be increased, and it is clear that some correction may have to be made, using out-of-plane displacement data from the numerical model or experimental out-of-plane displacement data from the Twyman-Green interferometer.

The size scale at which such surface slopes occur can be estimated from the

output of the finite–element calculation. Figures 25 through 28 show contours of surface slope for the three highest loads at which comparisons will be made between the experiment and the finite–element model (Section 4.3), 35000 Newtons, 52300 Newtons and 73500 Newtons. The ultimate failure load of the three–point–bend specimen was about 78000 Newtons.

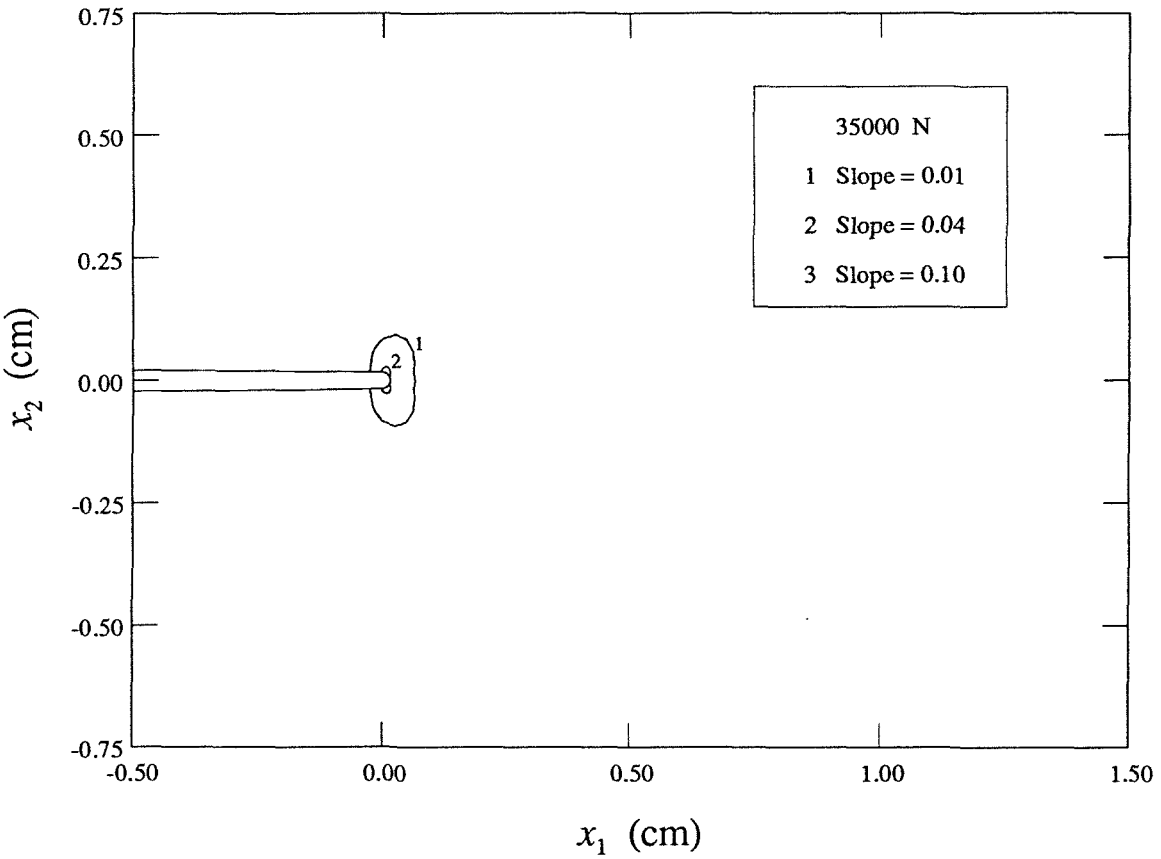


FIGURE 25. Contours of surface slope around the notch tip from the finite–element calculation. The load is 35000 Newtons; the ultimate failure load in the experiment was about 78000 Newtons. The plastic zone is about twice as large as contour 1, which corresponds to a slope of 0.010.

Contours representing surface slopes of 0.01, 0.04 and 0.10 are shown in the figures. In each figure, the plastic zone is about twice as large as the contour corresponding to a slope of 0.01. The contour for a slope of 0.04 represents the angular aperture limitation of the experiment, and matches well with the shadow spots observed in the experiment. From Figure 25, it can be seen that the surface

slope does not attain a magnitude of 0.10 anywhere on the specimen at a load of 35000 Newtons. From Figures 26 and 27 it can be seen that the slope ahead of the notch is not as severe as the slopes on either side of the notch tip. This is a result of the high shear deformation, which is present at the "corners" of the notch tip.

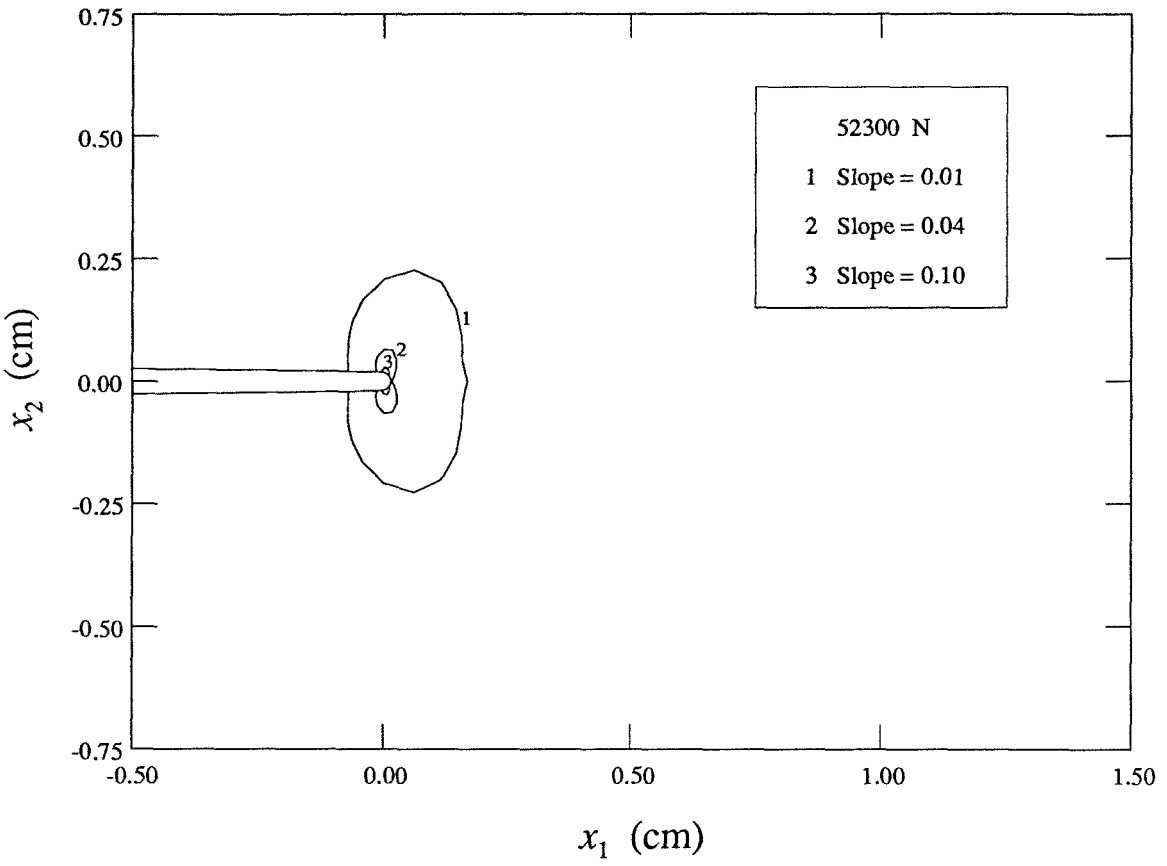


FIGURE 26. Contours of surface slope around the notch tip from the finite–element calculation. The load is 52300 Newtons; the ultimate failure load in the experiment was about 78000 Newtons. The plastic zone is about twice as large as contour 1, which corresponds to a slope of 0.010.

Figures 27 and 28 both show contours of surface slope at 73500 Newtons. The difference in the figures is that Figure 27 was calculated from a model that includes tunneling at loads above 63500 Newtons, and Figure 28 was calculated from a model that did not include tunneling. The tunneling of the crack is discussed in more detail in Section 4.5. The tunnel at 73500 Newtons is about 0.8 millimeter deep in the center of the specimen and only 0.03 millimeter long at the free surface. The profile

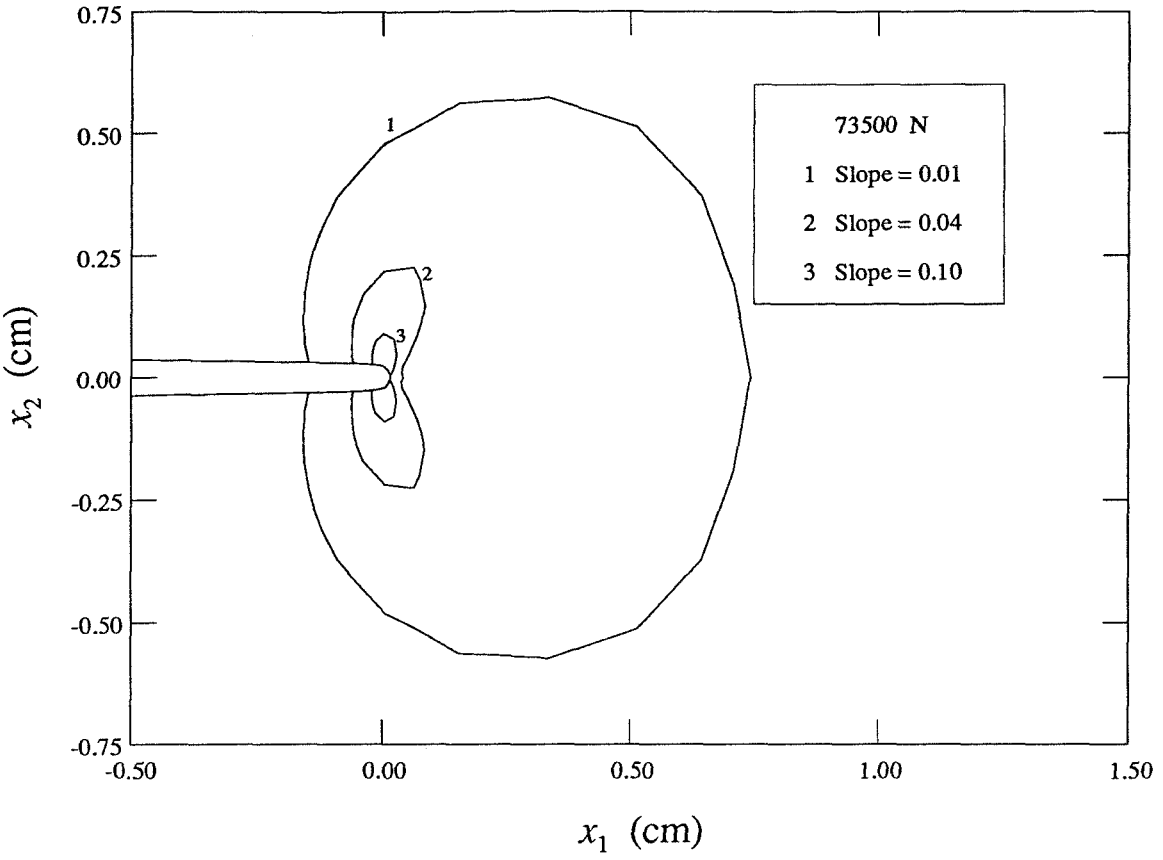


FIGURE 27. Contours of surface slope around the notch tip from the finite–element calculation. The load is 73500 Newtons; the ultimate failure load in the experiment was about 78000 Newtons. The plastic zone is about twice as large as contour 1, which corresponds to a slope of 0.010. This diagram is calculated from a model that includes the tunneling of the crack along the original plane of the crack. The tunnel at this load is about 0.8 millimeter deep in the center of the specimen, but only 0.03 millimeter long on the free surface.

of the tunnel is bell–shaped, somewhat like a Gaussian distribution. The two figures are included here to demonstrate that the effect of the tunnel is felt at distances much larger than the tunnel itself.

The contours representing surface slopes of 0.04 in Figures 25 through 27 match fairly well with the shadow spots observed in the photographs from the experiment. Actually, the shadow spots should be slightly larger than predicted by the contours, as the angular aperture limitation of 0.08 is determined at the center of the field of view. Away from the center of the field of view, the allowable angular deviation of

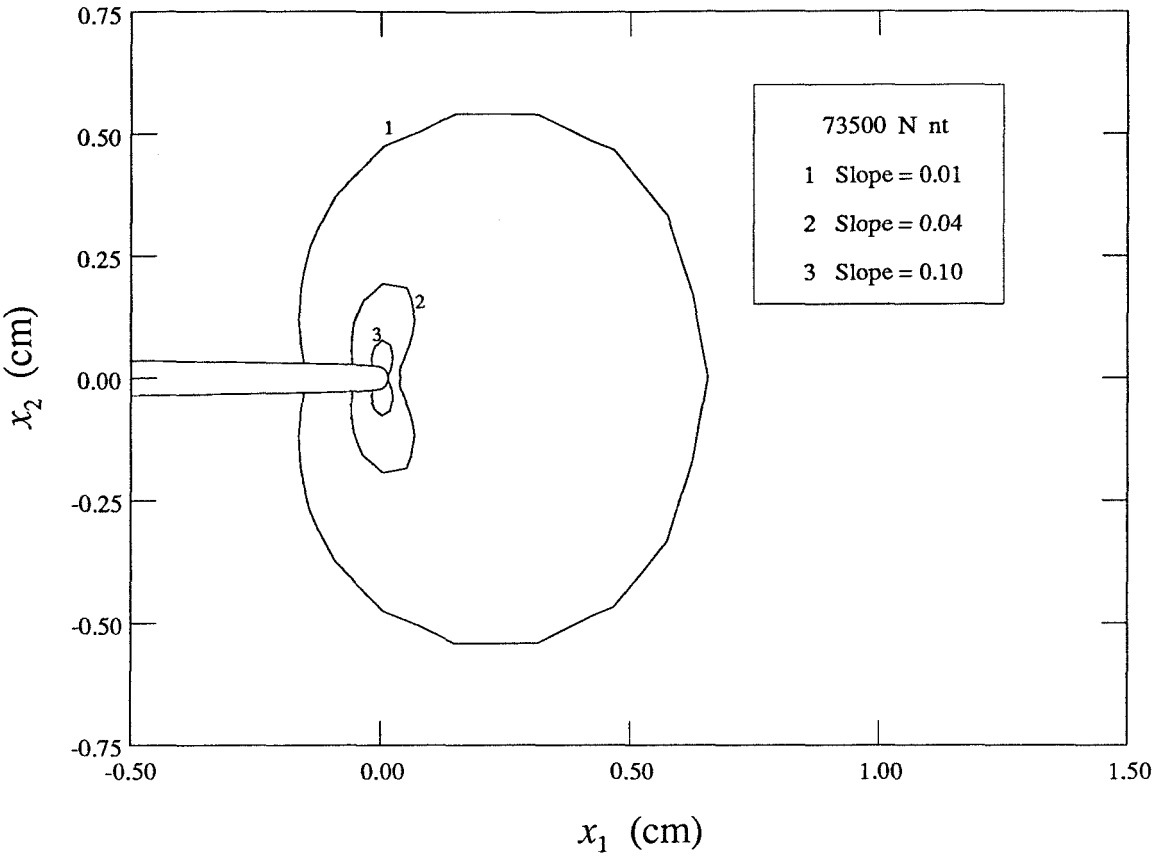


FIGURE 28. Contours of surface slope around the notch tip from the finite–element calculation. The load is 73500 Newtons; the ultimate failure load in the experiment was about 78000 Newtons. The plastic zone is about twice as large as contour 1, which corresponds to a slope of 0.010. This diagram is calculated from a model that does not include tunneling of the crack. This diagram is included to indicate that the tunnel has an effect on the deformation field at distances much larger than the tunnel itself.

light rays from the camera axis changes, becoming smaller on one side of the lens and larger on the other side of the lens. In most cases, at least one of the intersecting laser beams is deflected away from the lens when the surface slope is below 0.04, particularly at 73500 Newtons, where the out-of-plane displacement gradients are fairly large over a wide field.

2.3.14 A Final Word on the Effects of Tilting

It is clear that tilting of the plane of even an undeformed diffraction grating will introduce fringes. It would be useful if the fringes introduced corresponded to the displacement of the grating surface caused by the rotation. A simple counter–example shows that this is not the case. Refer to Figure 29, where the grating having wavelength w_0 (as indicated by the circles) has undergone a rigid rotation of γ_2 about the X_2 (or x_2) axis. The crossed laser beams form an interference pattern consisting of planes of constant intensity parallel to the X_2X_3 plane, so as to measure the displacement parallel to the X_1 axis. The displacement parallel to the X_1 axis of a point on the grating surface initially at position a measured along the X_1 axis is given by $a(\cos\gamma_2 - 1)$, assuming that the point at the origin of the coordinate frame remains fixed. The spacing of the rotated grating projected onto the fixed X_1X_2 plane is given by $w_0 \cos\gamma_2$.

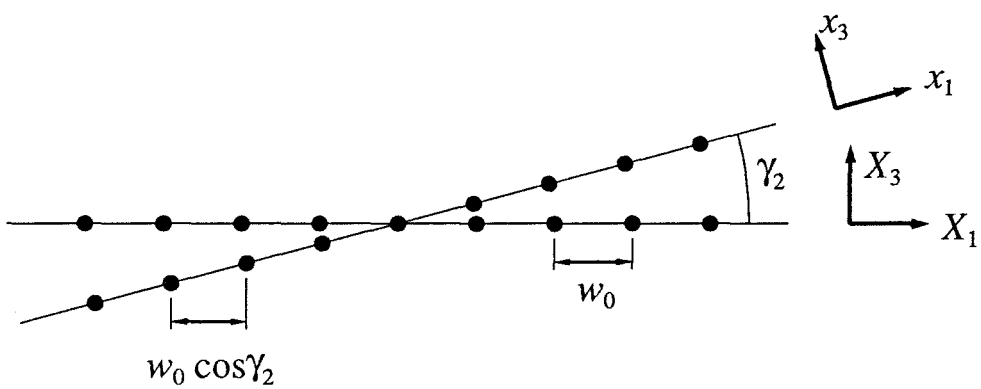


FIGURE 29. A rigid rotation of the grating by γ_2 about the X_2 axis. The spacing of the rotated grating with respect to the fixed X_1X_2 plane is given by $w_0 \cos\gamma_2$.

If the fringes observed using the moire interferometer do correspond to the displacement parallel to the X_1 axis caused by the rotation, then the fringe spacing and inclination can be found by assuming that this situation is equivalent to the

situation of two moire grids, a reference grating with spacing $w_0/2$ formed by the interfering laser beams, and a specimen grating with spacing $w_0 \cos\gamma_2$. Referring to Section 2.2.1, it is seen that the fringe spacing, δ, should be found as

$$\delta = \frac{w_0 \cos\gamma_2}{2(1-\cos\gamma_2)} \tag{2.3.69}$$

The fringe inclination, ϕ, should be identically zero for any nonzero γ_2. For $\gamma_2 = 0$, the fringe spacing is infinite, and the fringe inclination is undefined as $\tan\phi = 0/0$, since the gratings match exactly.

Equation (2.3.69) is the source of Post's estimation of the sensitivity of the moire interferometer shown in equation (1.2.1), using a Taylor expansion for $\cos\gamma_2$ and neglecting terms of order γ_2^2 with respect to unity. Note that he writes the equation in terms of the frequency of the fringes and the frequency of the virtual reference grating formed by the crossed laser beams; the frequency is just the inverse of the wavelength. The rotation of the grating will always cause a displacement that corresponds to a compression, and Post suggests that the frequency of these fringes will be

$$F = f\gamma_2^2/2 \tag{2.3.70}$$

where f is the frequency of the virtual reference grating, and γ_2 is the angle of rotation about the X_2 axis.

It will be shown that fringes observed using the moire interferometer do not correspond to the displacement that is due to rotation, even in this simple case. To show this is true, the fringe spacing of equation (2.3.69), and the accompanying fringe inclination ($\phi = 0$), will be compared to the observed fringe spacing determined from the equations calculated using the method of Section 2.2.1.

The actual fringe spacing and inclination that are due to the rigid rotation cannot be found using the small–angle approximation. In this case, the angles γ_1 and γ_3 are identically zero, and there is no stretching, so $w = w_0$. When these values for w, γ_1 and γ_3 are used in equations (2.3.65) and (2.3.68), they yield an infinite fringe spacing and an undefined fringe inclination, as would be the case if $\cos\gamma_2$ were replaced by 1 in equation (2.3.69) above. The small–angle approximation fails

because of the coupling of the various terms in equations (2.3.65) and (2.3.68). Interestingly, this means that if there is a *small* rotation about the X_2 axis only, there are no fringes introduced by that rotation. A similar result occurs if there is a *small* rotation about the X_1 axis only. Actually, even a large rotation about the X_1 axis does not introduce any fringes, as the diffracted laser beams will still overlap exactly. However, a large enough rotation will cause all or part of the diffracted beams to fall outside the camera aperture.

It is necessary to return to equations (2.3.31) and (2.3.32) to investigate the effect of a finite rotation about the X_2 axis only. Making the substitutions $\gamma_1 = \gamma_3 = 0$ in equations (2.3.31) and (2.3.32), the equations for the unit vectors describing the diffracted plane waves are given by

$$\begin{aligned} u_1 &= \frac{\lambda}{w} - \sin\alpha\cos\gamma_2 + \cos\alpha\sin\gamma_2 \\ u_2 &= 0 \\ u_3 &= [1-(u_1)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.71}$$

$$\begin{aligned} v_1 &= -\frac{\lambda}{w} + \sin\alpha\cos\gamma_2 + \cos\alpha\sin\gamma_2 \\ v_2 &= 0 \\ v_3 &= [1-(v_1)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.72}$$

The components of the difference vector $\mathbf{z} = \mathbf{u} - \mathbf{v}$, which characterizes the fringe pattern in the rotated grating coordinate frame, $x_1x_2x_3$, are then found as

$$\begin{aligned} z_1 &= 2\left(\frac{\lambda}{w} - \sin\alpha\cos\gamma_2\right) \\ z_2 &= 0 \\ z_3 &= [1-(u_1)^2]^{\frac{1}{2}} - [1-(v_1)^2]^{\frac{1}{2}} \end{aligned} \tag{2.3.73}$$

For a rigid rotation of the grating, the deformed grating wavelength w is replaced by the undeformed grating wavelength w_0 in equations (2.3.71), (2.3.72) and (2.3.73). It should be recalled that $\sin\alpha = \lambda/w_0$.

As was shown earlier, the fringe planes are perpendicular to the vector $\mathbf{z}$ with a spacing δ given by $\delta = \lambda/|\mathbf{z}|$.

The unit vectors parallel to the fixed X_1, X_2 and X_3 axes, $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$ are found as

$$\begin{aligned} \mathbf{E}_1 &= \cos\gamma_2 \mathbf{e}_1 + \sin\gamma_2 \mathbf{e}_3 \\ \mathbf{E}_2 &= \mathbf{e}_2 \\ \mathbf{E}_3 &= -\sin\gamma_2 \mathbf{e}_1 + \cos\gamma_2 \mathbf{e}_3 \end{aligned} \tag{2.3.74}$$

where $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ are the unit vectors aligned with the rotated $x_1x_2x_3$ grating coordinate frame.

As in Section 2.3.8, the observed fringes are determined by the vector that is the projection of $\mathbf{z}$ on the X_1–X_2 plane. Thus, the *observed* fringe spacing, δ^*, is given by

$$\delta^* = \frac{\lambda}{|\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)\mathbf{E}_3|} \tag{2.3.75}$$

which can also be written as

$$\delta^* = \frac{\lambda}{[\mathbf{z}\cdot\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)^2]^{\frac{1}{2}}} \tag{2.3.76}$$

The observed fringe inclination, ϕ^* is determined as

$$\tan\phi^* = \frac{[\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)\mathbf{E}_3]\cdot\mathbf{E}_2}{[\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)\mathbf{E}_3]\cdot\mathbf{E}_1} = \frac{\mathbf{z}\cdot\mathbf{E}_2}{\mathbf{z}\cdot\mathbf{E}_1} \tag{2.3.77}$$

From equations (2.3.73), (2.3.74) and (2.3.77), it can be seen that ϕ^* is identically zero for nonzero γ_2, and for $\gamma_2 = 0$, ϕ^* is undefined as $\tan\phi^* = 0/0$.

It remains only to show that δ^* is not equal to the fringe spacing of equation (2.3.69). Again making use of a two–term Taylor expansion for u_3 and v_3, so that $u_3 = 1 - (u_1^2/2)$ and $v_3 = 1 - (v_1^2/2)$, the observed fringe spacing δ^* is found as

$$\delta^* = \frac{w_0}{2(1-\cos\gamma_2)\,|\cos\gamma_2 - \cos\alpha\sin^2\gamma_2|} \tag{2.3.78}$$

The use of the Taylor expansion for u_3 and v_3 is quite good for very small values of γ_2; the deviation between the δ^* of equation (2.3.78) and the true δ^* calculated without any approximation for u_3 or v_3 is less than 5% for γ_2 less than 0.5 radian (30 degrees), and less than 0.2% for γ_2 less than 0.25 radian (15 degrees).

On the other hand, the deviation between δ as calculated in equation (2.3.69) and δ^* as calculated in equation (2.3.78) is on the order of 50% for $\gamma_2 = 0.5$ radian (30 degrees), and about 12% for $\gamma_2 = 0.25$ radian (15 degrees). For $\gamma_2 = 0.1$ radian (6 degrees), the deviation is about 2 %.

It may thus be seen that the fringes caused by tilting of the grating plane are not equivalent to fringes that correspond to the displacement parallel to the X_1 axis caused by such tilting, even in this simple case. However, the deviation between the two fringe patterns would not be too large for angles up to about 0.1 radian.

A similar result holds if the grating is allowed to stretch or shrink from the initial wavelength w_0 to a new wavelength w before the rotation about the X_2 axis occurs. The fringe spacing predicted from a comparison of a moire grid having spacing w_0 to a grid whose spacing is the projection of the final wavelength on the X_1X_2 plane, $w\cos\gamma_2$, is

$$\delta = \frac{w_0\, w\cos\gamma_2}{2\,|w_0 - w\cos\gamma_2|} \tag{2.3.79}$$

This is analogous to the fringe spacing calculated in equation (2.3.69) from the situation shown in Figure 29. The fringe inclination ϕ is identically zero for $w\cos\gamma_2 \neq w_0$; otherwise, it is undefined.

Using the procedure outlined above in equations (2.3.71) through (2.3.78), the observed fringe spacing in the case where the grating is allowed to stretch or shrink from wavelength w_0 to wavelength w is given by

$$\delta^* = \frac{ww_0}{2\,|w_0 - w\cos\gamma_2|\,|\cos\gamma_2 - \cos\alpha\sin^2\gamma_2|} \tag{2.3.80}$$

The observed fringe inclination, ϕ^*, is identically zero for $w\cos\gamma_2 \neq w_0$; otherwise, it is undefined.

It can be seen from comparing equations (2.3.79) and (2.3.80) that although the fringe spacing predicted using the two approaches do not match, the trend in the fringe spacing is similar in that the fringes in both cases depend on the relative magnitude of w_0 as compared to $w\cos\gamma_2$. The term $|w_0 - w\cos\gamma_2|$ in equations (2.3.79) and (2.3.80) acts to include the effect of the displacement that is due to the

rotation on the fringe spacing. As seen in Figure 29, the effect of rotation should be to make the rotated grating seem to have a smaller wavelength compared to the unrotated grating. If the grating has been compressed so that w is less than w_0, the rotation about the X_2 axis should make the grating wavelength seem even smaller, which it does as $w \cos\gamma_2 \leq w$. This produces fringes that are closer together. If the grating has been stretched so that w is greater than w_0, the effect of the rotation causes the grating to appear to have stretched less than it actually has, so that the fringes move farther apart until $w \cos\gamma_2 = w_0$. At that point, the fringe spacing is infinite. If γ_2 continues to increase, the projected wavelength $w \cos\gamma_2$ will be less than w_0, corresponding to a contraction of the grating, and the fringes will again move closer together. If w_0 is identical to $w \cos\gamma_2$, the fringe spacing is infinite in both equation (2.3.79) and equation (2.3.80).

The other term in the denominator of equation (2.3.80), $|\cos\gamma_2 - \cos\alpha \sin^2\gamma_2|$, is always less than or equal to unity, if γ_2 is still small enough so that the Taylor expansion approximation for u_3 and v_3 is appropriate. On the other hand, equation (2.3.79) contains the term $\cos\gamma_2$ in the numerator, and $\cos\gamma_2 \leq 1$ for the range of allowable γ_2. Thus, the fringe spacing predicted by equation (2.3.80) will always be greater than the fringe spacing predicted by equation (2.3.79), for nonzero γ_2. The deviation between the fringe spacing calculated from equation (2.3.79) and the fringe spacing calculated from equation (2.3.80) is also on the order of 50% for $\gamma_2 = 0.5$ radian (30 degrees), and about 12% for $\gamma_2 = 0.25$ radian (15 degrees). For $\gamma_2 = 0.1$ radian (6 degrees), the deviation is about 2 %. This is the same as for the case where there is no change in the grating wavelength.

If γ_2 is small, both $\cos\gamma_2$ and $|\cos\gamma_2 - \cos\alpha \sin^2\gamma_2|$ are close to unity, so that the fringe spacing is primarily controlled by the magnitude of the term $|w_0 - w\cos\gamma_2|$.

For these simple examples of tilting the grating plane, the fringe spacing observed using the moire interferometer does not exactly correspond to the fringe spacing that would be seen if the displacement that was due to the rotation were included in the analogy of comparing two moire grids. However, the fringe spacing does follow the trend of the displacement by comparing the undeformed grating

wavelength, w_0, with the projection of the deformed grating wavelength on the fixed X_1X_2 plane, $w\cos\gamma_2$. Thus, for small out-of-plane displacement gradients, a reasonable first approximation to including the effects of the rotation of the grating plane may be to assume that the fringe patterns also capture the displacement that is due to the rotation of the grating plane.

If there is no change in the plane of the diffraction grating, the equations describing the fringes observed using the moire interferometer are valid even in the presence of large deformations. It also appears that the displacement that is due to rotation of the grating plane is also nearly accounted for in the fringe patterns, if the rotation of the grating is small enough. Thus, it seems that moire interferometry can be used in the presence of large in-plane deformations as long as the out-of-plane deformation gradients are small enough, if it is recognized that the fringes are observed on the surface of the deformed configuration. The greatest difficulty with the technique is then introduced by the necessity of relating the displacements mapped on the deformed configuration back to the initial undeformed configuration.

2.3.15 Inhomogeneous Deformations

The equations describing the fringe spacing and inclination for an arbitrary (nonplanar) deformation have been derived under the assumption that the deformation is homogeneous; this assumption implies that the deformed grating remains periodic. For continuum deformation there will be a size scale at which any deformation can be considered locally homogeneous, but that size scale may be actually smaller than the size scale of the grating wavelength. If such is the case, the formation of the fringe patterns cannot be investigated using the plane wave assumption of Sections 2.3.1 through 2.3.3, which relies on the diffraction of plane waves from a deformed, but still periodic, grating. Presumably, if the inhomogeneous deformation is smooth enough on some scale relative to the scale of the grating, the directions of the diffracted waves will also vary smoothly, and the interference pattern will still provide a map of the deformation field. It is this assumption of sufficient smoothness in the deformation and interference patterns that is used to extend the application

of moire interferometry in the presence of inhomogeneous deformations.

Lack of periodicity of the grating will certainly occur at edges of the grating as well as at discontinuities in the deformation field, if any. However, even smoothly varying displacement fields would result in nonperiodic gratings on some scale. It would be necessary to examine the formation of fringes by diffraction from a grating that is not periodic in order to determine the validity of the assumption that smoothness in the deformation field results in smoothly varying interference fringe patterns. In effect, the question is whether the diffraction angle predicted by equation (2.3.23) remains valid at any point where the deformed grating wavelength, w, is not constant but a smoothly varying function of position.

Unfortunately, analytical solutions of general diffraction problems are difficult to obtain, as the closed–form solutions for diffraction problems invariably rely on a convenient coordinate frame in which a separation of variables may be employed. This method of solution cannot be used in case of diffraction from a grating, so some sort of approximate technique must be employed.

Most of the work analyzing diffraction gratings is aimed at predicting the diffraction efficiency of a given grating profile. A periodic grating profile implies periodicity in the diffracted light waves, and if the incident light waves are plane waves, then the diffracted light waves are also plane waves aligned with the directions of the various, discrete diffraction orders [81]. The periodicity of the grating ensures that the diffracted light will be distributed through the diffraction orders in plane waves, and the plane waves provide a basis for a series approximation to the diffraction problem. Petit *et al.*[55] use this result to predict the diffraction efficiency of various grating profiles by assuming a series solution in plane waves along each of the discrete diffraction orders and matching the amplitudes of those plane waves to the boundary conditions imposed by the grating profile. (There is some question as to whether this approach is valid near the grating surface, as the plane wave solutions resulting from the periodicity of the grating profile may not apply in the vicinity of the grating surface, particularly in deep grooves, as discussed in Reference [55].)

One possible way to estimate the effect of inhomogeneous deformation is suggested by Figure 10, in which the diffraction grating is compared to a series of equally spaced line sources. Perhaps the inhomogeneous problem could be attacked by using a series of line sources that are not equally spaced, at least in some finite region, and summing over the cylindrical monochromatic waves that emanate from the sources. The phases of the sources would be matched to the phases of the plane waves of the laser beams that impinge on the grating from opposite sides. These line sources could be distributed even over a curved surface, and the line sources themselves curved (or made up of smaller line segments) to include all displacement components.

Another approximate technique for attacking such a problem might be through a finite–element approach, or perhaps a boundary–element approach, in which the deformed grating profile would be modeled as a boundary condition for the solution of the harmonic wave equation in free space. The boundary–element approach might be preferred, as only the grating surface itself would need to be discretized.

2.3.16 Concluding Remarks

One important feature of the moire interferometer should be stressed, which is that the equations describing the fringe spacing and inclination do not depend on any assumption that the deformation is small, providing that there is negligible tilting of the specimen surface. The fringes can be thought of as contours of relative displacement with respect to the reference grating at any deformation level. However, at large displacement levels, one must be concerned about locating the fringes on the deformed specimen surface, and about how to relate locations on the deformed body with respect to the reference configuration of the undeformed body, if that should be necessary.

In undertaking the derivation of the equations describing the fringes for a general (not necessarily planar) deformation, the goal was to determine the effect of deformations that rotate the plane of the grating surface. This investigation was motivated by the fact that the surface near the tip of a crack dimples inward.

The dimple is confined to a small region, and even within this region, the surface slopes appear to be well approximated by small angles over most of the dimple, as shown by the finite–element model, and from data taken using the Twyman-Green interferometer.

Two approximations were made to obtain the fringe spacing and inclination in equations (2.3.65) and (2.3.68). The first approximation was a two–term Taylor expansion for the normal component of the unit vectors representing the diffracted beams. The second approximation was a small–angle approximation for the angles representing the tilt of the plane of the grating, letting $\sin\gamma = \gamma$ and $\cos\gamma = 1$.

The two–term approximation for the out-of-plane component of the unit vectors representing the diffracted beams, u_3 and v_3, is quite good for the experimental moire interferometer. Initially, u_3 and v_3 are arranged so that $u_3 = v_3 = 1$, which means that the diffracted beams travel along the surface normal, and the expansion for u_3 and v_3 is made about this initial condition. The error in this approximation is less than 1 percent for deviations of the diffracted beams away from the surface normal up to 30 degrees, while the theoretical limit that can be captured using the current configuration of the interferometer is about 20 degrees.

The small–angle approximation to the general equations simplifies the equations somewhat and helps to determine the leading sources of discrepancy in the fringe–formation equations, but the equations retaining the sine and cosine terms could be used as well. It appears from the finite–element analysis that to a large extent the surface rotation will be in the range where small–angle theory is valid. In the experiment, any large surface rotation also means large deviations from the original paths taken by the diffracted laser beams to the camera, which means that some light will be lost. Additionally, when the surface rotations are beyond small angles, the out-of-plane displacement measurement made using the Twyman-Green interferometer may need to have a correction because of the large angles.

This investigation has treated the tilting of the grating plane as a source of error in the fringe patterns that would be expected for the same in-plane deformation of the grating (*i.e.*, the same values of the deformed grating wavelength w and rotation

γ_3). The tilting is considered as a rigid rotation applied to a deformed grating, with the goal being the determination of the deformation before any rigid rotation takes place. This seems to be a reasonable way to look at things if the tilting is very small (nearly zero). If the tilting is not so small, it would be better to follow up the investigation of Section 2.3.14, in which the displacement that is due to the rotation is included in the comparison between fringe patterns for the special case of rotation about only one axis. The fringes formed by an arbitrary (homogeneous) deformation and rotation of the grating would then be compared to the fringes predicted through the comparison of two moire grids, a reference grating formed by the intersecting laser beams and a specimen grating with spacing and orientation determined as the projection of the deformed grating on the original grating plane. Differences between these two fringe patterns would be the discrepancy between a pattern corresponding to the deformation that includes the rotation component and a pattern that would be observed using the moire interferometer. The equations derived in this section may be used to make this comparison, but the comparison should be carried out with those equations that do not include the small-angle approximations.

2.4 Description of the Apparatus and the Experimental Technique

In order to measure surface displacements using moire interferometry, a diffraction grating must first be attached or etched onto the surface of the object to be measured. This diffraction grating is a periodic, reflective, surface height variation; its period is on the order of the wavelength of the light used in the measurements. Additionally, two laser beams must be brought to the specimen at a precise angle and held fixed in space. These laser beams must originate from the same source and must follow paths of nearly equal length to the specimen surface.

The displacement information is contained in interference fringes that have been shown to represent contours of constant in-plane displacement. The increment of displacement between adjacent contours is proportional to the period of the diffraction grating and so is itself on the order of the wavelength of light used; the smallest theoretically possible displacement increment between contours is one-half of that wavelength. This means that moire interferometry is very sensitive to small displacement gradients, and that the fringe density becomes very high when the strains become large. This measurement technique displays the displacement contours over a field that is limited by the size of the diffraction grating and the diameter of the laser beams used to create the interference pattern. However, since the light that travels to the observer is nearly collimated, the collection device must have an aperture as large as the measurement area in order to collect all of the information from the object. As the specimen deforms, the light diverges from its initial path, and at some point the aperture will not be large enough to capture all of the information from the specimen. Ordinarily, this situation would occur first at the edge of the field of view, but in the case of fracture, the deformation may be so large at the center of the field of view near the crack tip that this region also diverts the light outside the range of the collection apparatus.

The out-of-plane deformation gradients may also cause the light to be diverted beyond the range of the optical apparatus. Again, this phenomenon is most apparent at the tip of a crack in fracture mechanics, leading to a shadow spot around the

crack tip.

Such a shadow spot should not be confused with a caustic. The shadow spot observed with the moire interferometer is due to a loss of information caused by a failure of the camera optics to transmit all of the light from the specimen to the recording film. A caustic is the result of a folding of the mapping of the light from the specimen to the film, so that light from different points on the specimen falls on the same point of the film, and some of the film is left unexposed. Caustics can appear in moire interferometry if the camera is not well focussed on the surface of the specimen, but the size of the caustic can be kept smaller than the shadow spot caused by the inability of the camera to capture some of the light from the specimen.

2.4.1 Fringe Quality

The measurement relies on the interference pattern between two laser beams diffracted from the surface of the specimen. In order to produce interference patterns of high contrast, the two laser beams must be mutually coherent. Actually, this is somewhat of a circular definition, since coherence between two sets of light waves is determined by measuring the contrast of the interference pattern between them. In effect, coherence is a measure of how much the two sets of waves resemble each other. If the two waves are similar, and if they are out of phase with one another, they will cancel completely. In order to assure coherence, the two laser beams must originate from the same laser and travel paths of similar length to the object. Although lasers are nearly monochromatic, the wavelength of the laser beam does vary slightly with time (and therefore with distance from the laser). This variation is often approximated by considering the laser light to be the sum of two waves with wavelengths slightly above and below the nominal laser wavelength, causing a beat phenomenon. These two waves interfere with each other and cause the wave pattern to vary slightly in time (and in space as the waves travel), while still retaining the nominal laser wavelength. If the difference in wavelength between these two waves is $\Delta\lambda$ and the nominal wavelength of the laser is λ_0, the path-length difference

between the two beams ΔL should be such that

$$\Delta L < \frac{1}{8}\frac{\lambda_0^2}{\Delta\lambda} \tag{2.4.1}$$

for high contrast fringes. Although the laser output is more complex than the sum of two waves, this relation is still a good estimate of the allowable path–length difference, as described by Vest [86], if $\Delta\lambda$ represents the bandwidth of the light waves from the laser. This allowable path–length difference is called the coherence length of the laser, and for a Helium-Neon laser, it is on the order of 20 cm, so that for interferometric purposes the difference in path lengths between the two beams of the interferometer should be less than 20 cm and as nearly equal to zero as possible. This notion of the coherence length is actually a description of the temporal coherence of the laser because it describes the variation of the laser output with time. It is related to a length because of the propagation of the light waves and because that length is of a scale more easily comprehended (light travels over the coherence length in 6.67×10^{-10} seconds).

A second important coherence measurement is the spatial coherence of the light waves, which describes the similarity of the waves across the width of the beam. A laser source has excellent spatial coherence, in contrast to a diffuse monochromatic source, for instance, such as a sodium–vapor lamp, where the light emitted from two separate points on the source may have different phases, amplitudes and polarizations, even though they have the same wavelength.

The coherence of the interfering light waves is the reason that the two interfering laser beams must come from a single source, since matching the fluctuation in the output from two different lasers is effectively impossible.

Besides matching the path lengths to maintain the temporal coherence of the laser beams, the amplitude and polarization of the two laser beams should be similar to provide high–contrast fringes, as seen in Section 2.3.3, where the spatial distribution of the fringe planes was determined. The contrast of the interference fringes between two monochromatic plane waves was shown to be

$$\text{contrast} = \frac{\mathrm{I}_{max} - \mathrm{I}_{min}}{\mathrm{I}_{max} + \mathrm{I}_{min}} \tag{2.4.2}$$

where I_{max} is the maximum intensity observed in the interference pattern, and I_{min} is the minimum intensity observed in the interference pattern. In Section 2.3.3, the electric–field vectors of the two interfering plane waves were used to determine the intensity distribution of the interference. For harmonic, monochromatic plane waves, the electric–field vector can be decomposed into a constant amplitude vector multiplying a harmonic function of time and position. Since the intensity is a time average of the energy carried by the light, the harmonic function averages out, and only the amplitude vector is needed to determine the intensity. Let $\mathbf{A}_1$ be the amplitude vector for the electric field of the first set of plane waves, and $\mathbf{A}_2$ be the amplitude vector for the electric field of the second vector. Assuming that the light is linearly polarized, the vectors $\mathbf{A}_1$ and $\mathbf{A}_2$ are real–valued. The contrast is then found to be

$$\text{contrast} = \frac{2|(\mathbf{A}_1 \cdot \mathbf{A}_2)|}{|\mathbf{A}_1|^2 + |\mathbf{A}_2|^2} \tag{2.4.3}$$

From equation (2.4.3), it may be seen that if the light waves are polarized so that the electric fields are at 90° to each other, the contrast of the fringes is zero. Also, if the magnitude of the vectors is very different, the contrast also tends to zero. The maximum contrast occurs if the vectors are aligned exactly and the amplitudes match exactly, in which case the contrast is unity.

2.4.2 The Moire Interferometer

The moire interferometer used in these experiments employs two Spectra–Physics *Model 120 S* 15 mW Helium–Neon lasers, one for each of the two in-plane displacement components to be measured. It also depends strongly on adjustable mirrors and beam splitters for precise initial alignment with the diffraction grating. The mirrors, beam splitters and alignment hardware were all obtained from the Newport Corporation, Fountain Valley, California. For each displacement component, the laser beam passes through a beam expander and lens to produce a collimated laser beam of 50 mm diameter. This beam is then reflected from a mirror and passed through a beam splitter, where half of the beam is reflected and the other half transmitted. Each of the two resulting beams from the beam split-

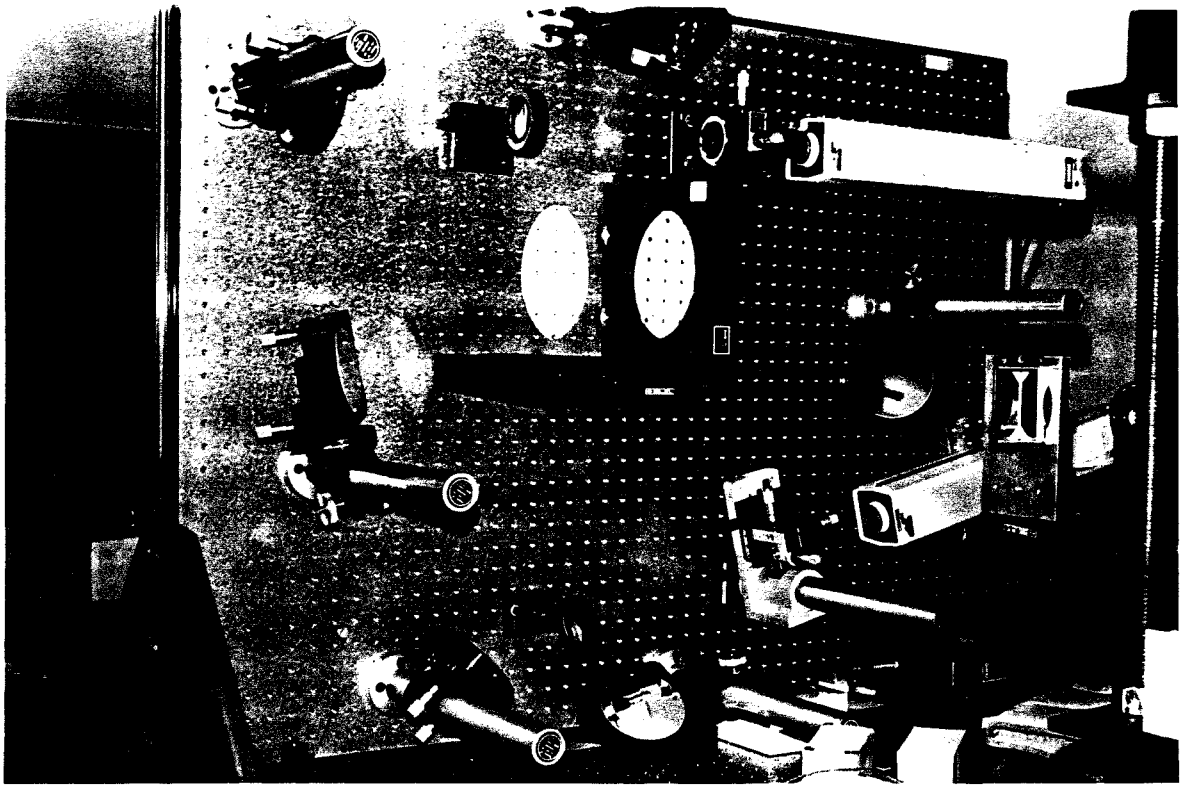


FIGURE 30. The moire interferometer. In the center are the two six–inch beam splitters. The beams used to determine the horizontal displacement are directed by the mirrors at left and right; the beams for the vertical displacement are at top and bottom. All optics are mounted on a Newport Corporation breadboard that measures three feet by four feet. The breadboard is held in a vertical plane on top of a Newport Corporation vibration isolated optical table. All four laser beams intersect approximately one meter from the breadboard surface.

ter travels to a mirror that directs the beam to the specimen surface. See Figures 30 through 35. This arrangement, consisting of a laser, a beam splitter and three mirrors is used twice, once to provide laser beams that cross in a vertical plane to measure the vertical component of displacement, and once to provide laser beams that cross in a horizontal plane, measuring the horizontal component of displacement. All of the mirrors and beam splitters are angularly adjustable by means of micrometers, which rotate the optical components about fixed pivots in their mounts. This feature of the interferometer makes it quite flexible and provides ease of alignment of the laser beams so that they intersect at the specimen surface at

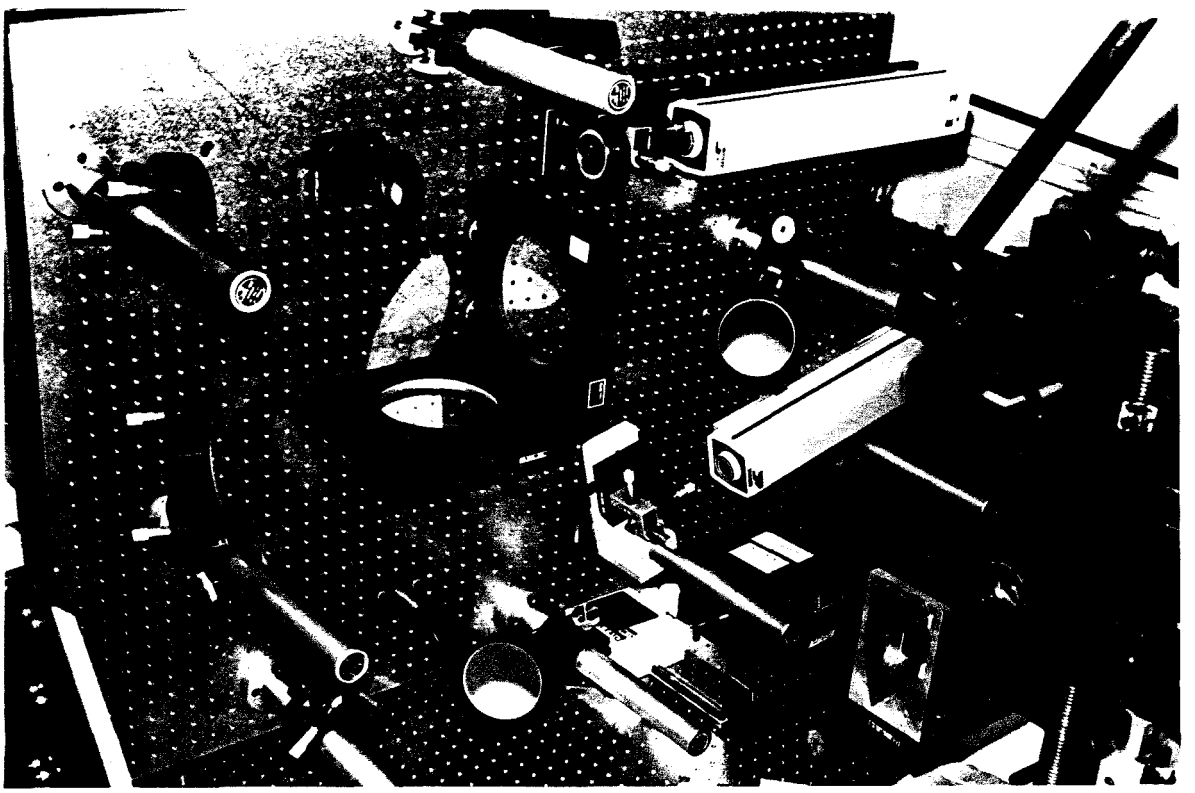


FIGURE 31. The moire interferometer. A second view shown from above. At the lower right can be seen the prism used to direct the diffracted beams to the camera, and the camera lens itself. The breadboard and optical table are covered by tapped holes on one inch centers.

the proper angle and at a specific location.

Photographs of the moire interferometer used in the experiments are shown in Figures 30 and 31. The arrangement of optical components is spelled out more clearly in Figures 32 and 33, where the two interferometers, one each for the horizontal and vertical components of displacement, are displayed separately.

All of the components are mounted on a Newport Corporation breadboard measuring three feet by four feet, and the breadboard is held upright in a vertical plane on a vibration–isolated optical table, also from the Newport Corporation; the table is ten feet by four feet by eighteen inches thick.

As mentioned, for high fringe contrast, the paths of each pair of laser beams must be nearly equal, the amplitudes of the laser beams must be nearly equal, and

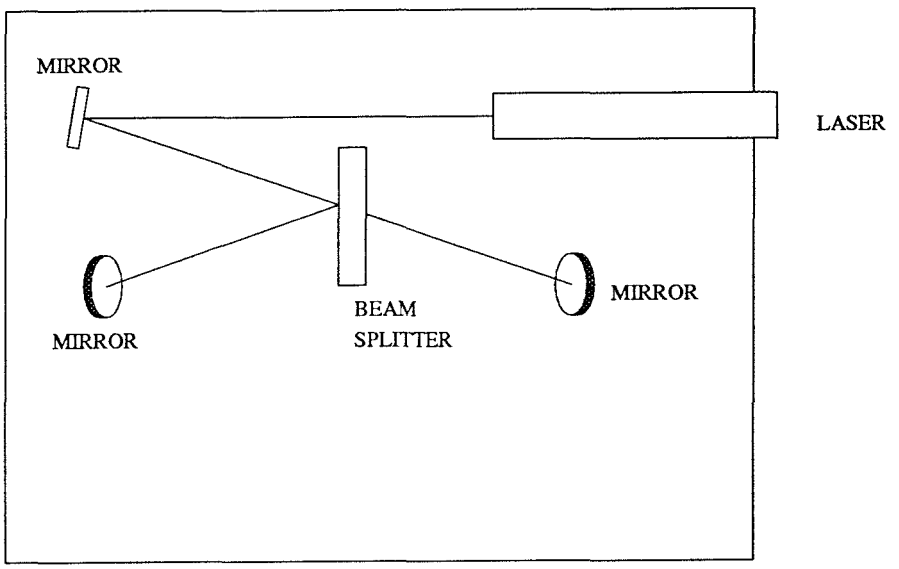


FIGURE 32. Diagram of the interferometer configuration to measure the horizontal component of displacement.

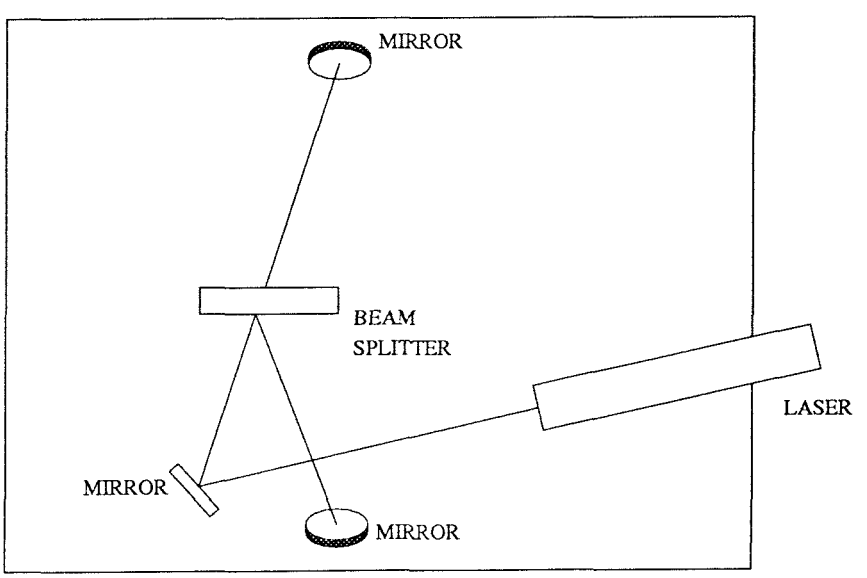


FIGURE 33. Diagram of the interferometer configuration to measure the vertical component of displacement.

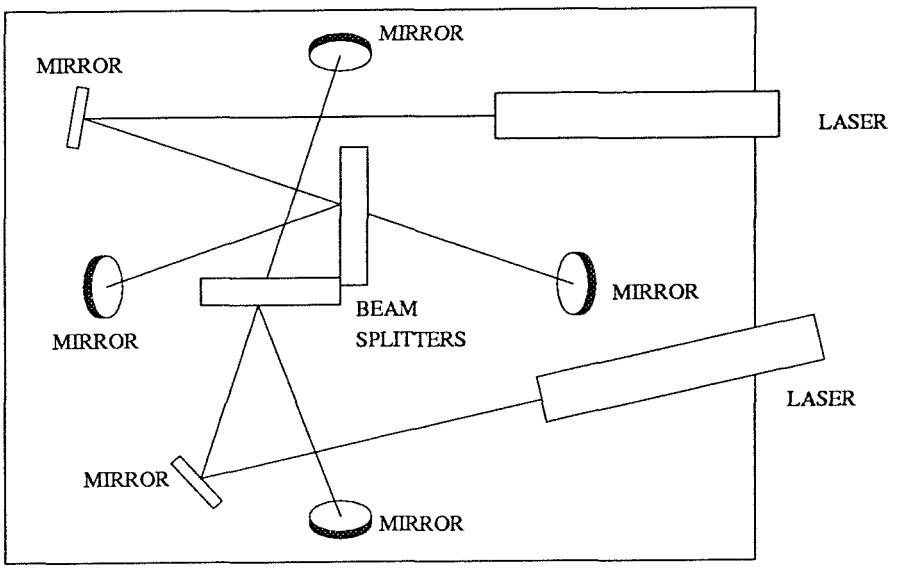


FIGURE 34. Diagram of the interferometer configuration to measure both the horizontal and vertical components of displacement.

the polarization of the laser beams should be in the same direction. As can be seen in Figures 32 and 33, the lengths traveled from the laser to the two mirrors at the end of each traced path are equal. The additional distance to the specimen and then to the camera is also equal for all four beams. The balancing of the amplitudes of the beams depends on the beam splitters, which have a ratio of reflected intensity to total intensity that varies between about 0.4 to 0.5, depending on the angle between the laser beam and the beam splitter and on the polarization of the laser beam. For the interferometer shown in the photographs, the ratio of the reflected intensity to total intensity is about 0.4. The difference in polarization between the two laser beams can be seen in Figures 33 and 32; the angle between the directions of polarization is approximately the angle between the paths on either side of the beam splitter. For the interferometer in the figures, this angle is 38.5°, the cosine of which is 0.782, which is required by the inner product of equation (2.4.3). The angle between the polarization directions of the incoming laser beams is retained in the diffracted laser beams.

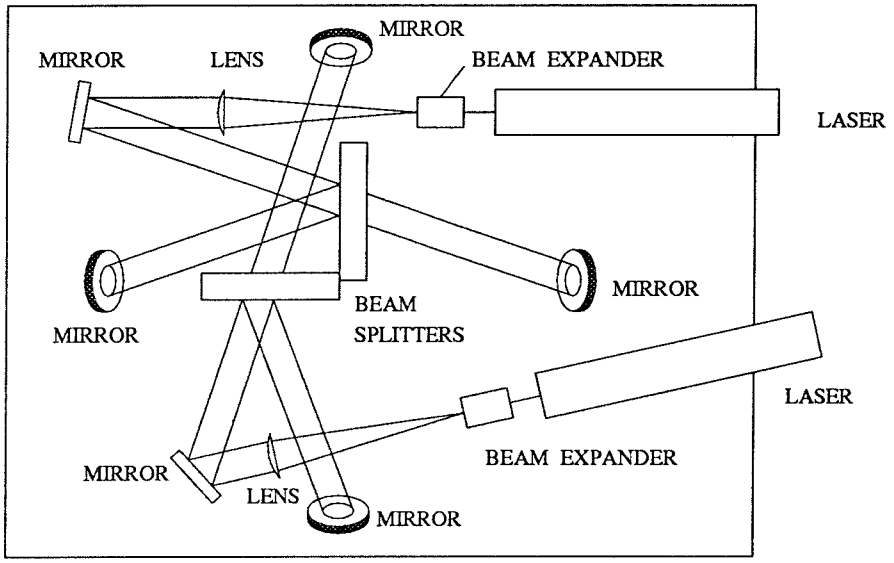


FIGURE 35. Diagram of the interferometer configuration to measure both of the displacement components with optics for beam expansion and collimation.

2.4.3 Interferometer Alignment

The interferometer alignment is begun by directing the unexpanded laser beams to the specimen. The beam that passes through the beam splitter is aligned first by adjusting the angles of the mirror that lies between the laser and the beam splitter (Mirror A in Figure 36) and the the mirror which the laser beam reaches after passing through the beam splitter (Mirror B in Figure 36). By adjusting these mirrors alternately, the laser beam can be brought to the specimen surface at the desired location and at the proper angle. The angle can be checked by tracing the path of the second–order beam diffracted from the specimen. Since the first–order diffracted beam should exit along the normal to the specimen surface, the second–order diffracted beam should return along the path of the incoming beam. Although this order carries considerably less of the energy than the zero or first–order beams, it is visible when using the unexpanded laser beam. After the beam that passes through the beam splitter is aligned, the beam reflected from the beam splitter is similarly aligned by alternately adjusting the angles of the beam splitter

and the mirror the beam reaches after reflecting from the beam splitter (Mirror C in Figure 36). Slight angular changes in the orientation of the beam splitter do not significantly alter the path of the laser beam passing through it.

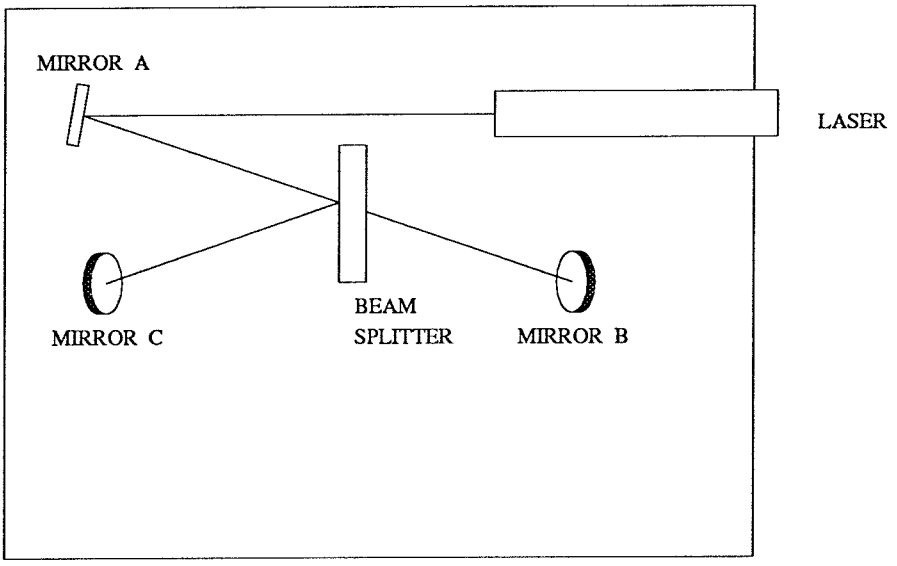


FIGURE 36. Diagram of interferometer configuration to measure the horizontal displacement components. Mirrors A, B and C and the beam splitter can be angularly adjusted about two axes with micrometers. Mirrors A and B are adjusted to bring the the beam that passes through the beam splitter to the specimen at a precise location and at the proper angle so that the first–order diffracted beam exits along the normal to the specimen surface. Then the beam splitter and mirror C are adjusted to bring the beam that reflects from the beam splitter to the correct location at the proper angle. (Changing the angle of the beam splitter slightly does not cause the beam that passes through the beam splitter to deviate significantly from its path.) The beam expander is then inserted into the beam, and mirror B or C adjusted to minimize the number of fringes. The interferometer configuration for measuring the vertical displacement component is adjusted in same way.

Once the unexpanded laser beam has been aligned, the collimating lens is inserted into the path and adjusted so that the beam passing through it is not deflected from its original route and so that the reflection from the lens travels back to the laser. This adjustment ensures that the lens is centered about the laser beam and that it is placed perpendicularly with respect to the path of the beam. The

beam expander, a microscope objective and pinhole arrangement, is then aligned. The microscope objective lens focuses the laser beam to a small diameter, and the pinhole filters out noise. In essence, the microscope objective acts as a Fourier transform on the laser light in spatial coordinates, and the pinhole filters all but the central peak of the beam in the spatial frequency domain, leaving a smooth beam that consists only of components of low spatial frequency. The distance from the collimating lens to the pinhole is adjusted so that the pinhole lies at a focal point of the lens; the light that leaves the lens then consists of plane waves. At this stage the overlapping beams diffracted from the specimen can be adjusted to produce a null field; that is, the number of fringes visible is minimized, and made zero if possible. If there is an initial pattern, it is hoped that it is negligible compared to the pattern obtained when the specimen is deformed. For steel, this initial pattern consists only of one or two random fringes. For polymers, however, the initial pattern can be several fringes.

2.4.4 The Grating

The specimen gratings are applied after the method of Post [62]; see also Post and Baracat, [63]. Figure 37 was taken from Post [62], and describes the process. A holographic plate is exposed to two crossed laser beams. The crossing of the beams produces an interference pattern of standing waves, which creates a periodic light–intensity variation across the surface of the holographic plate as shown in Figures 8 and 9. This period can be changed by changing the angle at which the laser beams cross. Again, the laser beams must originate from the same source and travel paths of similar length to create an interference pattern of high contrast. During the exposure, the structure of the emulsion on the holographic plate changes. Where the emulsion is exposed to light, the silver halide crystals suspended in the emulsion absorb energy and form metallic silver; this happens with more frequency in regions where the light is bright than where it is dark [37]. During the development process, the conversion to metallic silver is completed in those silver halide crystals that have begun to transform. When the plate is fixed, the unchanged silver halide crystals are removed, causing the emulsion to shrink in the regions where there is less metallic

silver. This shrinkage therefore follows the irradiance pattern of the interfering laser beams, creating a periodic surface–height variation that acts as a reflective diffraction grating. By coating this rippled surface with optical quality aluminum in a thin film, the reflectivity is enhanced and so is the diffraction efficiency. This rippled surface variation and the aluminum coating can be transferred to a specimen if the adherence between the aluminum film and the holographic emulsion is weak.

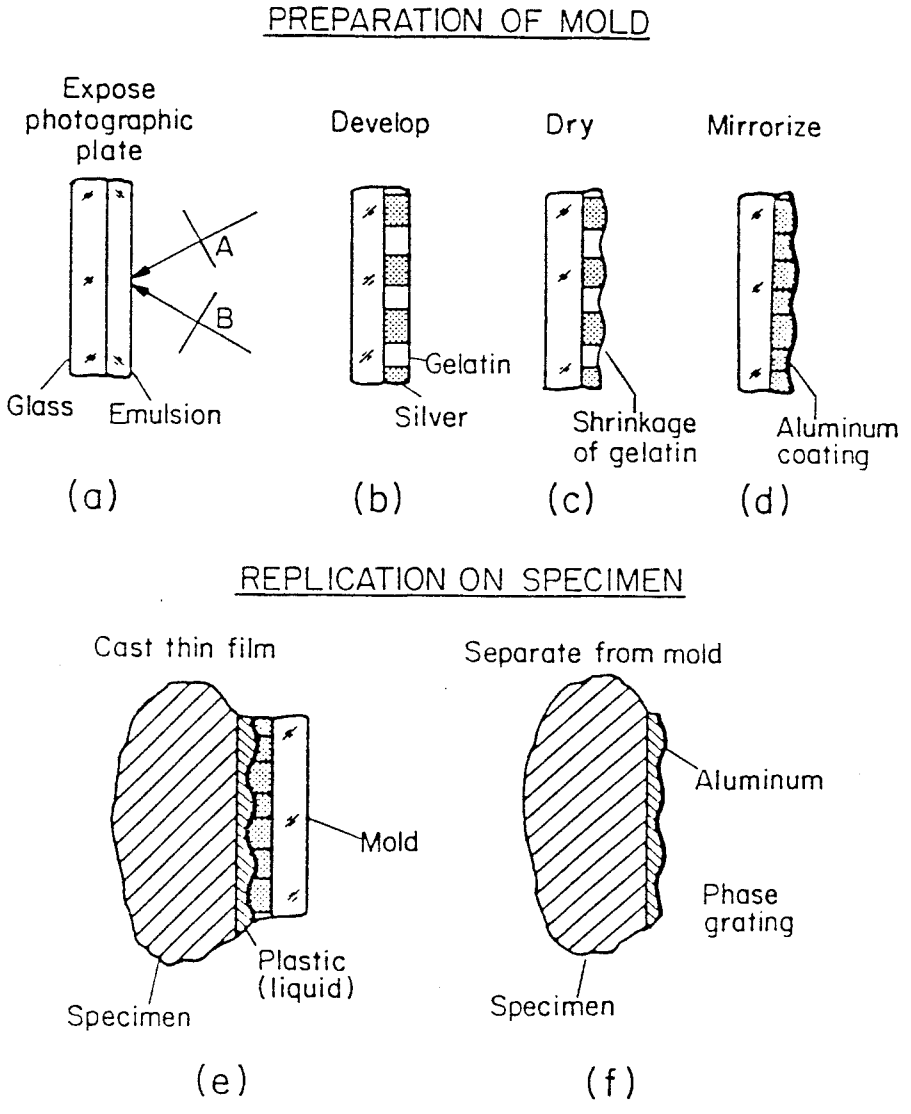


FIGURE 37. The process for preparing the specimen grating and transferring it to the specimen. (After Post [62].)

The details of the preparation of the specimen grating follow. Kodak Type 649

F spectroscopic holographic plates are exposed twice to two crossed laser beams emanating from a 15 mW Spectra–physics Helium-Neon laser, with the plates being rotated by 90 degrees between exposures to produce orthogonal sets of lines, making two orthogonal specimen gratings. The exposure time in each case is two seconds. The Kodak 649 F plates are sensitive to the 632.8 nm Helium-Neon laser wavelength and have an emulsion of 17 μms thick with a nominal sensitivity of 70 $\mu J/cm^2$ and a resolution of 2000 lines/mm. The plates are antihalation–backed to reduce reflection from the rear surface of the plates. The plates are developed as recommended by Kodak, 6 minutes in Kodak D-19 developer, 30 seconds in Kodak indicator stop bath, 3 minutes in Kodak rapid fixer, rinsed in water for 25 minutes and dip for 30 seconds in Kodak Photo-flo solution. After the plates have dried, they are dipped a second time in the Photo-flo solution for 30 seconds. When the plates are completely dry, they are coated with a thin coating of reflective aluminum in a vacuum coating chamber. This aluminum coating is on the order of 200 nm in thickness. The grating is transferred to the specimen using Photolastic PC-1 epoxy adhesive. (Photolastic Division, Measurements Group, Inc., P.O. Box 27777, Raleigh, North Carolina 27611, (919) 365-3800.) The double dip in the Photo-flo solution provides a very weak interface between the aluminum film and the emulsion on the holographic plate, thus facilitating the release of the aluminum from the photographic emulsion.

The gratings produced for this study have a wavelength of 1.667 μm or 600 lines/mm. The gratings are produced to a known wavelength by aligning the angle of the crossed laser beams to a commercially produced grating of wavelength 3.333 μm or 300 lines/mm in exactly the same way that the moire interferometer is aligned to the specimen grating. If there are no fringes visible over a 25.4 cm field of view, the wavelength of the laser interference is equal to twice that of the alignment grating to within 0.01 %. (The grating was obtained from Diffraction Products, Inc., P.O. Box 645, Woodstock, Illinois 60098. (815)338-6768.)

The efficiency of the diffraction gratings used for this study is about 10%. Loewen in [55] and Maystre, Neviere and Petit (also in [55]) discuss the diffraction efficiencies of sinusoidal and blazed diffraction gratings. For sinusoidal aluminum

coated gratings, Loewen predicts a maximum efficiency of nearly 90%. However, these efficiencies are obtained for gratings used in a specific orientation determined by the wavelength of the grating and the wavelength of light. Additionally, the efficiencies are given by Loewen for a single grating, whereas the moire interferometer requires a crossed grating, which introduces additional diffraction orders in which the light may be distributed. McPhedran, Derrick and Botten in Reference [55] discuss the theory of crossed gratings, but they are mainly concerned with predictions of certain anomalies, such as Wood's anomalies, observed in diffraction by gratings. Wood's anomalies are angles at which the diffracted light in a given order drops sharply, owing to a redistribution of energy into electromagnetic surface waves traveling through the grating itself.

2.4.5 Requirements and Limitations for the Collection Optics

The maximum strain that can be measured corresponds to the minimum fringe spacing that can be captured by the optical system and recorded on the film. Since the diffracted laser beams are still approximately collimated, the collection optics must have an aperture at least as large as the field to be viewed in order to record information from the entire field. The fringes observed in the experiment are the result of angular deviations of the diffracted beams from the original paths. Higher fringe densities, corresponding to higher displacement gradients (strains), are the result of relatively large angular deviations from the original specimen normal. Hence, in order to capture large–strain behavior, the angular aperture of the collection optics must be large.

The angular aperture is a measure of the largest angle a ray may attain without being lost by the optical system; it is inversely related to the f number of the optical system, so that a large angular aperture corresponds to a small f number. For a camera focussed infinitely far away, the f number of the camera lens is the ratio of the focal length of the lens divided by its aperture diameter.

The ability to capture fringes representing large strains is also limited by the resolution of the film used to record the image. For the Kodak Technical Pan film

used in the experiments, the minimum fringe wavelength that can be recorded on the film is about 5 microns [38]. However, the modulation transfer function of the film does drop off quickly as the fringe spacing decreases, so that the contrast of closely spaced fringes will be less than the contrast of widely spaced fringes.

The modulation transfer function is a measure of how well the contrast of the fringes is transferred through the optical apparatus to the recording film. Each optical component has a modulation transfer function, and when an optical device is cnstructed, the individual modulation transfer functions assemble in a complicated fashion to produce a modulation transfer function for the entire system. All modulation transfer functions tend toward zero as the fringe wavelength becomes smaller. Even a simple aperture without a lens shows this feature because diffraction from the edges of the aperture will cause the fringes to blur when the fringe wavelength becomes sufficiently small. The modulation transfer function for an optical system is determined using a known sinusoidal intensity distribution. The optical system produces an image of the known intensity distribution, and the contrast of the image is determined. The ratio of the contrast of the image to the contrast of the known distribution is the modulation transfer function for the spatial frequency of the known sinusoid. The contrast is defined by equation (2.4.2). Since film is a discrete recording medium where the discretization depends on the size of the silver halide grains in the photographic emulsion, the drop–off in the modulation transfer function at high frequencies is due to the limited sampling capability of the film.

The optical system can also be limited by aberrations in the lens, as well as by diffraction from the edges of any aperture through which the light must pass. These aberrations cause a blurring of the image, which will smear out the light–intensity distribution, destroying fringes of short wavelength. The Nikon camera lenses used in the experiment are optimized to be diffraction–limited; in other words, the most significant aberration in the lens is diffraction that is due to the aperture, not to defects in the lens. As a first estimate of the resolution limits of the optical recording system, only the effects of the angular aperture limit on the maximum observable diffraction angle will be considered. Diffraction from the edges of the aperture and

the effect of aberrations will be neglected †. The angular aperture will be assumed to be dominated by the lens aperture closest to the specimen, so that any ray passing through the lens aperture closest to the specimen will be assumed to pass through the rest of the lenses in the system and to fall on the film plane. In this case, the angular aperture is only a function of the distance between the specimen surface and the initial lens aperture and is assumed not to depend on the internal workings of the lens system.

If the camera aperture has a diameter A and the distance from the specimen to the aperture is L, as in Figure 38, the angular aperture, ϕ, is found from

$$\tan\phi = \frac{\mathrm{A}}{2\mathrm{L}} \tag{2.4.4}$$

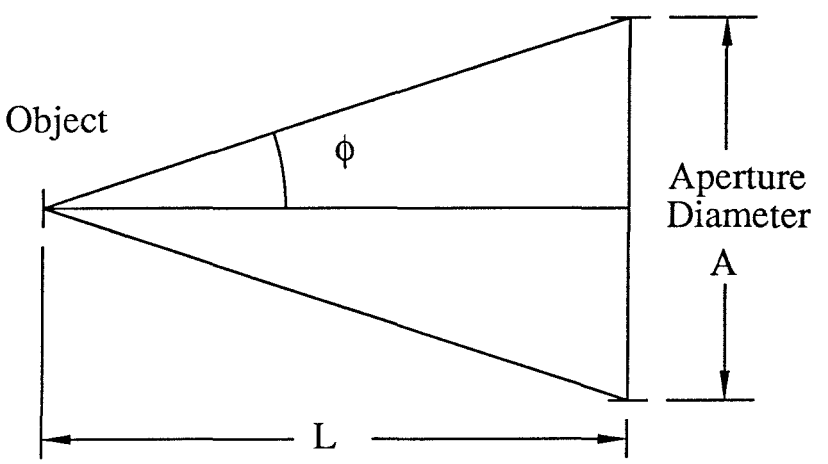


FIGURE 38. The angular aperture, ϕ, is a measure of the largest angle that a ray from the center of the field may achieve and still pass through the aperture of the lens closest to the specimen. L is the distance from the specimen to the aperture, A is the aperture diameter, and $\tan\phi = A/(2L)$.

† These effects are all included in the modulation transfer function for the system.

If two diffracted light rays (representing the normals of locally plane light waves) leave the specimen surface from the center of the field of view and just pass through the aperture of the lens, they will give rise to fringes with a wavelength δ, which depends on the angular aperture ϕ according to

$$\delta = \frac{\lambda}{2 \sin \phi} \tag{2.4.5}$$

where λ is the wavelength of light used. In planar deformation, the light waves are diffracted symmetrically about the normal to the specimen surface, so the two rays intersect the aperture at either end of a diameter of the aperture, which leads to the relation in equation (2.4.5) above. For a small angular aperture ϕ, $\tan \phi \approx \sin \phi$. This is a good approximation for the camera arrangement used in this study, which had an angular aperture of about 0.08 radian (about 5°). For small angular aperture ϕ, then, δ can be approximated by

$$\delta = \frac{\lambda \mathrm{L}}{\mathrm{A}} \tag{2.4.6}$$

For an optical system that has a magnification M, the fringes that arrive at the film plane have a wavelength $\mathrm{M}\delta$. The film is able to record fringes having wavelengths as small as $5\,\mu$m. Using $5\,\mu$m as the limit of the film, the fringes must then be magnified so that

$$5\,\mu\mathrm{m} < \mathrm{M}\delta \tag{2.4.7}$$

in order to record fringes of spacing δ at the specimen surface. From these equations, the required magnification can be related to the angular aperture, and thus to the camera aperture and specimen-to-aperture distance) as

$$\frac{5\,\mu\mathrm{m}}{\mathrm{M}} < \frac{\lambda \mathrm{L}}{\mathrm{A}} \tag{2.4.8}$$

For a Helium-Neon laser with $\lambda = 0.6328\,\mu$m this gives

$$\frac{7.90}{\mathrm{M}} < \frac{\mathrm{L}}{\mathrm{A}} \tag{2.4.9}$$

The required magnification M for a specimen-to-aperture distance L is shown in Figure 39 with an aperture A equal to 50 mm.

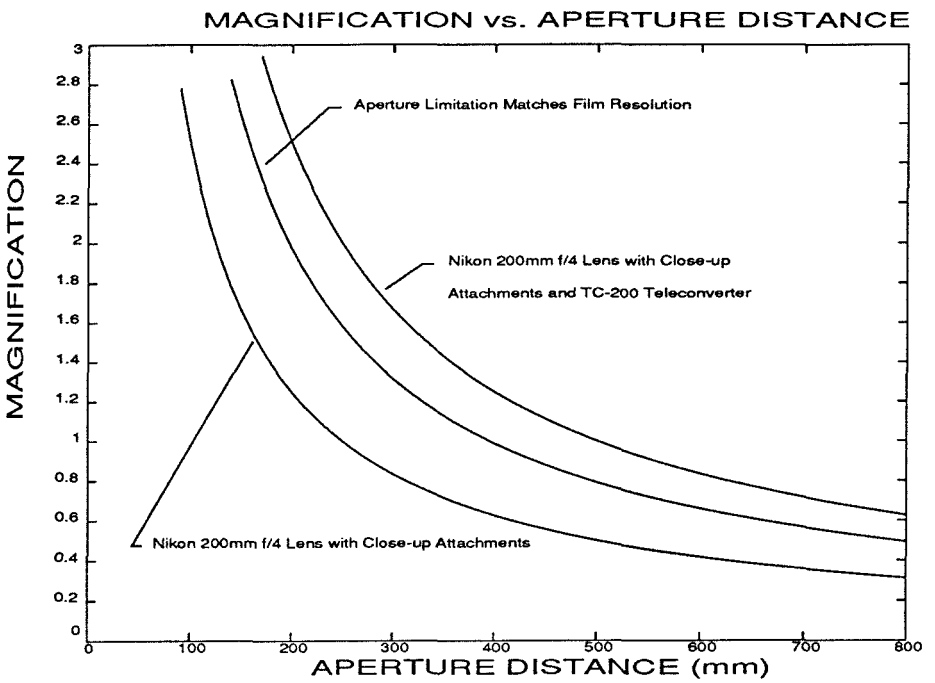


FIGURE 39. Required magnification to match film resolution to the aperture limitation. Performance of Nikon 200 mm f/4 with close-up attachments.

The Nikon Micro-Nikkor 200 mm f/4 Internal Focus lens used in the experiments has an aperture of 50 mm, and at its closest focus it has a magnification of 1/2 with a specimen-to-aperture distance of 500 mm. With the addition of the Nikon TC-200 Teleconverter, the magnification is doubled to 1 without changing the specimen-to-aperture distance. The lens alone does not provide enough magnification for the film to record the densest fringes that the lens can capture. However, with the addition of the teleconverter, the magnification is sufficient for the film to resolve the smallest fringes, but at the expense of imaging only one quarter of the area seen without the teleconverter. The minimum resolvable fringe wavelength can be improved by the use of one or more close-up attachment lenses, which decrease the specimen-to-aperture distance and improve the magnification by creating a virtual object farther away than the actual object, but proportionately larger than the actual object. These attachment lenses can be approximated by thin lens theory, creating a virtual object 500 mm away as seen by the primary 200 mm f/4 lens. Addition of such a lens relates the magnification M, the true specimen-to-camera

distance L, and the 500 mm distance observed by the primary lens through

$$M = \frac{500\,\text{mm}}{2L} \tag{2.4.10}$$

The magnification is doubled with the addition of the teleconverter. The required focal length f of the added close-up lens is given by

$$f = \frac{500\,\text{mm}}{2M - 1} = \frac{(L)500\,\text{mm}}{500\,\text{mm} - L} \tag{2.4.11}$$

The necessary magnification to match the film resolution with a 50 mm aperture is shown in Figure 39, along with the performance of the Nikon 200 mm f/4 lens with close-up attachments and teleconverter. It is necessary to be above the curve where the film resolution matches the aperture limitation to record all of the fringes that travel through the optical system.

The magnification can also be increased by use of an extension tube, which moves the film plane farther back from the lens. Use of such an extension also reduces the specimen-to-aperture distance. While this option is not normally recommended since the lens is optimized to focus on the film plane at a certain distance, the Nikon 200 mm f/4 lens is designed to give better performance at this off-design usage than other lenses. For this means of improving the magnification and angular aperture, the relation between the magnification and the specimen-to-aperture distance can be approximated by

$$M = \frac{167\,\text{mm}}{L - 167\,\text{mm}} \tag{2.4.12}$$

Equation (2.4.12) is a rough fit from the closest focus and magnification with no extension and the closest focus and magnification with a 52.5 mm extension. In fact, it follows the thin lens equation, assuming that when the 200 mm lens is adjusted to its closest focus, it can be approximated by a thin lens having a focal length of 167 mm. The 200 mm focal length applies when the lens is focussed to infinity. The approximate performance of the Nikon 200 mm f/4 with extensions is shown in Figure 40. The addition of the teleconverter would double the magnification without changing the specimen-to-aperture distance.

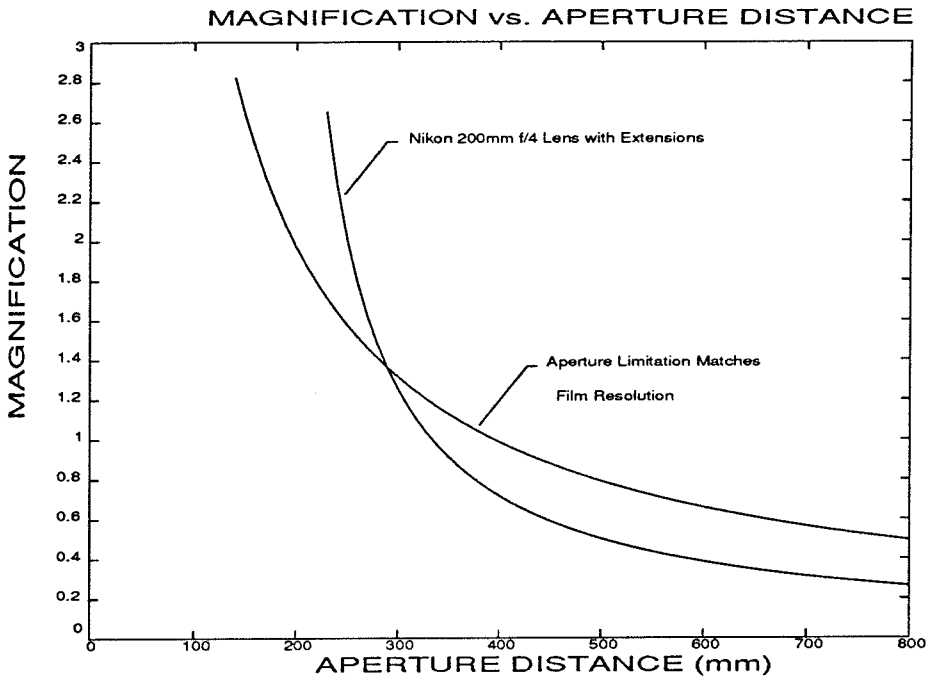


FIGURE 40. Required magnification to match film resolution to the aperture limitation. Performance of Nikon 200 mm f/4 with extensions.

2.4.6 Physical Limitation

The maximum angular aperture of the recording device is limited by the fact that it must not block the incoming laser beams, as illustrated in Figure 41. Additionally, if the aperture were to get in the way of the incoming beams, it would also be in a position to capture light from the reflected beams (the diffracted beams of order zero), and the diffracted beams of order two. As a reminder, the laser beams of interest are the diffracted beams of order one. For a given specimen–grating diameter of D, an aperture diameter of A, separation L between the grating and the aperture, and angle α between the normal to the grating surface and incoming laser beams, it can be seen from Figures 38 and 41 that

$$\tan\alpha = \frac{A + D}{2L} \tag{2.4.13}$$

$$\tan\phi = \frac{A}{2L} = \frac{A\tan\alpha}{A + D} \tag{2.4.14}$$

where ϕ is the maximum allowable angular aperture without blocking the incoming laser beams.The largest possible ϕ occurs when A becomes very large compared to

D, which gives $\phi = \alpha$. Of course, as A becomes large, L must also increase, for fixed α.

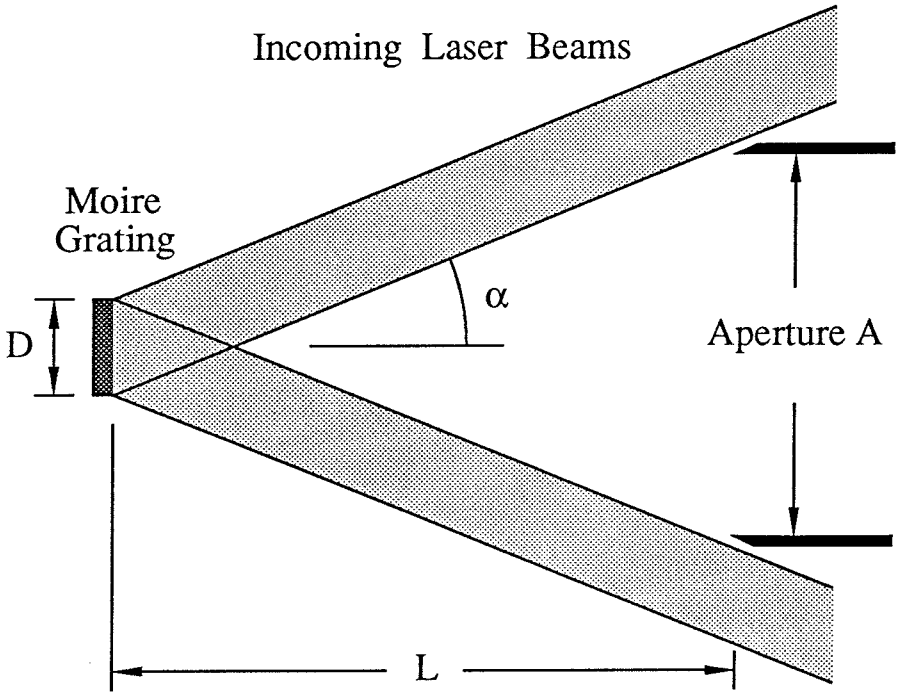


FIGURE 41. The aperture of the optical recording device may not be so large that it blocks the incoming laser beams. This constraint limits the angular aperture to be less than the angle α of the incoming laser beams. If the aperture lies just inside the incoming beams, $\tan\alpha = (\mathrm{D} + \mathrm{A})/(2\mathrm{L})$, while the angular aperture ϕ is determined from the relation $\tan\phi = \mathrm{A}/(2\mathrm{L})$.

2.4.7 Sensitivity

One measure of the interferometer performance is the sensitivity of the measurement. For this study, the sensitivity is defined as the displacement increment between adjacent black fringes. The sensitivity is a measure of the *smallest* displacement that can be detected by the interferometer. A Twyman-Green interferometer, which is used to measure out-of-plane displacements, has a fixed sensitivity of $\lambda/2$, where λ is the wavelength of light used. The sensitivity of geometric moire techniques depends on the spacing of the reference grating. The moire interferometer is similar to geometric moire in that its sensitivity can vary; the absolute lower limit on the sensitivity of the moire interferometer is $\lambda/2$.

The sensitivity of the moire interferometer depends on the wavelength of the undeformed diffraction grating, w_0. The displacement increment from one black fringe to an adjacent black fringe is $w_0/2$. The incoming laser beams are arranged in such a way that $\sin\alpha = \lambda/w_0$. From this relation it is seen that the minimum possible grating wavelength is $w_0 = \lambda$, and the absolute limit on the sensitivity is $\lambda/2$, identical to the sensitivity of the Twyman-Green interferometer. To obtain this sensitivity requires that $\alpha = \pi/2$, which means that the incoming laser beams are parallel to the surface. However, because the sensitivity is inversely proportional to the sine of the angle α, it is possible to make a reasonable approach to the limit with a significantly smaller value for α. At $\alpha = 3\pi/8$, the corresponding grating wavelength is $w_0 = \lambda/(0.924)$ and the sensitivity, $w_0/2$ is $\lambda/(1.848)$, which is only 8% greater than the limit of the sensitivity at an angle that is only 75% of the angle necessary to reach the limit.

2.4.8 Strain Measurement

A second measure of the interferometer performance is the maximum strain that can be determined. This measure corresponds to the *largest* displacement that can be recorded by the interferometer, and it is determined by the highest fringe density that can be captured and recorded on film.

The minimum fringe spacing can be converted to a measure of the maximum strain if the deformation is assumed to have a simple form. If it is assumed that the fringes are due entirely to stretching (no shear or rotation), the fringe spacing δ is found to be

$$\delta = \frac{w w_0}{2|w - w_0|} \tag{2.4.15}$$

where w_0 is the initial grating wavelength, and w is the current grating wavelength. The engineering strain, e, is defined by the stretch ratio $(w - w_0)/w_0$, which leads to the equation

$$\delta = \frac{w_0(1 + \mathrm{e})}{2|\mathrm{e}|} \tag{2.4.16}$$

$$\frac{|\mathrm{e}|}{1 + \mathrm{e}} = \frac{w_0}{2\delta} \tag{2.4.17}$$

The largest stretch ratio that can be measured corresponds to the minimum δ, which depends on ϕ, the angular aperture of the camera. The parameter $w_0/(2\delta)$ is the ratio of the sensitivity of the measurement to the minimum fringe spacing, and it is also proportional to the angular aperture of the system, since

$$\frac{w_0}{2\delta} = \frac{\sin\phi}{\sin\alpha} \tag{2.4.18}$$

Note that $\sin\phi/\sin\alpha$ varies between zero and unity. The variation of e with $w_0/(2\delta)$ is shown in Figure 42. Figure 43 shows a plot of the same data with the engineering strain, e, replaced by the "true strain," ϵ. The engineering strain, e, and the "true strain," ϵ, are related by

$$\epsilon = \ln(1 + \mathrm{e}) \tag{2.4.19}$$

which is the same equation used in the reduction of the uniaxial stress–strain data (See Section 3.3). Of course, for large stretch ratios the nonlinearity of the strain–displacement relation must be accounted for, but the observable stretch ratio is a parameter that will quickly show how well the interferometer can perform, and how close it is to its limits.

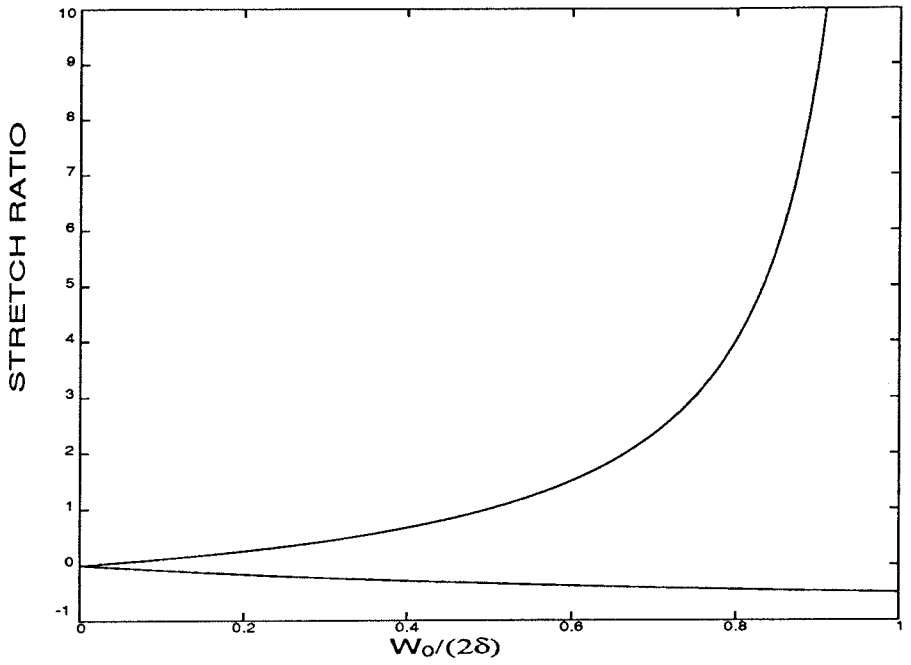


FIGURE 42. The stretch ratio $(w - w_0)/w_0$ as a function of fringe spacing δ. The parameter $w_0/(2\delta)$ is proportional to the angular aperture of the optical system. The fringe spacing δ and the undeformed grating wavelength w_0 are always positive quantities.

It is interesting that the observable stretch ratios do not behave symmetrically in extension and contraction, because the absolute value of the change in the grating wavelength, $w - w_0$, is used in equation (2.4.15). If the grating is stretched, so that the grating wavelength increases, in the limit the grating will eventually appear like a flat surface, and the diffracted beam will approach the zero–order, reflected beam. In this case, the strain can approach infinity while the angle of the diffracted ray is bounded. If the grating is compressed so that the wavelength decreases, the diffracted ray has a different kind of constraint. As the wavelength of the grating decreases, the diffracted ray may keep rotating until it is parallel to the grating surface. At that point, no solution to the diffraction equation exists (it implies that $\sin\phi > 1$), but the grating wavelength may continue to decrease.

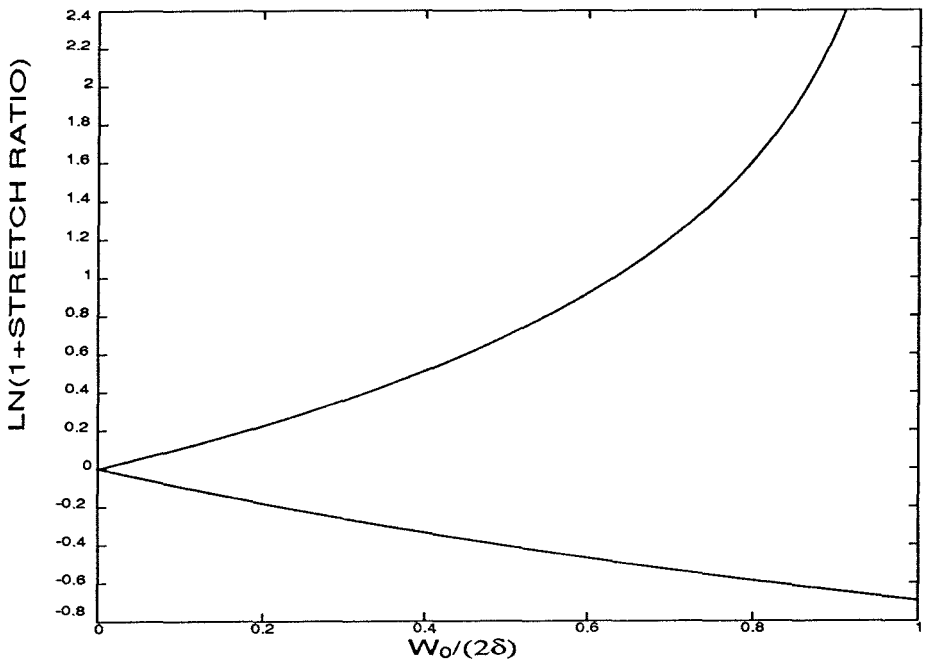


FIGURE 43. The function $\ln(w/w_0)$ as a function of fringe spacing δ. In a uniaxial test, the log of the 1 plus the stretch ratio is called the "true strain." The fringe spacing δ and the undeformed grating wavelength w_0 are always positive quantities.

2.4.9 Compromises

A compromise must be struck between the desire for high sensitivity in the measurement technique and the desire to measure large displacement gradients. There are a number of parameters that affect these quantities that must be determined. The wavelength of light is fixed at 0.6328 μm by the use of a Helium-Neon laser source. It is assumed that the aperture diameter A and the grating diameter D are equal; for this experiment it was chosen to be 50 mm. If the camera is capable of a magnification of 2, and the film can record fringes of 5 μm, the smallest fringe spacing on the specimen is 2.5 μm. This combination of parameters requires an angular aperture of $\phi = 0.127$ radian $= 7.3°$. Since A and D are equal, equations (2.4.13) and (2.4.14) indicate that $\tan\alpha = 2\tan\phi$, so α must be greater than or equal to 0.250 radian $= 14.3°$. This value for alpha yields a grating wavelength $w_0 = 2.56\,\mu$m. The parameter $w_0/(2\delta)$ is then equal to 0.51, and the corresponding

observable stretch ratios range from -0.34 to 1.04, or "true strains" between -0.36 and 0.72.

On the other hand, if the sensitivity is chosen to be about equal to the wavelength of light, so that the moire interferometer has half of the sensitivity of the Twyman-Green interferometer, $w_0 = 1.266\,\mu$m. Using this value for w_0 and $2.5\,\mu$m for δ yields measurable stretch ratios between -0.20 and 0.34, or "true strains" between -0.22 and 0.29.

The specimen gratings are made with a precise wavelength by aligning the intersection angle of the laser beams at the holographic plate with a commercially produced grating. A compromise was made by choosing a grating with a wavelength of $3.333\,\mu$m for alignment. The specimen grating produced has a wavelength half that of the alignment grating, so $w_0 = 1.667\,\mu$m. With $\delta = 2.5\,\mu$m, the measurable stretch ratios lie between -0.25 and 0.50, and the "true strains" lie between -0.29 and 0.41.

The camera magnification of 2 is chosen somewhat arbitrarily, although with a 50 mm field, a magnification of 2 would fill a 4 inch by 5 inch plate for a framing camera. If larger magnifications can be made, the lower limit on the fringe spacing grows smaller. Still assuming that A=D=50 mm and $\lambda = 0.6328\,\mu$m, and with $w_0 = 1.667\,\mu$m, the smallest δ that can be captured without blocking the incoming laser beams is $1.57\,\mu$m, assuming that the magnification is high enough to allow the film to record the fringes. This is the smallest fringe spacing that can be observed with the limitations imposed on A, D, λ and w_0. The stretch ratios are then between -0.35 and 1.11, and the "true strains" between -0.43 and 0.75.

In the experiment for this study, the highest magnification used was about 1.4, A=D=50 mm , L=320 mm. This gives $\tan\phi = 0.08$ so $\delta = 3.8\,\mu$m. The measurable stretch ratios are then predicted to lie between -0.18 and 0.28, and the "true strains" between -0.20 and 0.25.

2.4.10 Performance

With the case of $w_0 = 1.667\,\mu$m, the minimum possible fringe spacing is $\delta = w_0/2 = 0.8333\,\mu$m for an infinitely large aperture. With the limitation to a finite aperture of 50 mm, the minimum fringe spacing increases to $1.57\,\mu$m. Thus, the introduction of an aperture has increased the lower bound on the fringe spacing by a factor of only 1.9, assuming that the aperture is just inside the limit for blocking the beams. In the experiment for this study, with a predicted minimum fringe spacing of $3.8\,\mu$m, the minimum δ is 4.5, which equals the lower bound for an infinite aperture, and 2.4 times the lower bound for a 50 mm aperture. Should another experiment be performed, as hoped, the next experiment will be done as close to the 50 mm aperture limit as possible.

2.4.11 Future Improvements

The simplest and most effective improvement that can be made is to move the camera up to the closest point to the specimen allowed without blocking the incoming laser beams. This arrangement enables the camera to capture light from the specimen that is otherwise lost because of the angular deflections caused by both the in-plane and the out-of-plane displacements. The next improvement to make is to use a larger format camera, such as a 4 inch by 5 inch or 8 inch by 10 inch frame, rather than the 35 mm frame currently used, so that the photographs can be taken at a higher magnification and still include the entire field covered by the grating. Assuming that the test described in Chapter 4 will be repeated, these two improvements will be made. A third improvement in the camera system would be the addition of a lens with a larger aperture. A lens with a larger aperture will allow for the capture of more light than a lens with a smaller aperture, even if both are placed at the limit of blocking the incoming laser beams.

It is also possible to rotate the lasers so that the polarization of the incoming and diffracted laser beams is more nearly parallel, leading to higher fringe contrast. However, this action will change the polarization of the laser beams passing through the beam splitters, which will change the ratio of the transmitted light to the

reflected light. It should be possible to optimize this trade off for the highest fringe contrast.

Instead of transferring the grating from a holographic plate to the specimen, the grating could be produced on the specimen surface by means of a photoresist. A photoresist is basically a form of photographic emulsion that can be applied to the specimen and cured in place. Exposing the photoresist to crossed laser beams and developing the photoresist would produce a grating in the manner described in Section 2.4.4. The grating would then be coated with aluminum to improve the reflectivity. There are two advantages to this procedure. The first is that the quality of the grating is known immediately. Transferring a grating from a holographic plate is not routine, and the transfer may go awry. The success of the transfer is unknown while the epoxy is curing, and much time is wasted if the transfer does not succeed. A holographic plate is wasted as well, since the plates cannot be reused, and the plates are rather expensive. The second advantage to the photoresist is that since it is comparatively inexpensive, the exposure and development procedures may be changed to optimize the diffraction efficiency in the desired order. The disadvantages to the photoresist are that it requires shorter wavelengths and longer exposure times. Photoresists are produced mainly for printed circuit masks that are exposed using ultraviolet light. The photoresists are insensitive to the light from a Helium-Neon laser ($\lambda = 632.8\,\text{nm}$). In order to expose the photoresist, we would have to use the Ion–Argon laser tuned to its smallest wavelength. Additionally, more energy is required to expose the photoresist so that the exposure times must be longer, which makes vibrations more of a problem. Additionally, the specimen material is restricted in that it cannot be a material that is affected by the procedure used to develop the photoresist after exposure.

Initially, the moire interferometer was constructed under the assumption that we would study large areas, so the mirrors are all four inches in diameter, and the beam splitters are six inches in diameter. However, the experiments have all been conducted on areas of about two inches in diameter. If the area of interest is this small or smaller, it would be possible to reduce the size of the optical components, making the interferometer more compact. Further, the use of fiber optics, includ-

ing fiber optic beam splitters, would eliminate the need for all of the mirrors and beam splitters. Instead, four matching beam expander — collimator arrangements would be connected to a laser that could be optically switched between vertical or horizontal interferometers. In this arrangement, the key would be holding the optical components rigidly fixed with respect to one another. The mounting of the breadboard on the vibration–isolated optical table has very good vibration–damping properties, so that vibrations have not been much of a problem with the interferometer in its current incarnation. A moire interferometer using the fiber optic beam splitter and separate beam expanders and collimators is now commercially available.

2.5 Sources of Uncertainty or Error

Experimental uncertainty or error may be divided into two categories, systematic error and random error. Taylor [84] offers a good illustration of the two types of error, the example being the measurement of a repeatable event with a stopwatch. An example of random error is the uncertainty in the measurement introduced through the reaction time necessary for starting and stopping the watch at the beginning and end of the event. An example of systematic error is the uncertainty that would be introduced if the watch ran consistently 5% faster than it should.

The difference between the two types of uncertainty is that random error can be revealed through repetition of the measurement, while systematic error cannot. The error caused by starting and stopping the watch should vary between successive tests, and that variation in the measurement may be treated as a random quantity. If the measurement is repeated, the magnitude of the random error introduced into the measurement can be estimated through a statistical analysis of the repeated measurements. Thus, if a measurement can be repeated, a number of trials should produce a reliable estimate of the exact length of event (the mean value of the repeated measurements) as well as a good idea of the random uncertainty in the measurement. The systematic error in the example above cannot be discovered through a repetition of the measurement. Instead, the fast-running watch will produce a mean value and an estimate of the random error that are both 5% too large.

Systematic errors propagate through the measurement. They will not be revealed through repetition, but they may be removed from the measurement if they are discovered. For instance, if the stopwatch is checked against a reliable clock and found to be 5% fast, all of the measurements made with the stopwatch may be corrected. Systematic errors must be minimized at every turn because they cannot be removed or even estimated through a statistical procedure. Any sources of systematic error should be identified and removed, if possible.

2.5.1 Systematic Error

There are several sources of systematic error that may propagate through a measurement made using the moire interferometer. These sources of error arise in the process for making the diffraction grating and transferring the grating to the specimen, aligning the laser beams during the experiment, and reducing the data from the photographs taken during the experiment.

In making measurements using the Twyman–Green interferometer, one has the benefit of an effectively absolute length reference, the wavelength of the laser. (The mirrors and beam splitters are specified to be flat to much less than the wavelength, λ, of the light used, usually to $\lambda/10$ or better.) The measurements made using the moire interferometer have a less absolute reference, the initial wavelength of the diffraction grating. Uncertainty in the grating wavelength is due to the procedure for reproducing the diffraction gratings and aligning the interferometer. The diffraction gratings used in the experiment, with a nominal wavelength of 1.667 μm, are created on a holographic plate by exposure to the interference pattern between two crossed, collimated laser beams. These laser beams are aligned with a commercially produced master grating having a wavelength of 3.333 μm (300 grooves per millimeter). The grating was purchased from Diffraction Products, Inc. (P.O. Box 645, Woodstock, Illinois 60098, (815)338-6768). These reflective diffraction gratings, also called phase gratings, are periodic height variations with a period or wavelength on the order of the wavelength of light. The height variation is usually much smaller than the grating wavelength. The laser interference pattern has a wavelength of one half of the grating wavelength when it is aligned so that the first–order beam is diffracted along the normal to the grating plane. The exposed holographic plate also records a grating with one-half the wavelength of the master grating. This commercially produced master grating was scribed on an interferometrically controlled ruling engine. The interferometric control keeps the random error in the spacing of the grating profile below 0.0025 times the wavelength of a Helium–Neon laser source, or about 1.6 nanometers. This random error is non-cumulative, so the position of any groove will be within the random variation of its intended position. The error in the master grating will be assumed to be negligible compared to the errors

introduced in aligning the interferometer and reproducing the diffraction gratings.

Aligning the crossed laser beams to a grating with wavelength w_0 is intended to produce an interference pattern of wavelength $w_0/2$, which is then transferred to the holographic plate and used as the specimen grating in the experiment. Let the interference pattern have a wavelength of w_1, which is intended to be equal to $w_0/2$. The fringe spacing, δ, is related to the grating wavelength, w_0, and the wavelength of the interference pattern between the two laser beams, w_1, by

$$\delta = \frac{w_0 w_1}{[4w_1^2 + w_0^2 - 4w_1 w_0 \cos\gamma]^{\frac{1}{2}}} \tag{2.5.1}$$

where γ is the angle between the lines of the grating and the lines of the interference pattern. The maximum error in the wavelength occurs when the lines of the interference pattern are exactly parallel to the lines of the grating, so that $\gamma = 0$. The wavelength of the laser interference pattern is then given by

$$\begin{aligned} w_1 &= \frac{w_0}{2 - w_0/\delta} \qquad \text{for} \quad w_1 > w_0/2 \\ w_1 &= \frac{w_0}{2 + w_0/\delta} \qquad \text{for} \quad w_1 < w_0/2 \end{aligned} \tag{2.5.2}$$

If the fringe spacing δ were infinite, the wavelength of the interference pattern, and of the resulting specimen grating produced on the holographic plate, would be exactly $w_0/2$. The commercially produced master grating is 25 mm square. The laser beams are adjusted until there are no fringes visible, so that the minimum fringe spacing is 25 mm. The wavelength of the master grating is $w_0 = 3.333\mu$m, so $w_0/\delta \leq 1.333 \times 10^{-4}$. Thus, the specimen grating produced has a wavelength of $w_0/2 \pm 0.007\%$, or $w_1 = 1.667\mu\text{m} \pm 0.00001\mu\text{m}$.

A similar systematic error may be introduced when the experiment is performed. In the experiment, crossed laser beams are aligned to the grating that has been produced on a holographic plate and transferred to the specimen. This specimen grating has the wavelength w_1 described above. The laser interference pattern aligned to the specimen grating has a wavelength w_2. It is intended that the interference pattern have a wavelength of $w_1/2$ (actually $w_0/4$). This interference pattern is the basis for the measurements made using the moire interferometer.

as the deformation of the grating is revealed by comparing the deformed grating with the fixed positions of the laser beams. The fringes observed during the experiment represent displacement contours with the displacement between adjacent fringes differing by one interference pattern wavelength, w_2, which is intended to be exactly $w_0/4$. The specimen grating produced on the holographic plate is about 50 mm in diameter, but the initial alignment to this grating is not quite as good as the alignment to the commercial grating, because the quality of this holographically produced grating is not as high as that of the commercial grating. However, the minimum fringe spacing (δ_1) observed in this alignment is about 20 mm, and usually somewhat larger. Additionally, since the grating is more sensitive ($w_1 \approx w_0/2$), the error in this alignment is actually less than the error in the alignment to the commercial grating. The error in this alignment is estimated in the same way as shown in equation (2.5.2), so that w_2 is found as

$$w_2 = \frac{w_1}{2 - w_1/\delta_1} \qquad \text{for} \quad w_2 > w_1/2$$

$$w_2 = \frac{w_1}{2 + w_1/\delta_1} \qquad \text{for} \quad w_2 < w_1/2 \tag{2.5.3}$$

Making the substitutions $w_1 = 1.667\mu$m and $\delta_1 = 20$mm, the interference fringe pattern (displacement contour increment) is found to be $w_2 = w_1/2 \pm 0.004\%$. This error would approximately add to the error involved in the alignment to the master grating, so the laser interference pattern should have a wavelength of $w_2 = w_0/4 \pm 0.011\%$ or $w_2 = 0.8333\mu\text{m} \pm 0.00009\mu\text{m}$.

Actually, the error in aligning the specimen grating is not expected to be correlated with the error in aligning the master grating, so that one alignment error may yield a wavelength that is smaller than intended, and the other alignment yield a wavelength that is longer than intended. For uncorrelated errors of this sort, Taylor [84] indicates that the uncertainties should be added in quadrature, so that the uncertainty in w_2 is given by $\sqrt{(0.007\%)^2 + (0.004\%)^2} = 0.008\%$ in which case the laser interference pattern should have a wavelength of $w_2 = 0.8333\mu\text{m} \pm 0.00007\mu\text{m}$. In any case, the error in the wavelength of the laser interference pattern is found only in the fifth significant figure.

Another source of error that might be involved in the grating alignment procedure is the lack of collimation of the laser beams. The laser beams are collimated by passing them through a beam expander consisting of a microscope objective and spatial filter (pinhole) to a 50 mm diameter, cemented achromat lens. The collimation is checked by measuring the diameter of the beam close to the 50 mm lens, and also at a distance of several meters away from the lens. The distance between the lens and the beam expander is adjusted until the beam diameter remains constant. The most likely deviation in the laser beams would be a slightly spherical wave front, rather than a strictly plane wave front. However, spherical waves would simply produce a laser interference pattern with a wavelength that varies smoothly across the grating. The variation of the wavelength of that interference pattern must also be within the 0.008 % estimate above; otherwise, there would be fringes caused by the mismatch of the grating with the smoothly varying, laser–interference pattern resulting from the spherical wave fronts.

Error in the grating wavelength might be measured using an scanning electron microscope, or one of a class of tunneling electron microscopes. However, the grating wavelength would have to be measured over a distance of several wavelengths to provide a reliable measurement. The difficulty with the use of the scanning electron microscope is that the specimen size must be fairly small (about 1 cm), as the specimen must be placed in a vacuum chamber. Thus, the use of a scanning electron microscope to check the wavelength of the grating is not feasible, unless as a quality control measure, using a sample from one holographically produced grating as a check on a group of similar gratings. The tunneling electron microscope might be used to check the grating in place, as it can be operated in air and can be mounted in such a way that the specimen size is not restricted.

The uncertainty introduced in the alignment procedure described above is of low spatial frequency, either a constant discrepancy between the intended grating wavelength and the true grating wavelength, or a smoothly varying discrepancy across the grating surface that is due to a slightly spherical wave front rather than a true plane wave. Further systematic errors that have a higher spatial frequency may be introduced into the diffraction grating by effects that may be termed noise

in the crossed laser beams used to produce the grating on a holographic plate. This noise is generated by diffraction of the laser beams by the aperture of the lens, diffraction by dust particles in the beam, and by the effect of ghost reflections from the lens and holographic plate. These effects cannot be entirely eliminated, but they can be minimized. The diffraction pattern caused by the lens aperture is on the order of two millimeters around the edge. The effect of this diffraction pattern can be seen in the experiment photographs as a wavy modulation to the fringes at the edge of the field of view. The holographic plates are 400 mm by 425 mm, so larger optics could be used if the edge effects are a problem. Diffraction by dust particles can be reduced by keeping the room as clean as possible and avoiding air currents that could stir up the dust. The ghost reflections are minimized by using anti-reflection coatings on all lens surfaces, and through the antihalation backing on the KODAK holographic plates that reduces reflection from the back surface of the plate.

Along with stirring up the dust, air currents themselves can cause noise in the holographic recording. Any density variations in the air through which the laser beams travel will cause localized changes in the paths of the laser beams, causing a relative shift in the phases of the laser beams, creating fringes. This effect can be clearly seen if one holds one's hand under either of the two laser beams after they have been aligned so that no fringes are visible. The heat from the body causes air currents that change the phase of one beam relative to the other, causing fringes. This is similar to the effects observed using a Schlieren optical system in a wind tunnel. Thus, the holographic plates must be exposed in a quiet environment, after the heating or air conditioning has been turned off. However, since the exposure times for the holographic plates are about two seconds, any transient fluctuations in the air should be averaged out over the length of the exposure. On the other hand, the exposure times for the film recording the fringe patterns during the experiment are fairly short, about one one-hundredth of a second, so transient effects might be a problem, except for the fact that the fringe patterns caused by the deformation overwhelm any effects caused by small density variations in the air. Commercially manufactured holographic gratings are

produced in a vacuum to eliminate the problems of air currents and dust.

Similarly, the effects of vibrations of the optical components must also be minimized. This is primarily accomplished by mounting all of the equipment on a Newport Corporation optical table, which is pneumatically isolated from the floor. The optical table itself is also highly damped. For the long exposure times necessary for the holographic plate, any vibrations are expected to be damped out relatively quickly. For the short exposure times used in photographing the fringe patterns, the photographs are taken only when the fringe patterns are still. (Clicking the shutter on the camera does cause significant vibration of the fringe patterns, which can be seen after the picture has been taken, but the exposure of the film has been completed before the vibrations emanating from the camera reach the specimen or other optical components.)

Actually, the most significant source of error in the alignment of the optics would be a drift in the position of one or more components before the process of producing the grating on the holographic plate has been completed. This error can be minimized only by ensuring that all of the optical components are held firmly in place after the alignment procedure has been completed.

There are other sources of error in the chemical processes necessary for developing the holographic plate and transferring the grating to the specimen. The development process removes silver halide crystals from the emulsion on the glass plate; these crystals are taken from those regions where too few photons have arrived to convert the silver halide into metallic silver. The local shrinkage of the emulsion as these unconverted crystals are removed causes the height variation of the phase grating. If the shrinkage is more widespread, it may change the wavelength of the phase grating over a larger area, altering the wavelength of the laser interference pattern used as a reference in the experiment. The emulsion of the KODAK Type 649F holographic plates lies on a glass substrate one millimeter thick. This substrate must be sufficiently rigid for use in holographic interferometry, in which the reconstructed object viewed through the holographic plate is compared interferometrically with the real deformed object. The epoxy used to transfer the

phase–grating profile to the specimen may also be subject to shrinking or swelling while it cures. However, the surface of the specimen is lapped flat, and the holographic plate is pressed against the specimen by another steel plate that has been lapped flat. The epoxy used in the experiment is designed for use in photoelasticity experiments, in which a layer of photoelastic material is bonded to the structure to be tested. Thus, the epoxy is designed to cure in thin layers under large plates and to hold the relatively soft, photoelastic material firmly in place on the surface of the structure without appreciable swelling or shrinking. In addition, this epoxy is intended to be used with structures undergoing relatively large deformations. The epoxy is allowed to cure for about one week, and then the holographic plate is peeled away from the surface immediately before the test. The use of epoxy to transfer the grating also sometimes allows for air bubbles to be trapped under the grating, even though the epoxy is placed in a vacuum before it is poured onto the surface of the specimen. If the epoxy does shrink or swell during curing, the regions surrounding the air bubbles will produce a local concentration of that shrinking or swelling. However, the air bubbles may also provide a means for eliminating another source of systematic error, which will be discussed below. Errors that are due to the changes in the emulsion or epoxy might be revealed through electron microscopy, along with any deviation in the grating wavelength that is due to optical misalignment or drift, or air currents or vibrations.

A different type of error that may be introduced during the manufacture of the grating is nonorthogonality of the crossed gratings. In order to measure two independent in-plane displacement components, the surface must have gratings with ridges running in two directions. It is intended that the crossed gratings be orthogonal, so that the in-plane displacements measured are also orthogonal. The two gratings are produced on the holographic plate by exposing the plate to the laser interference pattern twice, with a 90° rotation between exposures. The rotation is controlled by mounting the plate holder on a Newport Corporation rotation stage, which is calibrated in 5 minute increments, so one exposure is made at $0° \pm 5'$ and the other exposure made at $90° \pm 5'$. Again, the error in each angle is uncorrelated, so that the cumulative error in the angle between the ridges of the two gratings

should be less than about 7 minutes of arc.

Although it is expected that the crossed gratings on the holographic plate are orthogonal within 7 minutes, the lines of the gratings may not coincide with the intended axes of the specimen, in this study the line defined by the crack. The lines of the grating are aligned with the edges of the holographic plate, by placing a piece of thread obliquely across both collimated laser beams and adjusting its position until there is only one shadow visible at the plane of the holographic plate, passing through the center of the spot illuminated by the intersecting laser beams. At this point, the shadow forms a line perpendicular to lines of the laser interference pattern. The holographic plate holder is mounted on a rotation stage attached to a lab jack, so that the plate holder may be adjusted up and down and rotated. A plate with a line drawn across the middle is placed in the plate holder, and the lab and rotation stage adjusted until the line on the plate coincides with the shadow of the thread. In this way the lines of the grating are aligned with the edge of the holographic plate to about ± 15 minutes of arc. The edge of the fracture specimen and the edge of the holographic plate are aligned during the transfer process through the use of a jig. This alignment is good to about ± 30 minutes of arc. Assuming that these two alignment uncertainties are uncorrelated, the uncertainty of the alignment between the fracture specimen and the lines of the grating is about ± 35 minutes of arc.

So far, the discussion about the sources of systematic error has focussed primarily on those errors that occur before the experiment has begun. Other sources of error arise from the adjustment of the interferometer during the experiment, from blowing up the photographs, and from the data–reduction procedure.

The interferometer was adjusted during the experiment in an effort to remove the effects of rigid rotations of the specimen as it was loaded. The criterion for the adjustment was symmetry of the fringe pattern about the line of the crack. The two micrometer screws on each of the mirror mounts used in the moire interferometer described in Section 2.4.2 allow for approximately orthogonal adjustment of the path of the laser beams. One micrometer adjusts the angle at which the laser beams

intersect; this micrometer was left unchanged during the experiment. The second micrometer adjusts the orientation of the plane in which the laser beams cross,; in effect, it rotates the planes of the laser interference pattern. This micrometer was adjusted in an effort to negate the rigid rotation of the specimen, by adjusting the fringe patterns until they were symmetric about the crack line.

The adjustment to remove rigid rotations from the fringe patterns is probably the least well–calculated source of error in the experiment. The amount of each adjustment was not recorded during the experiment, and the adjustment was made independently for the two interferometers measuring the vertical and horizontal displacements. The assumption that the adjustment of the path of one of the laser beams does not affect the angle at which the laser beams cross is valid for small adjustments, but that assumption may not be valid if a large adjustment is necessary. In future, it may be better to avoid adjusting the interferometer, and to remove the effects of small rigid rotations in the data–reduction procedure. This strategy was adopted in reducing the data from the experiment used to determine Poisson's ratio.

The effects of rigid rotation can be removed from the fringe patterns if some assumptions are made regarding the symmetry of the displacement fields. The displacement component parallel to the crack, u_1, is assumed to be symmetric about the crack line (the x_1 axis), so that $u_1(x_1, x_2) = u_1(x_1, -x_2)$. The displacement component perpendicular to the crack, u_2, is assumed to be antisymmetric about the crack line, so that $u_2(x_1, x_2) = -u_2(x_1, -x_2)$. Making these assumptions, the effect of a small rigid rotation can be measured and removed during the data reduction. For a small rigid rotation of γ about the x_3 axis, the displacements that are due to the rotation are approximately

$$\begin{aligned} u_1 &= -\gamma x_2 \\ u_2 &= \gamma x_1 \end{aligned} \tag{2.5.4}$$

Thus, the u_1 displacements may be corrected by subtracting a linear function of x_2 corresponding to the displacements that are due to the rotation, which may be measured from the anti-symmetric u_1 displacement along the positive and negative x_2 axis. The u_2 displacements may be corrected by subtracting a linear function

of x_1 corresponding to the displacements that are due to the rotation, which are evident in symmetric deformation above and below the crack along the negative x_1 axis (the crack itself). These are the locations where the fringes are nearly perpendicular to the axes, making the comparison easier.

Any rotational misalignment of the moire interferometer during the experiment may also be corrected in this way, although it was intended that the adjustment of the interferometer produce a symmetric fringe pattern about the x_1 axis. Small deviations from exactly symmetric fringe patterns are due to slight rigid rotations away from the optimum alignment, which may be removed using the symmetry assumptions above.

A further source of systematic error lies in blowing up the photographs of the fringe patterns so that the displacement data may be extracted. While it is intended that each picture be some exact magnification of the actual physical specimen, a slight drift in the positioning of the enlarger may change the magnification slightly between frames. Errors in the magnification will cause errors in the measured positions of the fringes that are linear in the in-plane dimensions. A 1% error in the magnification will lead to a 1% error in the measured distance between adjacent fringes.

This error may be eliminated through the use of the spots on the specimen surface caused by air bubbles in the epoxy, or other sources of noise that are fixed on the specimen surface, such as scratches. Points on the specimen ahead of the crack do not move appreciably under loading, so measurement of the relative positions of two or more points should provide a good control of the magnification of the photograph. Points behind the crack tip may move by relatively large amounts, because of large rotations behind the crack.

Further systematic errors may arise in the use of the digitizing tablet to measure the positions of the fringes. The center of curvature of the notch tip is used as the origin of the coordinate frame in which the locations of the fringe centers are measured. An overlay is used to orient axes with the line of the crack and the origin at the notch tip. The error in locating the position of the notch tip will be constant

for each photograph, but will vary between photographs. Error in this location will provide a constant offset to the locations of the fringes that depends on the magnification. A conservative estimate is that the error in positioning the origin of the axes on the overlay is within one half of one millimeter on the digitizing tablet. For a photograph with a magnification of five, which is the lowest magnification photograph used in the data reduction, this corresponds to an error of less than one tenth of one millimeter, or one hundredth of the plate thickness, on the specimen itself. The notch has a tip radius of 0.15 millimeters. The error in positioning the origin of the overlay axes should decrease as the magnification increases, as the tip is easier to see, and a small error in the position on the tablet is made even smaller on the object by the effect of the magnification.

A misalignment of the axes of the overlay grid with the line of the crack will cause an error in the position of the fringes which is approximately linear with the radius from the origin, much like the displacements that are due to rotation discussed above. This misalignment is also constant for each photograph, but may vary between photographs. This misalignment error should be less than about one degree of arc for a photograph with a magnification of five, but this could increase as the magnification is increased, as less of the crack is visible for aligning the axes of the overlay.

The errors in aligning the overlay might also be reduced or at least made identical between photographs by using the locations of air bubbles or other marks as alignment guides. In fact, three noncollinear marks on the specimen surface ahead of the crack could be measured once with respect to axes having their origin at the notch tip, and those three points used as benchmarks for each successive photograph. This would eliminate difficulties with repeating the magnification and alignment of the overlay. This strategy was also followed in reducing the data from the Poisson's ratio experiment.

2.5.2 Random Error

In contrast to sources of systematic error, there are fewer sources of random error that need to be addressed. As mentioned, random errors are revealed through multiple measurements of the same repeated event. There are few repeated or repeatable events in the alignment, manufacture or transfer of the grating to the specimen, or in aligning the interferometer before or during the experiment. Such random sources of error would be revealed only by repeating the entire experimental procedure, and would yield an estimate of the distribution of the systematic errors described above. For example, an estimate of the expected magnitude of the rigid rotation could be determined from several repetitions of the experiment, but having that estimate would neither eliminate the rotation nor provide a means for subtracting the effects from the data. To really address the sources of random error, the entire experiment would have to be repeated, perhaps many times.

One source of random uncertainty is the error in locating the centers of the fringes from the photographs using the digitizing tablet. This error could be estimated by repeated measurement of the location of the fringes, or perhaps by fitting a smooth curve through data taken along a line and estimating the magnitude of the average offset of the data points from the curve. The digitizing tablet has a resolution of 0.0254 mm (0.001 inch) between points on the pad. Estimation of the location of the center of one black fringe should certainly be less than the space between adjacent black fringes. Should this not be possible, a further magnification of the photograph should be made. A reasonable estimate of the minimum fringe spacing that can be seen and digitized without magnification is about 0.5 mm between black fringes. With the aid of a magnifying lens above the digitizing tablet, this could be improved to about 0.05 mm, perhaps. For relatively widely spaced fringes, estimate of the location of the center of fringe should be within about one tenth of the distance between adjacent fringes. For closely spaced fringes, there will be an absolute distance, determined by the digitizing hardware and by the visibility, which determines the minimum resolvable fringe spacing. The magnification of the photographs should be such that the hardware–determined minimum is less than the estimate for easily visible fringes of one tenth of the fringe spacing.

In the fit of the data from the Poisson's ratio experiment, the minimization of the fit function provided an estimate of random error in the position of the fringes of about three tenths of the distance between adjacent fringes. This is much larger than the estimate above. This discrepancy is much more likely caused by a poor choice of the fitting function as compared to the actual displacement fields than a true measure of the scatter in digitizing the positions of the fringes.

There may be a type of random error introduced into the fringe patterns by vibrations. Vibrations resulting in rigid translations of the body do not affect the shape of the fringe patterns, although they may change the phase of the diffracted laser beams enough to change a light fringe into a dark one, and vice versa. This change does not affect the data reduction, however, as the moire interferometric measurement is made only relative to other points in the field of view, unless the displacement of one or more points in the field is known from some other method.

Vibrations that manifest themselves as rotations of the body produce changes in the fringe pattern that will introduce error in the fringe patterns. If the photograph of the fringe pattern is taken in a sufficiently short time so that the fringes are effectively frozen, this effect may be taken care of in the data–reduction procedure outlined above.

CHAPTER 3

Material Properties

3.1 Introduction

A key element in making a close comparison between a finite–element model and an experiment is a proper representation of the constitutive behavior of the material for use in the numerical model. As a result, some care was taken to measure the material properties of the steel used in the experiment, in both a uniaxial test and a test to measure Poisson's ratio. One possible method of using the data from the measurement would be to provide a table of measured stress and strain to the finite–element code, so that the numerical model would exactly follow the measured constitutive behavior. At the same time, however, it was intended that a comparison would be made between the experimental measurements of the crack tip deformation and the analytical crack tip deformation field derived by Hutchinson [31] and by Rice and Rosengren [69] for plastic deformation, called the HRR field. The HRR field is intended to be applicable for materials that follow a power law relating the stress and strain when plasticity is present. With this comparison in mind, the measured elastic–plastic constitutive behavior of the steel used in the experiments is fit to a power–law representation in the nonlinear regime. Comparison will also be made with the linearly elastic deformation field derived by Williams [90], governed by the stress–intensity factor, K, called the K field. Both the HRR field and the K field are singular near the crack tip.

Careful measurement of the constitutive behavior of the material employed in this study was also intended to provide a basis for a comparison of the results of the current study with previous experimental and finite–element investigations performed at GALCIT, such as the work by Zehnder [91], Zehnder and Rosakis [92],

Narasimhan and Rosakis [50,51], and Narasimhan, Rosakis and Moran [52]. As this work became available for comparison, the choice of material and the three–point–bend fracture specimen geometry for the present study was redirected. Initially, it was intended that the present experiments performed using aluminum would be compared to finite–element results through a suitable scaling of the material properties such as Young's modulus, Poisson's ratio, yield stress and hardening exponent. The alternative was to proceed with a finite–element model which employed a constitutive relation based on measured quantities. The latter path was eventually chosen. Dr. Narasimhan and Dr. Rosakis kindly provided the numerical code used in the present study, a good description of their use of the model may be found in [52]. With Dr. Narasimhan's assistance, the model was altered somewhat in order to match the present experiment geometry and constitutive behavior as closely as possible.

In the first section to follow, a brief description of the incremental plasticity relation used in the finite–element code will be given. This description is provided as background for the discussion of the finite–element model in Section 4.3. It should be emphasized that the constitutive model is restricted to linearized small strains; this is one restriction that will need to be overcome in order to study the large deformations at the tip of the crack more thoroughly. It is also important to emphasize that the code employs an incremental plasticity law, which is generally thought to be a better representation of the behavior of real materials than deformation plasticity. In contrast, a deformation plasticity theory is used to derive both the HRR field and the J–integral, which is used to determine the strength of the singularity of the HRR field. A comparison will be made between the incremental–theory power–law formulation of the finite–element model and the deformation–theory power law used to calculate the HRR field quantities. Finally, specialization of the constitutive law used in the code to a uniaxial test indicates which material properties must be measured for input to the numerical code.

The measurement of these material properties through the uniaxial test are described in Section 3.3. The specimen geometry and testing procedure are described, as well as the procedure for fitting the plastic stress and strain behavior

to the power law used in the numerical code. Some comparison will be made between the measured constitutive behavior of the present study with other materials some from previous investigations, such as the work of Zehnder [91]. The results of these comparisons indicate that the yield stress is a good parameter to normalize the stress and strain behavior, at least for the small sample of 4340 steel used in the comparison, all of which was subject to similar heat treatment. An important deviation between the fit power law and the experimental data may be seen in the region near the yield stress, the effect of Lüder's bands passing through the test section. Lüder's bands result from the sudden release of pinned dislocations when the shear stress reaches some critical value, and this feature is not captured by the power law. The final topic of this section is an attempt to extend the uniaxial test data to large deformations, which could be used with a finite–strain finite–element code; measurements of the stress and strain were made through photographs. These data are taken in the region beyond the onset of necking, where the assumption of a uniform stress and strain state is no longer valid, and a correction for stress triaxiality was needed. An empirical correction taken from Bridgman [12] was used. His correction was derived specifically for specimens of circular geometry, rather than for the rectangular geometry used in these tests; the empirical correction he gives has been used as a first approximation to correcting the effects of triaxiality in the rectangular specimen.

Section 3.4 presents a description of tests of the surface hardness of various specimens. The surface–hardness comparison is based on the results shown in Section 3.3, which indicate that normalization of the measured uniaxial properties of different specimens by the yield stress would provide the best basis for comparison. The surface–hardness measurement is intended to provide a simple, quick estimate of the yield stress of a material; the measurement relies on a relation between the yield stress of a material and the hardness of the specimen measured with an indenter. In particular, these measurements were intended to ensure compatability between the three–point–bend specimen of the current study and several other three–point–bend specimens used to determine the extent of crack tunneling in the interior of the body.

In making use of the finite–element model to simulate the three–point–bend specimen, one phenomenon that needed to be addressed was tunneling of the crack in the interior of the body, before crack growth was visible at the surface. Such tunneling could change the specimen geometry radically, if the tunnel were deep, making the numerical model a poor simulation of the experiment. In order to estimate the extent of the tunneling, a multispecimen approach was used. The tunneling specimens were made from 4340 steel left over from previous investigations. In order to attempt to validate the comparison between the steel used in the three–point–bend experiment and the steel used in the tunneling tests, the hardness of several different specimens was measured for comparison with the three–point–bend specimen used in the moire interferometric study. The tunneling tests themselves will be described in Chapter 4. It must be recognized, however, that the tunnel tests are an investigation of the material response at the far end of the stress–strain diagram, where material failure is occurring, so that the comparison based on yield stress does not produce a clear answer as to the comparability of the materials.

In the final section of this chapter, a test to measure the Poisson's ratio of the material using the moire interferometer is described. The specimen was a section of a second three–point–bend specimen of the same material as the specimen used in the interferometric test. The second three–point–bend specimen had been used in a preliminary test. The Poisson's ratio test was performed in compression and the axial and transverse displacements measured with the moire interferometer. The measured displacements were fit to an assumed function of the applied load to determine Poisson's ratio. The compression loading caused considerable bending stresses through the thickness of the specimen. Poisson's ratio was estimated to be about 0.28, as compared to the value of 0.3 used by previous investigators. This is a difference of about 7%, and would introduce an error of about 7% in an out-of-plane displacement comparison, assuming plane stress. The error in the in-plane displacement comparison may be estimated from the K field of Williams [90], which indicates that the error in a comparison of calculated in-plane displacments should vary between 10 and 30%.

Much of the data in this chapter was generated to make comparisons between

various tests. Part of the need for such comparisons is a lack of foresight as to how much steel would be needed to make these tests. No allowance was made for the specimens needed for the tunneling tests, nor for the tests to measure Poisson's ratio.

3.2 Elastic and Plastic Constitutive Relations

In this section, a brief discussion of the elastic and plastic constitutive model is presented. The plastic constitutive relation requires a yield condition to determine when plastic flow occurs, and a rule relating the size and direction of the plastic strain increment to an applied increment of stress. Much of the discussion is pertinent to the HRR field and to the finite–element model discussed in Chapter 4, but the specialization of the plasticity model to a uniaxial test is required for the discussion of the results of the uniaxial test data which follow in Section 3.3.

This constitutive formulation is adapted from class notes of A.J. Rosakis, with further reference to the books by Hill [28] and Kachanov [34], and to the paper by Narasimhan and Rosakis [50].

3.2.1 Preliminaries

The deformation is restricted to admit only linearized (small) strains, so that no distinction is made between the position of a material particle in the undeformed (reference) configuration as compared to the position of that particle in a deformed configuration. In this case the components of the strain tensor at the position $\mathbf{x}$, $\epsilon_{ij}(\mathbf{x},t)$, with respect to a fixed Cartesian orthonormal coordinate frame ($i,j = 1,2,3$), are related to the components of the displacement vector, $u_i(\mathbf{x},t)$, as

$$\epsilon_{ij}(\mathbf{x},t) = \frac{1}{2}\left(\frac{\partial u_i(\mathbf{x},t)}{\partial x_j} + \frac{\partial u_j(\mathbf{x},t)}{\partial x_i}\right) \tag{3.2.1}$$

for any time t.

The stresses are required to satisfy the quasi-static equations of equilibrium in the absence of body forces. In components, using the same fixed Cartesian orthonormal coordinate frame,

$$\frac{\partial \sigma_{ij}(\mathbf{x},t)}{\partial x_j} = 0 \tag{3.2.2}$$

again for any time t.

The constitutive law must then relate the stress state of the body, $\boldsymbol{\sigma}(\mathbf{x},t)$, to the strain state, $\boldsymbol{\epsilon}(\mathbf{x},t)$, at the current time by including some way to quantify effect of the accumulated plasticity incurred over the history of the deformation.

3.2.2 Strain Decomposition

For a given current strain state $\boldsymbol{\epsilon}(\mathbf{x},t)$, it is assumed that the strain can be decomposed into an elastic part $\boldsymbol{\epsilon}^e(\mathbf{x},t)$ and a plastic part $\boldsymbol{\epsilon}^p(\mathbf{x},t)$.

$$\boldsymbol{\epsilon}(\mathbf{x},t) = \boldsymbol{\epsilon}^e(\mathbf{x},t) + \boldsymbol{\epsilon}^p(\mathbf{x},t) \tag{3.2.3}$$

The elastic strain $\boldsymbol{\epsilon}^e(\mathbf{x},t)$ is related to the stress state $\boldsymbol{\sigma}(\mathbf{x},t)$ by the usual linearly elastic constitutive law for homogeneous, isotropic materials. This relation is valid in both the elastic regime and the plastic regime, and includes the possibility of elastic unloading from a plastically strained state. In terms of components, the elastic constitutive law is written as

$$\epsilon^e_{ij}(\mathbf{x},t) = \frac{(1+\nu)}{\mathrm{E}}\sigma_{ij}(\mathbf{x},t) - \frac{\nu}{\mathrm{E}}\sigma_{kk}(\mathbf{x},t)\delta_{ij} \tag{3.2.4}$$

where δ_{ij} is the Kroneker delta, and repeated indices kk imply summation over $k = 1,2,3$. E is Young's modulus, and ν is Poisson's ratio. This constitutive law may also be written as

$$\sigma_{ij}(\mathbf{x},t) = \frac{\mathrm{E}}{(1+\nu)}\epsilon^e_{ij}(\mathbf{x},t) + \frac{\nu\mathrm{E}}{(1+\nu)(1-2\nu)}\epsilon^e_{kk}(\mathbf{x},t)\delta_{ij} \tag{3.2.5}$$

In more compact notation, equation (3.2.5) may be written in general form as

$$\sigma_{ij}(\mathbf{x},t) = C_{ijkl}\epsilon^e_{kl}(\mathbf{x},t) \tag{3.2.6}$$

where C_{ijkl} is a constant, positive–definite, elastic, compliance four-tensor for homogeneous, isotropic materials. From equation (3.2.5), C_{ijkl} may be found as

$$\begin{aligned} C_{ijkl} &= \frac{\mathrm{E}}{(1+\nu)}\delta_{ik}\delta_{jl} + \frac{\nu\mathrm{E}}{(1+\nu)(1-2\nu)}\delta_{kl}\delta_{ij} \\ &= 2\mu\delta_{ik}\delta_{jl} + \lambda\delta_{kl}\delta_{ij} \end{aligned} \tag{3.2.7}$$

where μ and λ are the Lamé constants of the material.

3.2.3 Assumptions Regarding the Plastic Strain

The plastic strain is assumed to be independent of the hydrostatic pressure, so that the material is plastically incompressible. This means that the plastic strain $\epsilon^p(\mathbf{x},t)$ is only a function of the deviatoric stress $\mathbf{S}(\mathbf{x},t)$, where $\mathbf{S}(\mathbf{x},t)$ is the current stress state minus the hydrostatic pressure. In components,

$$S_{ij}(\mathbf{x},t) = \sigma_{ij}(\mathbf{x},t) - \tfrac{1}{3}\sigma_{kk}(\mathbf{x},t)\delta_{ij} \tag{3.2.8}$$

The incremental form of the plastic constitutive relation will be written in rate form, with time as a parameter. The numerical code uses a specified load history and time steps to produce load steps. Thus, the elastic and plastic strain increments at any point will be converted to strain rates, $\dot{\epsilon}^e(\mathbf{x},t)$ and $\dot{\epsilon}^p(\mathbf{x},t)$, and the stress increment to a stress rate, $\dot{\sigma}(\mathbf{x},t)$. However, neither the elastic nor the plastic constitutive laws are explicitly physically dependent on time or loading rate, so that time does not appear in the formulation of the constitutive law used in the numerical model. The load history input to the model could be any continuous, increasing function of time to simulate the loading of the experiment. The use of time as a parameter in the numerical code allows in principle for the inclusion of explicit time or rate dependence, although this feature of the code was not used in this model.

The history of the deformation will be contained in the scalar quantity $\overline{\epsilon}^p(\mathbf{x},t)$, the accumulated equivalent plastic strain. The rate of change of the accumulated equivalent plastic strain, $\dot{\overline{\epsilon}}^p(\mathbf{x},t)$, will be defined by

$$\dot{\overline{\epsilon}}^p(\mathbf{x},t) = \left(\tfrac{2}{3}\dot{\epsilon}^p_{ij}(\mathbf{x},t)\dot{\epsilon}^p_{ij}(\mathbf{x},t)\right)^{1/2} \tag{3.2.9}$$

from which it follows that the current accumulated equivalent plastic strain, $\overline{\epsilon}^p(\mathbf{x},t)$, is described by an integral over the loading history:

$$\overline{\epsilon}^p(\mathbf{x},t) = \int_{-\infty}^{t} \left(\tfrac{2}{3}\dot{\epsilon}^p_{ij}(\mathbf{x},\tau)\dot{\epsilon}^p_{ij}(\mathbf{x},\tau)\right)^{1/2} d\tau \tag{3.2.10}$$

3.2.4 Yield Surface — Condition for Plastic Deformation

The yield surface is a function of the current stress and strain states, $f(\boldsymbol{\sigma}, \boldsymbol{\epsilon})$, defined in such a way that the condition for plastic deformation to occur is that $f(\boldsymbol{\sigma}, \boldsymbol{\epsilon}) = 0$. If the stress and strain states are such that $f(\boldsymbol{\sigma}, \boldsymbol{\epsilon}) < 0$, the deformation is purely elastic. The yield surface expands during plastic deformation so that it is impossible for the stress and strain to reach a state for which $f(\boldsymbol{\sigma}, \boldsymbol{\epsilon}) > 0$.

It is assumed that the dependence of the yield surface on the current strain state is through the accumulated plastic strain, $\overline{\epsilon}^p$. Using the assumption of plastic incompressibility and the history parameter $\overline{\epsilon}^p$, the yield function $f(\boldsymbol{\sigma}, \boldsymbol{\epsilon})$ is equivalent to a function $\hat{f}(\mathbf{S}, \overline{\epsilon}^p)$.

Further assuming continuous material isotropy and symmetry, the yield condition can depend only on the invariants of the deviatoric stress, J_2 and J_3 ($\mathrm{J}_1 \equiv 0$). In particular, in the small–strain, incremental form of the plastic constitutive relation used in the numerical model, yielding is assumed to depend only on the second invariant (J_2) of the deviatoric stress tensor, $\mathbf{S}$, and on the history of loading, which is contained in the variable $\overline{\epsilon}^p$. The invariant J_2 is found to be

$$\mathrm{J}_2(\mathbf{x}, t) = \tfrac{1}{2} S_{ij}(\mathbf{x}, t) S_{ij}(\mathbf{x}, t) = \tfrac{1}{2} S_{ij}(\mathbf{x}, t) \sigma_{ij}(\mathbf{x}, t) \tag{3.2.11}$$

The yield surface chosen for this study follows the Huber–Von Mises criterion, making it a sphere in deviatoric stress space, specifically,

$$f\big(\boldsymbol{\sigma}(\mathbf{x}, t), \boldsymbol{\epsilon}(\mathbf{x}, t)\big) = 3\mathrm{J}_2(\mathbf{x}, t) - \overline{\sigma}^2\big(\overline{\epsilon}^p(\mathbf{x}, t)\big) \tag{3.2.12}$$

where $\overline{\sigma}(\overline{\epsilon}^p)$ is a function chosen to model the constitutive behavior of the material in the experiment, in this case a power law.

3.2.5 Constitutive Assumption — Power Law Formulation

The function $\overline{\sigma}(\overline{\epsilon}^p)$, which is a function of the loading history through the accumulated equivalent plastic strain $\overline{\epsilon}^p$, is assumed for the following computations to follow a power law of the form:

$$\frac{\overline{\epsilon}^p}{\epsilon_0} = \left(\frac{\overline{\sigma}}{\sigma_0}\right)^n - \frac{\overline{\sigma}}{\sigma_0} \tag{3.2.13}$$

In equation (3.2.13), σ_0 is the initial yield stress in uniaxial tension, and ϵ_0 is the corresponding strain at yield in uniaxial tension. The function $\overline{\sigma}$ acts as the current yield stress during plastic deformation. The initial yield stress, σ_0, strain at yield, ϵ_0, and hardening exponent, n, will be fit to a uniaxial measurement of the steel used in the experiment. Note that Young's modulus, E, is related to the initial yield stress, σ_0, and the strain at yield, ϵ_0, by $\mathrm{E} = \sigma_0/\epsilon_0$.

3.2.6 Normality

Applying Drucker's postulate [28,34], which requires that any plastic work done through a cycle of stress be greater than zero; *i.e.*,

$$\oint(\sigma_{ij} - \sigma_{ij}^A)d\epsilon_{ij}^p > 0 \tag{3.2.14}$$

where σ_{ij}^A represents the same initial and final stress state, with σ_{ij}^A on or inside the initial yield surface. If the stress cycle lies entirely inside the yield surface, no plastic work is done.

There are two consequences of requiring that Drucker's postulate hold during plastic deformation. The first consequence is that the yield surface is required to be convex, which is satisfied by the choice of $f(\boldsymbol{\sigma}, \epsilon)$ as a sphere in deviatoric stress space in the previous section. The second consequence of Drucker's postulate is that the plastic strain increment must follow an associated flow rule, which means that the plastic strain increment (in strain space) must be parallel to the normal to the yield surface (in stress space), at the current applied stress. Since the yield surface defined by equation (3.2.12) describes a sphere in deviatoric stress space,

the plastic strain increment in strain space will be parallel to the deviatoric stress in stress space. The unit normal to the yield surface is given by $S_{ij}/(S_{kl}S_{kl})^{1/2}$, which is equal to $S_{ij}/(2\mathrm{J}_2)^{1/2}$ from equation (3.2.11). Thus, the plastic strain rate can be written as

$$\dot{\epsilon}^p_{ij}(\mathbf{x},t) = \dot{\lambda}(\mathbf{x},t)S_{ij}(\mathbf{x},t) \tag{3.2.15}$$

where $\dot{\lambda}(\mathbf{x},t)$ is a function of $\mathrm{J}_2(\mathbf{x},t)$ and $\overline{\epsilon}^p(\mathbf{x},t)$; $\dot{\lambda}(\mathbf{x},t)$ contains the magnitude of the plastic strain rate. Note that $\dot{\lambda}(\mathbf{x},t)$ is greater than zero when plastic deformation is occurring, and equal to zero when the deformation is elastic, such as during unloading.

3.2.7 Hardening

The magnitude of the plastic strain increment is a function of the hardening behavior of the material, which in this case is assumed to follow the power law of equation (3.2.13). From equations (3.2.9) and (3.2.15), it is seen that $\dot{\lambda}(\mathbf{x},t)$, the magnitude of the plastic strain increment, is

$$\dot{\lambda}(\mathbf{x},t) = \sqrt{\frac{3}{4\mathrm{J}_2(\mathbf{x},t)}}\,\dot{\overline{\epsilon}}^p(\mathbf{x},t) = \sqrt{\frac{3}{4\mathrm{J}_2(\mathbf{x},t)}}\,\frac{d\overline{\epsilon}^p}{d\overline{\sigma}}\dot{\overline{\sigma}}(\mathbf{x},t) \tag{3.2.16}$$

The condition for plastic deformation to occur is given by equation (3.2.12), which implies that $\overline{\sigma}^2(\mathbf{x},t) = 3\mathrm{J}_2(\mathbf{x},t)$. Thus, $\dot{\overline{\sigma}}(\mathbf{x},t)$ can be expanded to yield

$$\dot{\lambda}(\mathbf{x},t) = \frac{3}{4\mathrm{J}_2(\mathbf{x},t)}\frac{d\overline{\epsilon}^p}{d\overline{\sigma}}\dot{\mathrm{J}}_2(\mathbf{x},t) \tag{3.2.17}$$

It is possible to define a general tangent modulus function, H, which relates the rate of change of the function $\overline{\sigma}$ to the rate of change of the accumulated plastic strain, $\overline{\epsilon}^p$, for any constitutive law, so that

$$H = \frac{d\overline{\sigma}}{d\overline{\epsilon}^p} \tag{3.2.18}$$

Using the general tangent modulus, H, the plastic strain rate can be written as

$$\dot{\epsilon}^p_{ij}(\mathbf{x},t) = \frac{3}{4H\mathrm{J}_2(\mathbf{x},t)}\dot{\mathrm{J}}_2(\mathbf{x},t)S_{ij}(\mathbf{x},t) \tag{3.2.19}$$

For the power–law form of the constitutive law used in equation (3.2.13), the tangent modulus function may be calculated explicitly, in which case the plastic strain rate is

$$\dot{\epsilon}^p_{ij}(\mathbf{x},t) = \frac{3\epsilon_0}{4\sigma_0 \mathrm{J}_2(\mathbf{x},t)}\left[n\left(\frac{\overline{\sigma}(\mathbf{x},t)}{\sigma_0}\right)^{n-1} - 1\right]\dot{\mathrm{J}}_2(\mathbf{x},t)S_{ij}(\mathbf{x},t) \tag{3.2.20}$$

also recalling that $\overline{\sigma}^2(\mathbf{x},t) = 3\mathrm{J}_2(\mathbf{x},t)$, and $\mathrm{J}_2(\mathbf{x},t) = [S_{ij}(\mathbf{x},t)S_{ij}(\mathbf{x},t)]/2$, so that $\dot{\mathrm{J}}_2(\mathbf{x},t) = S_{ij}(\mathbf{x},t)\dot{S}_{ij}(\mathbf{x},t)$.

3.2.8 Constitutive Relation for the Total Strain Rate

The *total* strain rate $\dot{\epsilon}(\mathbf{x},t)$ can be related to the stress rate $\dot{\sigma}(\mathbf{x},t)$ by combining the plastic and elastic parts. This formulation may be done through use of the constant, elastic–compliance four–tensor, C_{ijkl}, and the plastic tangent modulus function, H, as well as the current stress state $\sigma_{ij}(\mathbf{x},t)$ and the current yield stress, $\overline{\sigma}(\mathbf{x},t)$.

$$\begin{aligned}\dot{\sigma}_{ij}(\mathbf{x},t) &= \left[C_{ijkl} - \frac{C_{ijpq}S_{pq}(\mathbf{x},t)S_{mn}(\mathbf{x},t)C_{mnkl}}{S_{rt}(\mathbf{x},t)C_{rtuv}S_{uv}(\mathbf{x},t) + \frac{4}{9}\overline{\sigma}^2(\mathbf{x},t)H}\right]\dot{\epsilon}_{kl}(\mathbf{x},t) \\ &= C^*_{ijkl}(\mathbf{x},t)\dot{\epsilon}_{kl}(\mathbf{x},t)\end{aligned} \tag{3.2.21}$$

The quantity $C^*_{ijkl}(\mathbf{x},t)$ defined in equation (3.2.21) is the history–dependent compliance tensor, which may be used to relate the stress and strain rates during plastic deformation. During elastic deformation, the compliance $C^*_{ijkl}(\mathbf{x},t)$ reduces to the constant elastic compliance C_{ijkl} .

3.2.9 Deformation Plasticity

The crack tip deformation field of Hutchinson [31] and Rice and Rosengren [69], called the HRR field, is derived within the framework of deformation–theory plasticity. Deformation–theory plasticity employs the simplifying assumption that the plastic strain is directly related to the stress. This assumption contrasts with the incremental theory discussed above, in which the plastic strain rate is related to the stress rate. The advantage of deformation plasticity is that it is identical to a nonlinear elasticity law, in that the plastic strains depend only on the current state, independent of the loading history. In general, incremental theory is considered to be more correct than deformation theory, although for certain types of problems involving nonproportional loading, deformation plasticity appears to correspond more closely to experiments than does incremental plasticity. Also, the simpler formulation of deformation theory makes it more easily applied [28,34]. The finite–element code used in this study employs an incremental theory to model as closely as possible the physical body. However, two important tools in fracture analysis are derived within the framework of deformation theory, namely, the J–integral of Rice [65,66] and the HRR field approximation to crack tip deformation fields in power–law hardening materials. The J–integral is a measure of the energy input to the crack tip, and a critical value of the J–integral is a commonly used fracture criterion. Both the J–integral and the HRR field will be discussed in more detail in the next chapter.

Shih [77] has tabulated HRR field quantities for plane strain and plane stress and for a variety of hardening exponents (n). Shih's deformation–theory, constitutive law is expressed as

$$\frac{\epsilon_{ij}^{p}(\mathbf{x},t)}{\epsilon_0} = \frac{3}{2}\alpha\left(\frac{\overline{\sigma}(\mathbf{x},t)}{\sigma_0}\right)^{n-1}\frac{S_{ij}(\mathbf{x},t)}{\sigma_0} \tag{3.2.22}$$

where it will be recalled that σ_0 is the initial yield stress in uniaxial tension, and ϵ_0 is the strain at yield in uniaxial tension, also that $\overline{\sigma}^2(\mathbf{x},t) = 3\mathrm{J}_2(\mathbf{x},t)$, with $\mathrm{J}_2(\mathbf{x},t) = [S_{ij}(\mathbf{x},t)S_{ij}(\mathbf{x},t)]/2$. The parameter α is a material constant that may be used to fit the power law to a measured stress–strain relation; α will be taken as unity for this study. The HRR field approximation neglects any elastic strain in

comparison to the plastic strain.

For a given loading history, the resulting strains and displacements calculated using deformation theory will, in general, differ from the strains and displacements calculated using incremental theory. The exception to this rule is the case of proportional loading, in which the stress tensor can be written as a tensor that is only a function of position multiplied by a monotonically increasing scalar function of time. When the loading is strictly proportional, deformation theory and incremental theory coincide. However, Kachanov [34] also cites the case in which the loading path "develops in a definite direction" as being similar to proportional loading, meaning that the loading approaches proportional loading as time increases. In this case, the deformation theory and the incremental theory, while initially different, will become closer as time progresses. Additionally, for some specific yield surfaces having corners, there is a certain class of loading paths for which the deformation and incremental theories coincide [16].

Deformation theory and incremental theory diverge when there is unloading. For an incremental theory, the unloading path will follow a linearly elastic constitutive law, at least until the stress state reaches another point on the yield surface. The unloading path in deformation theory retraces its steps along the loading path. In the context of the present work, this feature may cause difficulties when there is tunneling of the crack within the fracture specimen. The unloading that is due to tunneling means that the assumption of deformation theory plasticity (or incremental plasticity, assuming proportional loading) used for J–integral or HRR field calculations is no longer valid.

3.2.10 Incremental Plasticity Under Proportional Loading

As mentioned, deformation and incremental plasticity theories coincide only in the case of proportional loading. Since the HRR field is derived using a deformation–theory plasticity, and since the finite–element model employs an incremental–theory plasticity, one issue that may be raised in making a comparison between the HRR field and the finite–element model is whether or not the loading history is proportional or at least "develops in a definite direction." A second issue that arises is whether the incremental constitutive model, assuming conditions of proportional loading, matches the deformation plasticity model employed by Shih [77] in tabulating the HRR field quantities, whose tabulated values will be used for comparison with the finite–element model.

As observed by Kachanov [34], proportional loading is strictly valid only for small–strain theory, in which no distinction is made between the position of a particle in the undeformed configuration compared with the position of that particle in a deformed configuration. Both the HRR field and the finite–element model are restricted to small–strain plasticity.

By assuming proportional loading, the deformation–theory analog to the incremental plasticity constitutive law of equation (3.2.20) may be derived. That analog will be used to compare the constitutive law used in the finite–element model with the deformation–theory constitutive law used by Shih [77].

Under conditions of proportional loading, the stress is assumed to be separable in space and time, taking the form $\sigma_{ij}(\mathbf{x},t) = \sigma^0_{ij}(\mathbf{x})g(t)$, where $\sigma^0_{ij}(\mathbf{x})$ is independent of time, and with $g(t) \geq 0$ and $\dot{g}(t) \geq 0$.

The constitutive law may now be written as a relation between the stress and strain states directly, as opposed to a relation between the stress rate and the strain rate. From equation (3.2.13), which relates the accumulated, equivalent plastic strain $\overline{\epsilon}^p(\mathbf{x},t)$ and the current yield stress $\overline{\sigma}(\mathbf{x},t)$ through a power law, the plastic strain may be found as

$$\frac{\epsilon^p_{ij}(\mathbf{x},t)}{\epsilon_0} = \frac{3}{2}\left[\left(\frac{\overline{\sigma}(\mathbf{x},t)}{\sigma_0}\right)^{n-1} - 1\right]\frac{S_{ij}(\mathbf{x},t)}{\sigma_0} \tag{3.2.23}$$

where it will be recalled that σ_0 is the initial yield stress in uniaxial tension, and ϵ_0 is the strain at yield in uniaxial tension, and also that $\overline{\sigma}^2(\mathbf{x},t) = 3\mathrm{J}_2(\mathbf{x},t)$, with $\mathrm{J}_2(\mathbf{x},t) = [S_{ij}(\mathbf{x},t)S_{ij}(\mathbf{x},t)]/2$. The elastic strains of equation (3.2.4) may be written in a similar form as

$$\frac{\epsilon^e_{ij}(\mathbf{x},t)}{\epsilon_0} = (1+\nu)\frac{S_{ij}(\mathbf{x},t)}{\sigma_0} + \frac{(1-2\nu)}{3}\frac{\sigma_{kk}(\mathbf{x},t)}{\sigma_0}\delta_{ij} \tag{3.2.24}$$

Note that if Poisson's ratio ν is equal to one half, the first term of equation (3.2.24) cancels the second term of equation (3.2.23).

Equations (3.2.23) and (3.2.24) may be compared with Shih's constitutive law of equation (3.2.22). If the elastic strains may be neglected in comparison to the plastic strains, the second term in equation (3.2.23) may be neglected in comparison to the first term, and equation (3.2.23) becomes identical to equation (3.2.22) (with α taken as unity). Hutchinson [verbal communication] believes that any difficulty with the HRR field comes from neglecting the elastic strains, particularly the dilatational strains.

3.2.11 Uniaxial Test to Determine Hardening Behavior

The yield stress, the strain at yield, and the hardening exponent, n, are determined using data obtained from a uniaxial measurement. As long as the specimen has not begun to neck, the uniaxial test is a case of proportional loading, so that the stress and strain may be related directly. In the uniaxial test, the stress and strain are measured along one axis, here chosen arbitrarily to be the x_1 axis. It is assumed that all stress components are zero except the tensile stress in the x_1 direction, and that the stress and strain in the test section are independent of position, so that the stress tensor may be written as

$$\sigma_{ij}(\mathbf{x},t) = \delta_{1i}\delta_{1j}\,\sigma_{11}(t) \tag{3.2.25}$$

where $\sigma_{11}(t) \geq 0$ and $\dot{\sigma}_{11}(t) \geq 0$. The deviatoric stress has the same time dependence, with $S_{ij}(\mathbf{x},t)$ given by

$$S_{ij}(\mathbf{x},t) = (\delta_{1i}\delta_{1j} - \tfrac{1}{3}\delta_{ij})\,\sigma_{11}(t) \tag{3.2.26}$$

The second invariant of the deviatoric stress, J_2, is thus $\frac{1}{3}\sigma_{11}^2(t)$. The condition for plastic yielding, $f(\boldsymbol{\sigma}, \boldsymbol{\epsilon}) = 0$ implies that

$$\overline{\sigma}^2(\overline{\epsilon}^p) = \sigma_{11}^2(t) \tag{3.2.27}$$

Since both $\sigma_{11}(t)$ and $\overline{\sigma}$ are positive, equation (3.2.27) is equivalent to

$$\overline{\sigma}(\overline{\epsilon}^p) = \sigma_{11}(t) \tag{3.2.28}$$

Thus, the current yield stress $\overline{\sigma}$ at time t is equal to the applied stress, $\sigma_{11}(t)$, if $\sigma_{11}(t)$ is greater than the initial yield stress, σ_0.

As mentioned, only the strain $\epsilon_{11}(t)$ is measured. Equations (3.2.23) and (3.2.24) may be used to determine the elastic and plastic parts of $\epsilon_{11}(t)$. The elastic strain is found as

$$\frac{\epsilon_{11}^e(t)}{\epsilon_0} = \frac{2(1+\nu)}{3}\frac{\sigma_{11}(t)}{\sigma_0} + \frac{(1-2\nu)}{3}\frac{\sigma_{11}(t)}{\sigma_0} = \frac{\sigma_{11}(t)}{\sigma_0} \tag{3.2.29}$$

for all $\sigma_{11}(t)$. The plastic strain is found as

$$\frac{\epsilon_{11}^p(t)}{\epsilon_0} = \left[\left(\frac{\sigma_{11}(t)}{\sigma_0}\right)^{n-1} - 1\right]\frac{\sigma_{11}(t)}{\sigma_0} \tag{3.2.30}$$

for $\sigma_{11}(t) \geq \sigma_0$. Thus, the total strain, $\epsilon_{11}(t)$, is given by

$$\frac{\epsilon_{11}(t)}{\epsilon_0} = \left(\frac{\sigma_{11}(t)}{\sigma_0}\right)^n \tag{3.2.31}$$

when $\sigma_{11}(t) \geq \sigma_0$.

Equation (3.2.31) is fit to the measured data from a uniaxial test to provide the hardening parameter n for the numerical simulation. The initial yield stress, σ_0, and strain at yield, ϵ_0, are determined as the end of the linear region from the measured data, and Young's modulus, E, is determined as $E = \sigma_0/\epsilon_0$. The hardening exponent is determined by fitting a straight line through a plot of $\log \sigma$ against $\log \epsilon$. The experimental measurement of Poisson's ratio will be discussed in Section 3.5.

3.3 Experimental Material Characterization

The material used in the three–point–bend study is 4340 steel, described in the ASM Metals Handbook [46] as an ultrahigh–strength steel: "AISI/SAE 4340 steel is considered the standard to which other ultrahigh–strength steels are compared. It combines deep hardenability with high ductility, toughness and strength." 4340 is a low alloy, medium carbon steel. The "43" in the designation indicates that it is a Nickel–Chromium–Molybdenum steel. The "40" in the designation indicates a nominal carbon content of 0.40% by weight, or about 2% carbon atoms compared to 98% iron atoms. The composition of the steel used in the experiments is listed below, as provided by the mill. The term "deep hardenability" refers to the material's ability to harden even in relatively thick sections, and is a result primarily of the nickel content of the steel. Hardening is accomplished through a rapid cooling of the steel from a high temperature, causing a phase transformation from Austenite to Martensite. This rapid cooling is accomplished by quenching the steel in an oil bath. For unalloyed carbon steels, the cooling rate must be very high to cause the proper phase transformation. The addition of the alloying elements reduces the cooling rate necessary to achieve the desired phase transformation, allowing hardening to take place in a deeper material section. The quenched steel has a high yield stress, but it is very brittle, having little ductility.

Tempering the steel, which means reheating the quenched steel, can increase the ductility without sacrificing all of the increase in the yield stress gained by quenching. The yield stress of 4340 steel can have a range from about 700 to 1900 MPa, depending on the choice of the tempering temperature. The higher the tempering temperature, the lower the yield stress and the higher the ductility. The yield stress of 700 MPa corresponds to a tempering temperature of about 700°C, and the yield stress of 1900 MPa corresponds to a tempering temperature of about 200°C, according to the Metals Handbook.

The 4340 steel was specified to be aircraft quality and vacuum–degassed. Aircraft quality is a standard descriptor indicating that certain specifications have been met in the production of the steel and that the steel is suitable for use in aircraft

Chemical Composition, % weight											
C	Mn	P	S	Si	Ni	Cr	Mo	Cu	Sn	Al	Fe
0.420	0.740	0.010	0.012	0.250	1.710	0.800	0.240	0.060	0.000	0.036	balance

Table 1. Composition of the 4340 Steel used in the experiments. Aircraft quality steel, vacuum–degassed, cold drawn.

structures. Vacuum degassing is a method used to remove oxygen and hydrogen from the molten steel. Hydrogen can collect in the grain boundaries and can cause brittle failure of the steel.

3.3.1 Uniaxial Tests

As has been mentioned, the uniaxial tests were initially planned to ensure that a meaningful comparison could be made between the measurements made for this study; the measurements by Zehnder [91] and Zehnder and Rosakis [92]; and the numerical simulation of that experiment by Narasimhan and Rosakis [51]. The material properties were measured in a manner identical to that used by Zehnder [91] to ensure that the measurements would be directly comparable.

Four fracture test specimens and two uniaxial test specimens were cut from the same piece of steel, a plate four inches wide by one-half inch thick by eight feet long. Two additional uniaxial test specimens were made from 4340 steel remaining from previous investigations. These specimens were included in order to estimate the variation of the uniaxial measurement between different sources of steel, particularly in comparison with the two specimens of steel purchased for this investigation. The uniaxial specimens were numbered in such a way that specimens 1 and 3 were made from the steel purchased for the current study, and specimens 2 and 4 were made from the leftover steel.

Each uniaxial test specimen was heat–treated at the same time as a fracture specimen, to ensure uniformity of the heat treatment. The steel was heat–treated

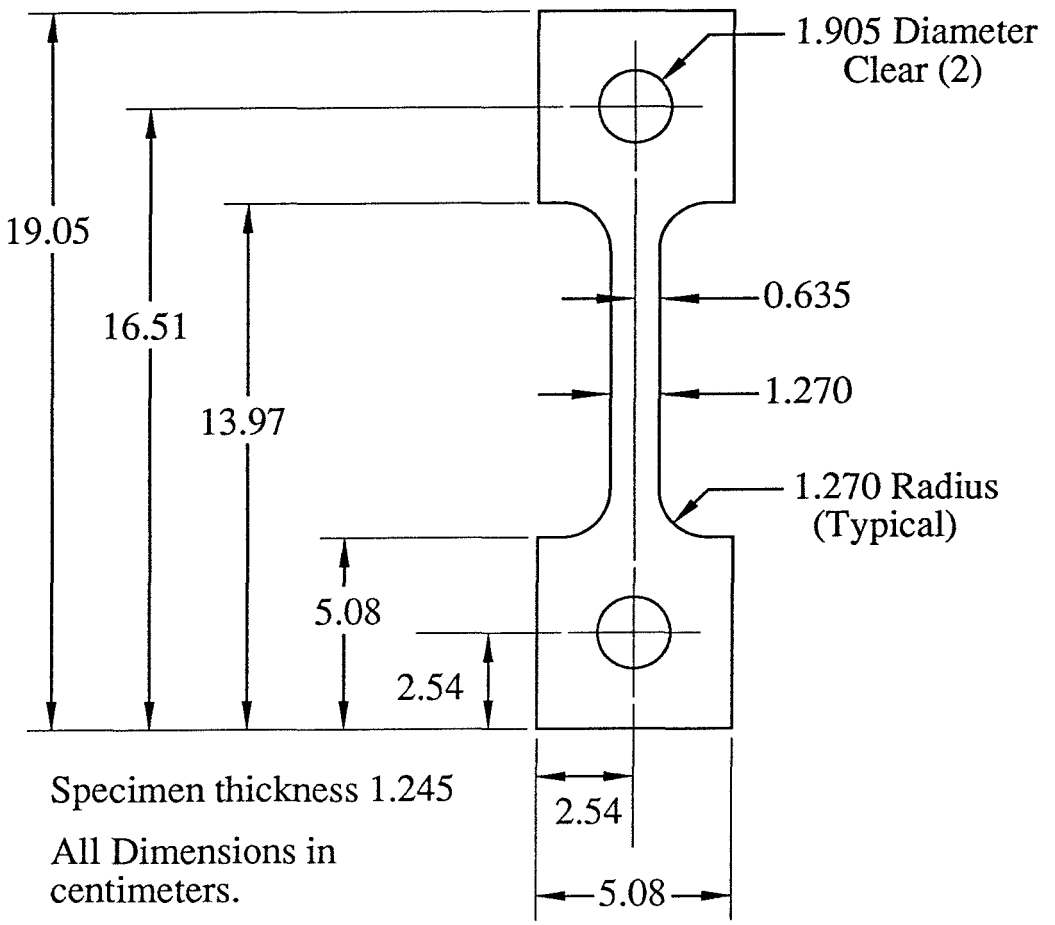


FIGURE 44. Uniaxial test specimen geometry. The material is a ductile heat treatment of 4340 steel. The test section is ground flat and parallel with a thickness of 1.245 cm.

to give a ductile response, first being "normalized" at 870 °C for one hour and air–cooled to remove any effects of rolling or machining, and leaving the steel in its stable equilibrium state. The steel was then heated to 815 °C for one and one-half hours and oil–quenched to produce a Martensitic transformation. Finally, the steel was tempered at a temperature of 540 °C for one hour and air–cooled. The uniaxial test geometry is shown in Figure 44. The specimen dimensions are adapted from ASTM Standard E 646–78 [6], with some modification to account for the relatively large specimen thickness. The calculation of the hardening exponent for use in the

power law is as indicated in ASTM Standard E 646–78 [6].

3.3.2 Testing Procedure

The uniaxial tests were done in the Baldwin hydraulic loading machine (capacity 1,000,000 N), which does not allow precise control of either the force or displacement history. The tests were done with the loading valve slightly open to give a slowly increasing load to failure of the specimen. The overall strain rate was less than 0.5%/sec in each case. The load was read from a SENSOTEC 400,000 N capacity load cell, and the strain was read from two Instron strain–gage extensometers, a 10% extensometer with a 2–inch gauge length and a 100% extensometer with a 1–inch gauge length. All three signals were recorded on a Masscomp computer data acquisition and control system using Laboratory Workbench software under a Unix operating system.

3.3.3 Data Reduction

The Laboratory Workbench software allows for the inclusion of multiplying factors and offset voltages during the recording of the data, so the calibrations of the load cell and extensometers could be used to convert the voltages to load and displacement units during the experiment. Also, the gauge length of each extensometer could be divided out to give the engineering strain $\mathrm{e} = \Delta L/L_0$. The engineering stress s was then calculated from the recorded load P and the initial cross–sectional area A_0 as $s = P/A_0$. When the specimen is deforming plastically, but before the onset of necking, the assumption of plastic incompressibility relates the current area, A, to the initial area, A_0, through the engineering strain as $A = A_0/(1+\mathrm{e})$. Thus, the true stress σ $(= P/A)$ is calculated by $\sigma = s(1+\mathrm{e})$. The true strain ϵ is calculated by noting that $\mathrm{d}\epsilon = \mathrm{d}L/L$. Integrating this quantity from the initial length L_0 to the current length L gives $\epsilon = \ln(L/L_0) = \ln(1+\mathrm{e})$. (For small strains, this relation can be linearized as $\epsilon = \Delta L/L_0$.)

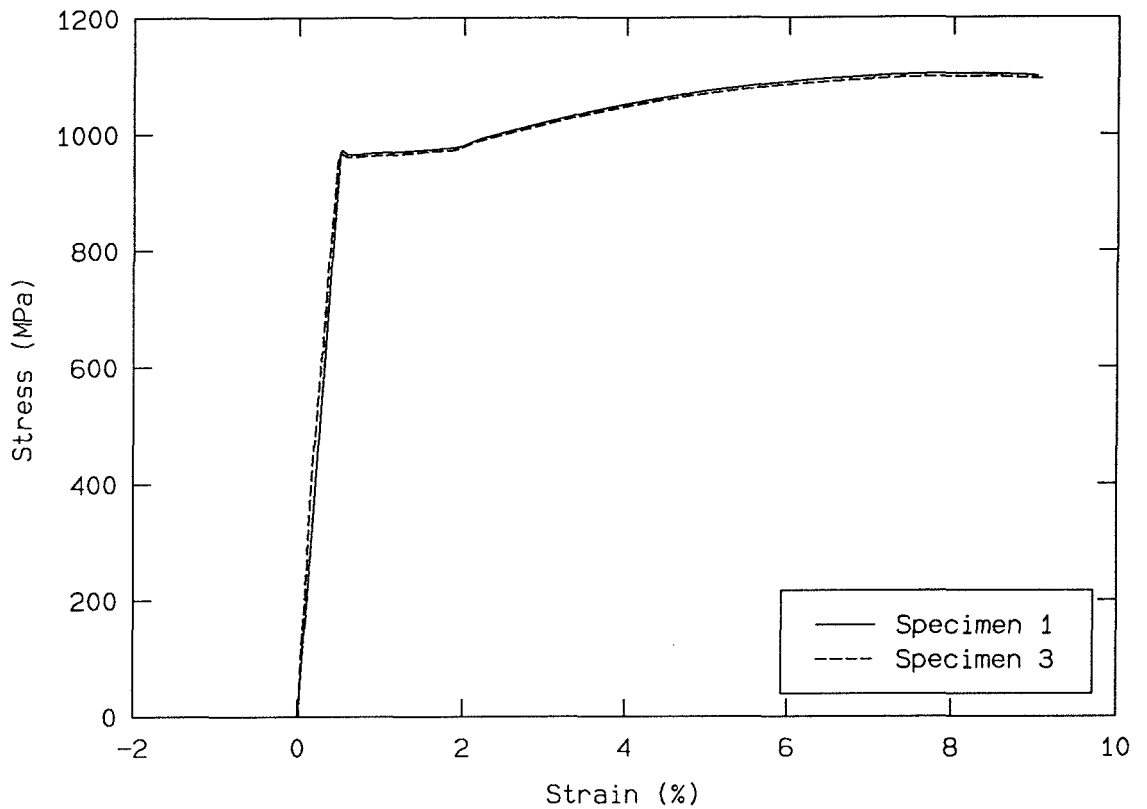


FIGURE 45. The stress–strain data for specimens 1 and 3, the steel used in the interferometric investigation. These data were taken using the 10% extensometer with a two–inch gauge length.

The resulting stress–strain curves for specimens 1 and 3 are shown in Figures 45 and 46. These data were taken using the 10% extensometer, which was released from the specimen as it reached the limit of its travel. The onset of necking can be seen at about 8% strain. From these data, the yield stress for the steel was determined to be 960.0 MPa, and the strain at yield was found to be 0.500%. Young's modulus is therefore found as 192.0 GPa.

A small "lip" can be seen in the curves in Figure 45, just at the yield point, where the stress level decreases sharply. The size of this drop in stress is sensitive to the rate of loading, and to the initial imperfections of the specimen. Since the size of this "lip" is specimen–dependent, the yield stress of the steel is usually given

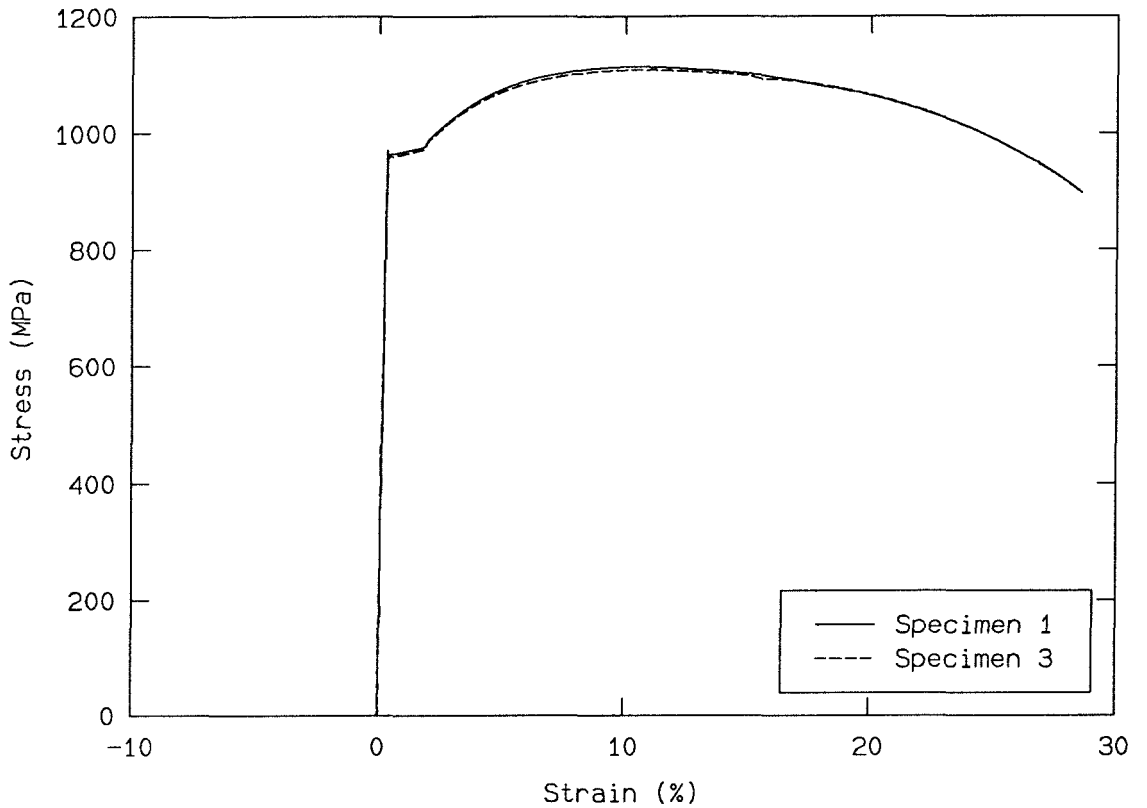


FIGURE 46. The stress–strain data for specimens 1 and 3, the steel used in the interferometric investigation. These data were taken using the 100% extensometer with a one–inch gauge length.

as the bottom of the stress drop, rather than the maximum stress.

The flat part of the stress–strain curve between yield and 2% strain is associated with the localized plastic flow seen as Lüder's bands [26,58]. Lüder's bands are seen at approximately 45° to the loading axis, along the lines of maximum shear. As discussed by Hertzberg [26], and by Polakowski and Ripling [58], the localized yielding is a result of the carbon and nitrogen residing in the interstitial regions between iron atoms. The carbon and nitrogen atoms tend to collect in the large gaps between planes of iron atoms at edge dislocations, pinning the edge dislocations under small shearing stresses. When the stress on such a pinned dislocation reaches a critical value, the dislocation is released. The released dislocations require a smaller

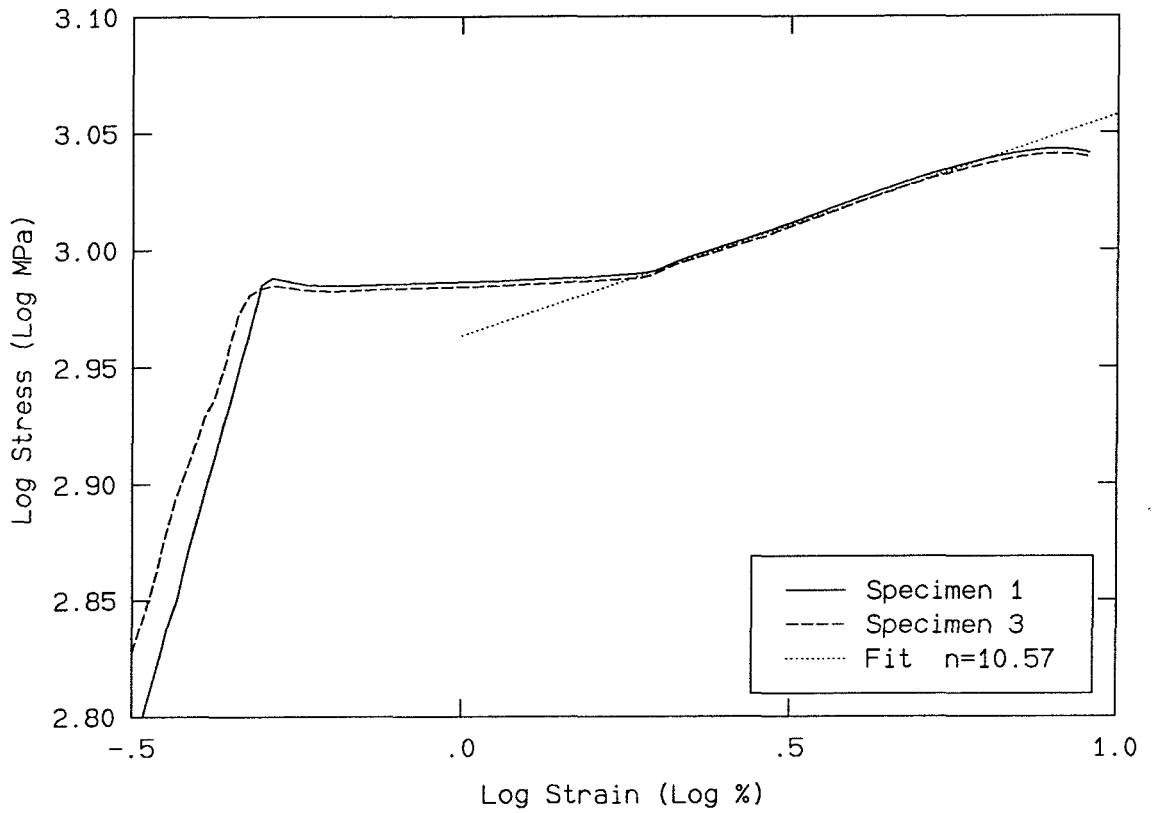


FIGURE 47. The stress–strain data for specimens 1 and 3, shown on logarithmic scales. After the Lüder's bands have passed through the test section, the plastic deformation is well described by a power law, in this case fit with a hardening exponent of 10.57.

stress to move through the steel. Strain hardening, which is due mainly to interactions between dislocations, will eventually halt these released dislocations, causing an increase in stress until all of the carbon– and nitrogen–pinned dislocations have been released, and the Lüder's bands have passed through the test section.

After the Lüder's bands have passed through the test section, the material continues to undergo homogeneous plastic deformation governed by strain hardening, until necking begins. This part of the stress–strain behavior is well described by a power law, as can be seen in Figure 47, where the stress and strain are shown on logarithmic scales. Figure 47 includes a line fit through the power–law part of the

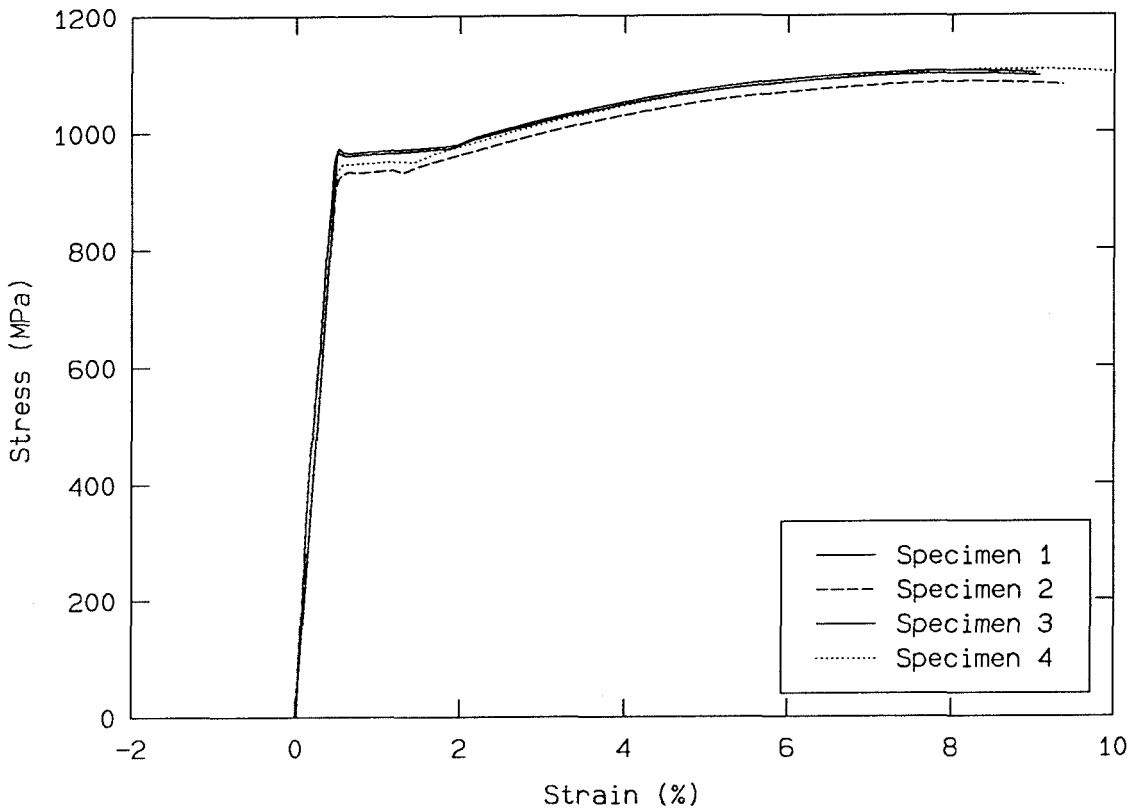


FIGURE 48. The stress–strain data for specimens 1 through 4,measured using the 10% extensometer with two–inch gauge length. Specimens 1 and 3 are made from the steel used in the interferometric experiment, while specimens 2 and 4 are made from steel left over from previous investigations.

stress–strain response.

The stress–strain behavior determined using the 10% extensometer for all four uniaxial specimens is shown in Figure 48. Specimens 1 and 3, corresponding to the steel used in the interferometric measurements, are shown in solid lines. Specimens 2 and 4, steel remaining from previous studies, are shown as the dashed and dotted lines, respectively. This figure is intended to display the scatter that is due to the different sources of steel, and to indicate the repeatability of the measurement between specimens 1 and 3.

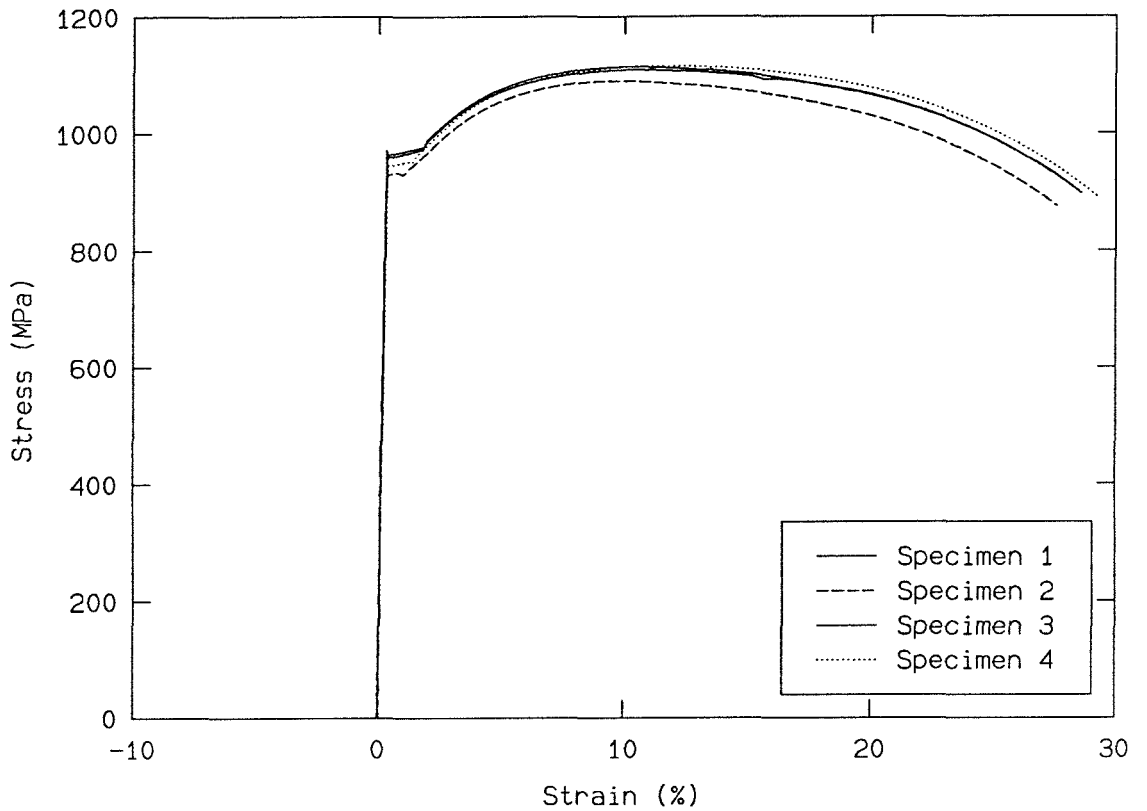


FIGURE 49. The stress–strain data for specimens 1 through 4 as measured by the 100% extensometer with one–inch gauge length.Specimens 1 and 3 are made from the steel used in the interferometric experiment, while specimens 2 and 4 are made from steel left over from previous investigations.

The stress–strain response measured using the 100% extensometer is shown in Figure 49. This is also intended to indicate the variation among the four specimens. These data are somewhat more suspect than those from the 10% extensometer as the 100% extensometer had a smaller gauge length (1 inch). This smaller gauge length made it possible for one arm of the extensometer to be within the necked region at higher strains. This is true for specimen 2 in particular, as one arm of the 100% extensometer was in the necked region.

The highest point on the curves indicates the onset of necking in the specimen. Until necking occurs, it is assumed that the stress and strain are homogeneous

throughout the length of the test section of the specimen. This assumption of homogeneity is used in the data reduction to estimate the cross–sectional area of the specimen from the measured extension (strain). Necking is a localization of the deformation, making the assumption of homogeneity invalid. Although the stress seems to decrease after necking, in fact, the area of the neck is reducing faster than the load is decreasing, causing an increase in the strain at the neck. The behavior of the material after necking will be discussed further in Section 3.3.7 below.

3.3.4 Power–Law Formulation

The formulation of the constitutive relation used in the finite–element code made it possible to input a power law to model the plastic hardening of the material. The uniaxial constitutive relation for input to the numerical model is expressed in the following way:

$$\frac{\epsilon}{\epsilon_0} = \begin{cases} \dfrac{\sigma}{\sigma_0} & \sigma < \sigma_0 \\ \left(\dfrac{\sigma}{\sigma_0}\right)^n & \sigma \geq \sigma_0 \end{cases} \tag{3.3.1}$$

where σ_0 is the yield stress, ϵ_0 is the corresponding strain at yield and n is the hardening exponent.

The power–law constitutive model is used in order to compare the experimental three–point–bend results with the analytic HRR field, an asymptotic, two–dimensional approximation to the deformation at a crack tip for a material having a power–law constitutive relation. However, this form of the power law for the finite–element model is constrained in that it has to pass through the stress and strain at yield. As seen in Figure 47, a power law does model the region of homogeneous plastic deformation quite well, but it does not pass through the yield stress and strain.

A power law of the form in equation (3.3.1) was least–squares fit, minimizing the error in stress as a function of strain for the stress-strain data from yield to a strain of 8 %, where necking can be seen to initiate. This least–squares fit incorporated the constraint of passing through the yield stress and strain. The resulting stress–

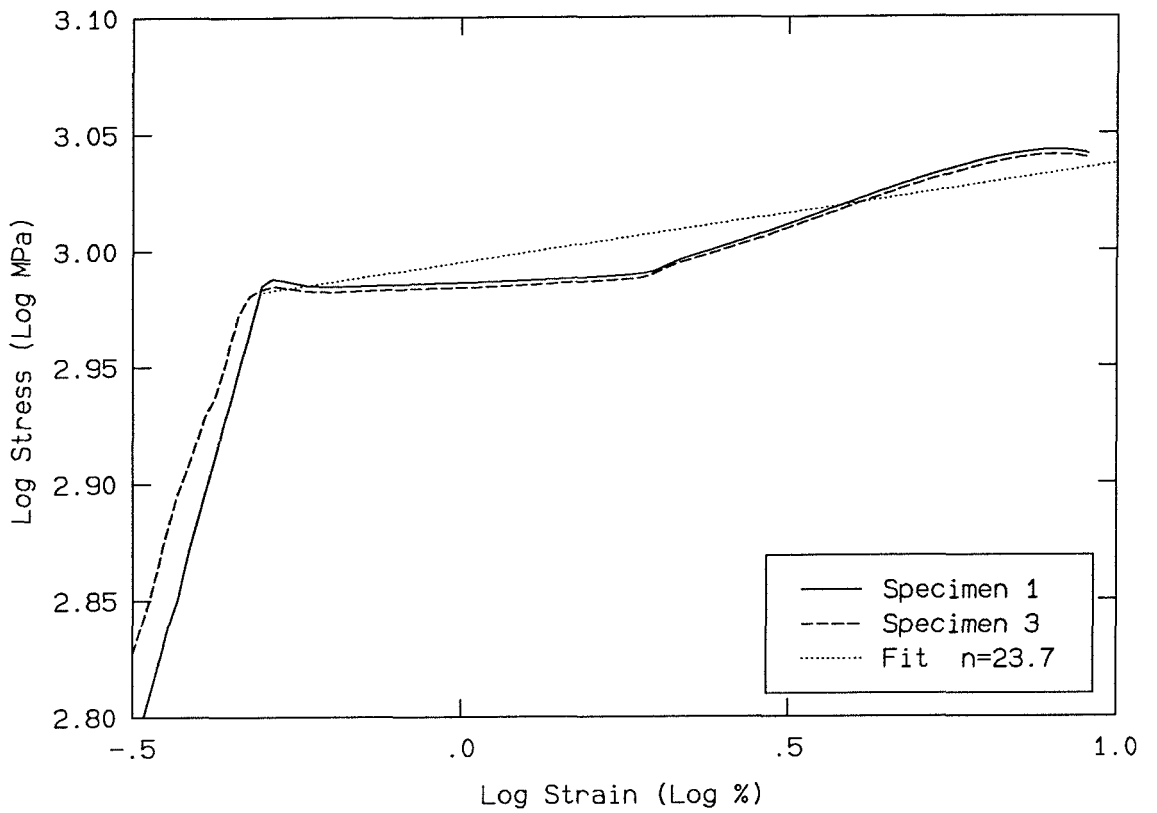


FIGURE 50. The stress–strain data for specimens 1 and 3 on logarithmic scales, including the power law that has been least–squares fit through the data after yielding. This power law is constrained to pass through the yield stress and strain.

strain law is compared with the experimental data in Figure 50 on a logarithmic scale, and in Figure 51 on a linear scale. The hardening exponent n is calculated to be $n = 23.7$.

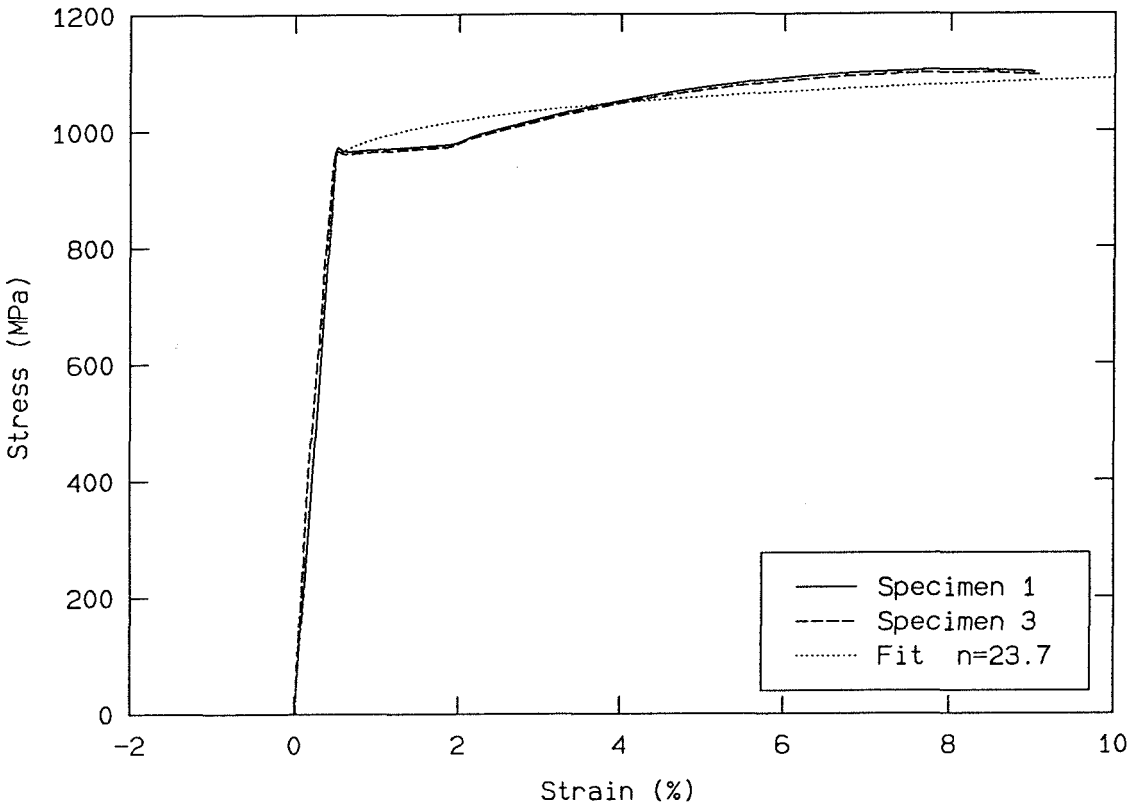


FIGURE 51. The stress–strain data for specimens 1 and 3, and the power law fit through the data after yielding.

3.3.5 Comparison with Previous Work

This investigation was in some part motivated by the previous work of Zehnder [91], Zehnder and Rosakis [92], and Narasimhan and Rosakis [51], as they had provided a basis with which to compare the results of the current study. In comparing the measured uniaxial response of specimens 1 and 3 to that of Zehnder, it was found that the yield stresses differed by about 10%, and the Young's modulus by about 5%. Zehnder's data and the current constitutive relation are shown in Figure 52. It can be seen that the shape of the two curves is similar, and that fact suggests that the stress–strain relations of the two studies can be compared by normalizing the data by the yield stress. The normalized data are shown in Figure 53. This normalization is a subset of normalizing the stress by the yield stress and

the strain by the strain at yield. Since the strain at yield is very nearly the same for the various stress–strain curves being compared, that normalization has been neglected.

In fact, since the finite–element model was rerun, and since the previous investigations [51,91,92] did not look at the in–plane displacements at the crack tip, little use has been made of the data available from their work for this thesis. However, comparisons can be made using the out-of-plane displacements, and it is interesting to see the range of material properties that can be attained with 4340 steel.

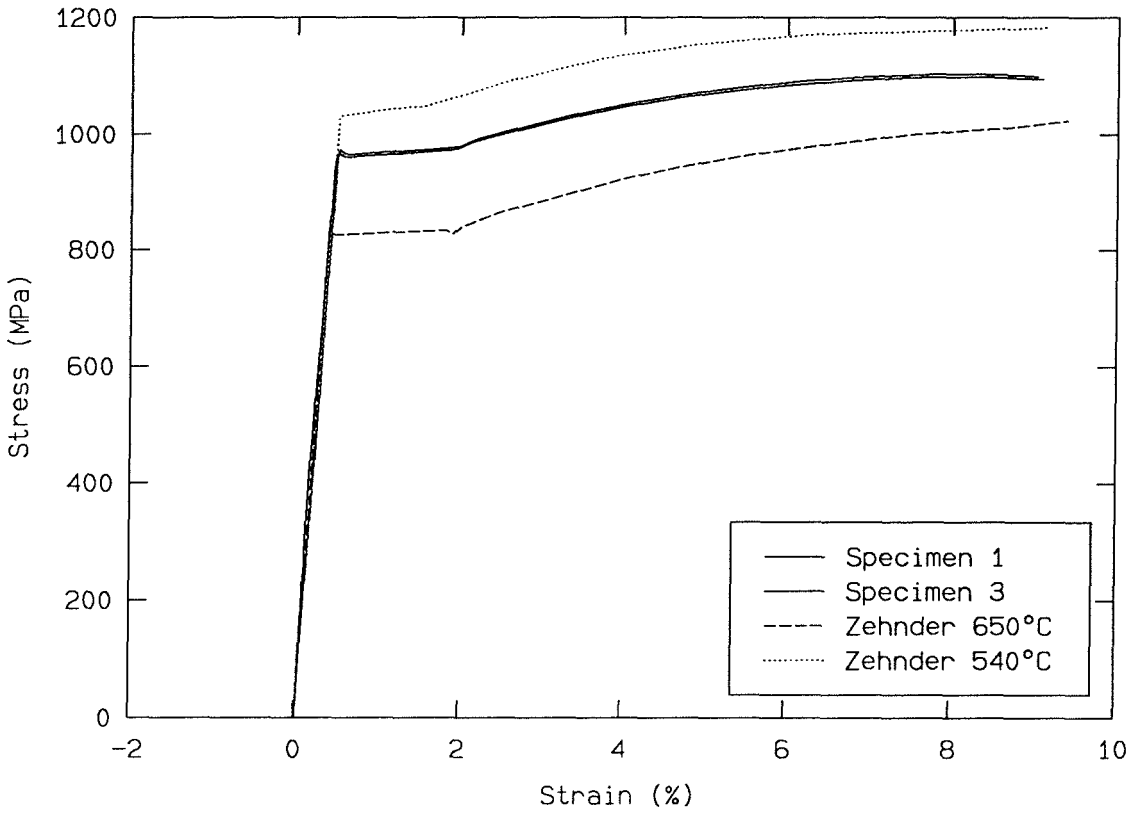


FIGURE 52. The stress–strain data for specimen 1 and 3, shown with data from Zehnder [91]. The temperature in the legend corresponds to the tempering temperature of the steel. Specimens 1 and 3 were also tempered at 540 °C.

In the numerical simulation of Narasimhan and Rosakis [51], the Young's Mod-

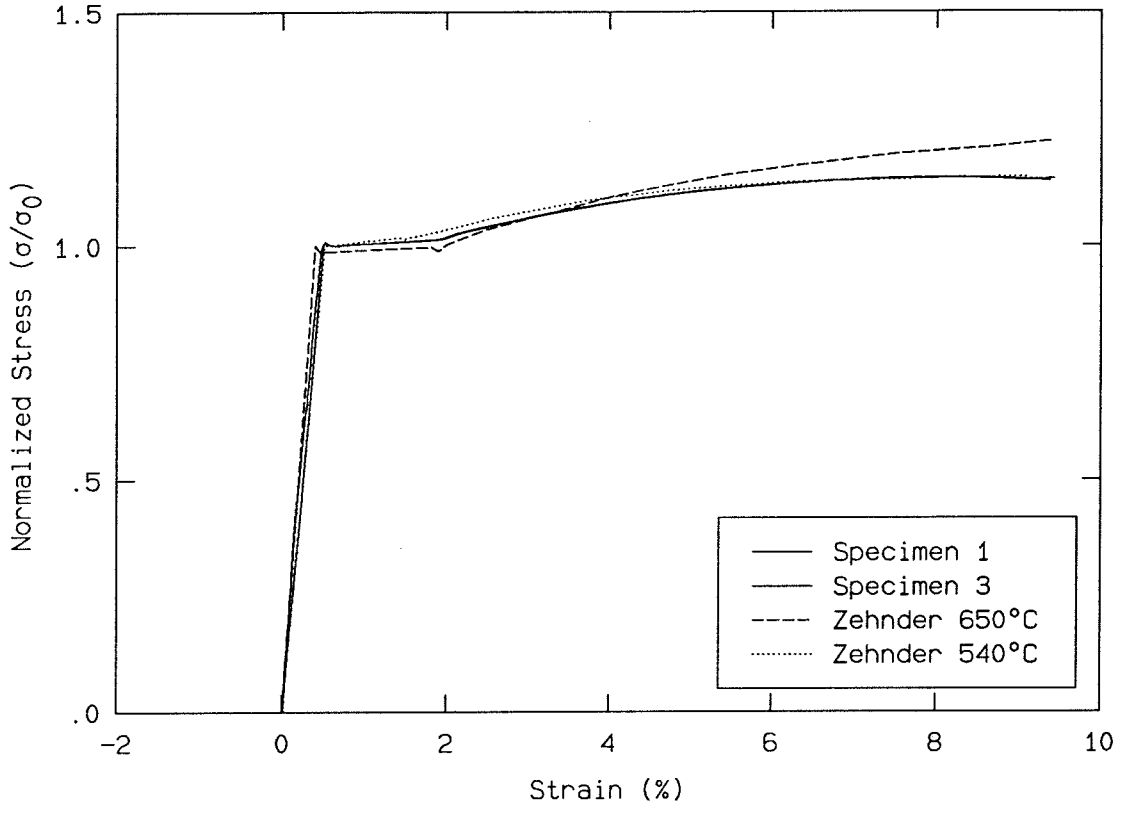


FIGURE 53. The stress–strain data for specimens 1 and 3, with the data from Zehnder [91], normalized by the yield stress.

ulus was 199.2 GPa, while the yield stress σ_0 and strain at yield ϵ_0 were determined to be $\sigma_0 = 1067$MPa and $\epsilon_0 = 0.536\%$. The hardening exponent n was found to be $n = 22.0$.

3.3.6 Comparison Between Specimens

The notion that a comparison of the constitutive behavior of different studies on 4340 steel could be made by normalizing the constitutive models by the yield stress is probably too simplistic, however. The data being compared here are from a small part of the range of material behavior possible with 4340 steel, which is controlled through the heat treatment. Figure 54 shows the result of normalizing the data from uniaxial specimens 1 through 4 by the yield stresses. It can be

seen that the normalization does not reduce the scatter much in the measurements away from the yield point. This is partly due to varying lengths of that part of the curve associated with Lüder's bands, perhaps caused by varying amounts of nitrogen dissolved in the steel (the carbon content should match closely). Also, it must be recognized that the final catastrophic fracture of the three–point–bend specimen is more closely related to that part of the stress–strain response farthest from the yield stress, at the largest strains. Figure 55 shows the data from the 100% extensometer for specimens 1 through 4 when normalized by the yield stress. Comparing Figure 55 to Figure 49, it can be seen that the normalization has little effect on reducing the variations at high strains.

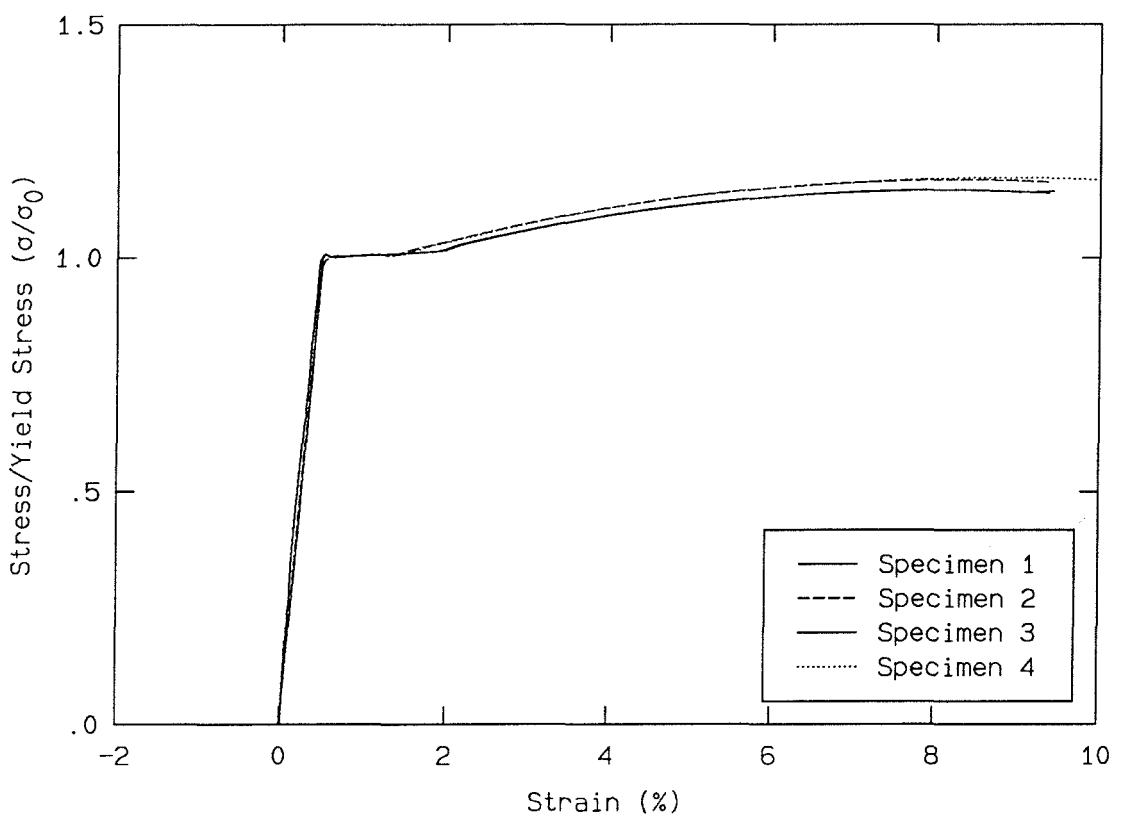


FIGURE 54. The stress–strain data for specimens 1 through 4 normalized by the yield stress. Data from the 10% extensometer.

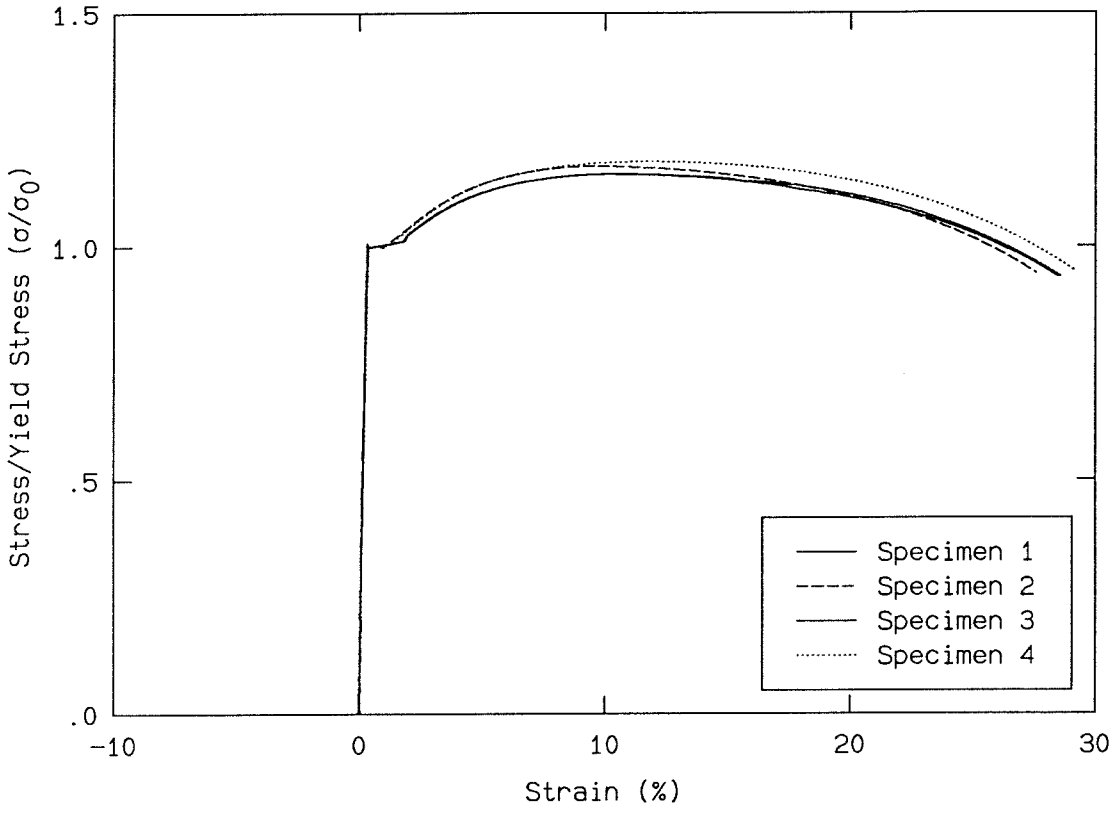


FIGURE 55. The stress–strain data for specimens 1 through 4 normalized by the yield stress. Data from the 100%extensometer.

3.3.7 Large Strain Behavior

As mentioned above, the onset of necking and the resulting inhomogeneous, localized plastic deformation means that the stress and strain plotted in Figure 49 are not valid after the curve begins to decrease. The strain in Figure 49 is derived from the extensometer reading under the assumption that the extension is homogeneous, and the current area is also derived from the extensometer reading, using plastic incompressibility. That current area is used to calculate the stress shown in Figure 49.

The data could be extended, however, if the cross–sectional area of the specimen could be measured independently of the extensometer and load records. To that end, photographs were taken of specimens 1, 2 and 3 to record visually the reduction in

area after necking began. A mirror was placed at the side of the specimen, angled at 45°, to reflect a view of the side of the specimen toward the camera. In this way, both a front view and a side view of the specimen could be photographed simultaneously. A photograph was taken at zero load to record the initial area. Additional photographs were taken after about 10% extension at intervals of about 4% extension up to failure.

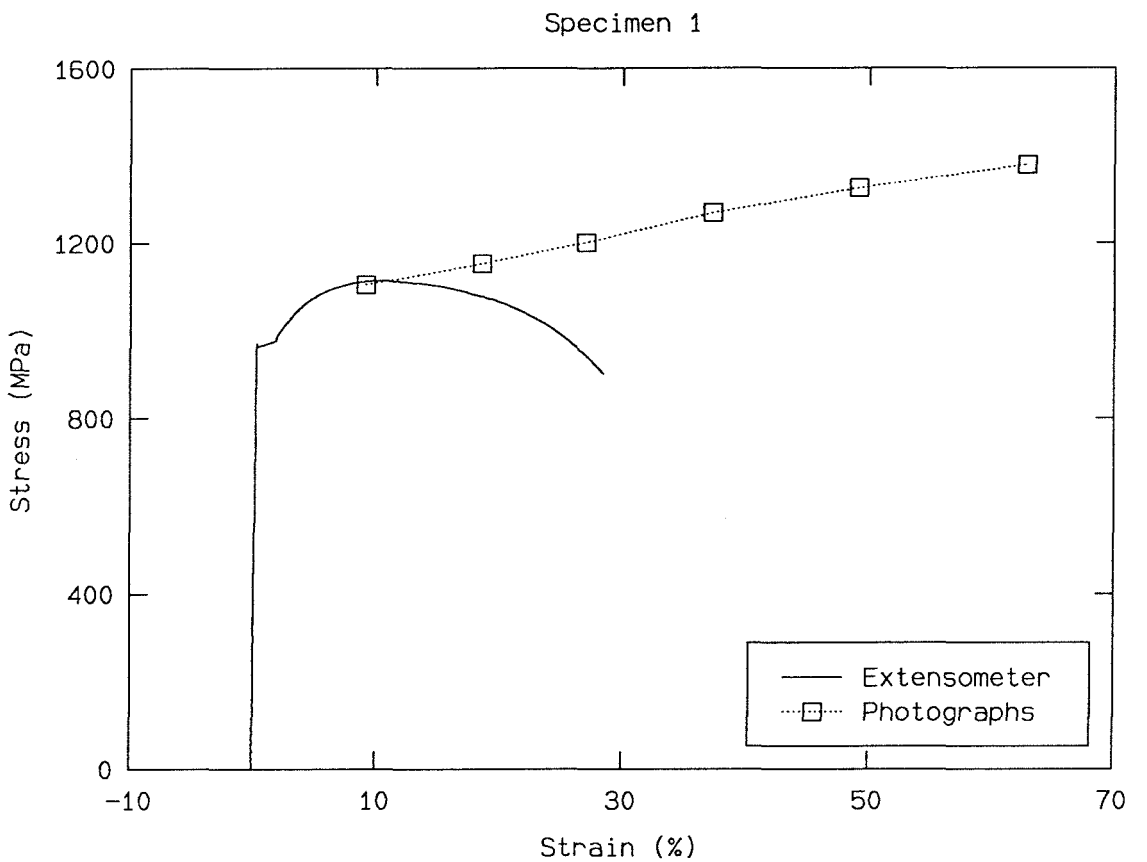


FIGURE 56. The stress–strain data for specimen 1, including the data calculated from the photographs.

The data reduction is somewhat similar to that described above, although in this case the area is the independent variable, and the strain is the dependent variable. The initial area, A_0, is measured from the first of the series of photographs, and the current area, A, is determined from each following photograph. The proper scaling of the area to correctly determine the stress is found by relating the actual

initial area of each specimen to the initial area determined from the photographs. (Some correction is needed because of the slightly different distances between the two specimen surfaces and the camera due to the angled mirror.) Again, assuming plastic incompressibility, the ratio of the initial area to the current area of the neck, A_0/A, is equal to the ratio of a current length to the corresponding initial length, L/L_0. The strain, ϵ, is given by $\epsilon = \ln(L/L_0)$, or equivalently, $\epsilon = \ln(A_0/A)$. The extension as recorded by the extensometer can be determined by measuring the distance between the arms of the extensometer, and the corresponding load can be determined from the computer file in which the load and extension have been stored. That load, P, is then used along with the current area, A, to find the stress, $\sigma = P/A$.

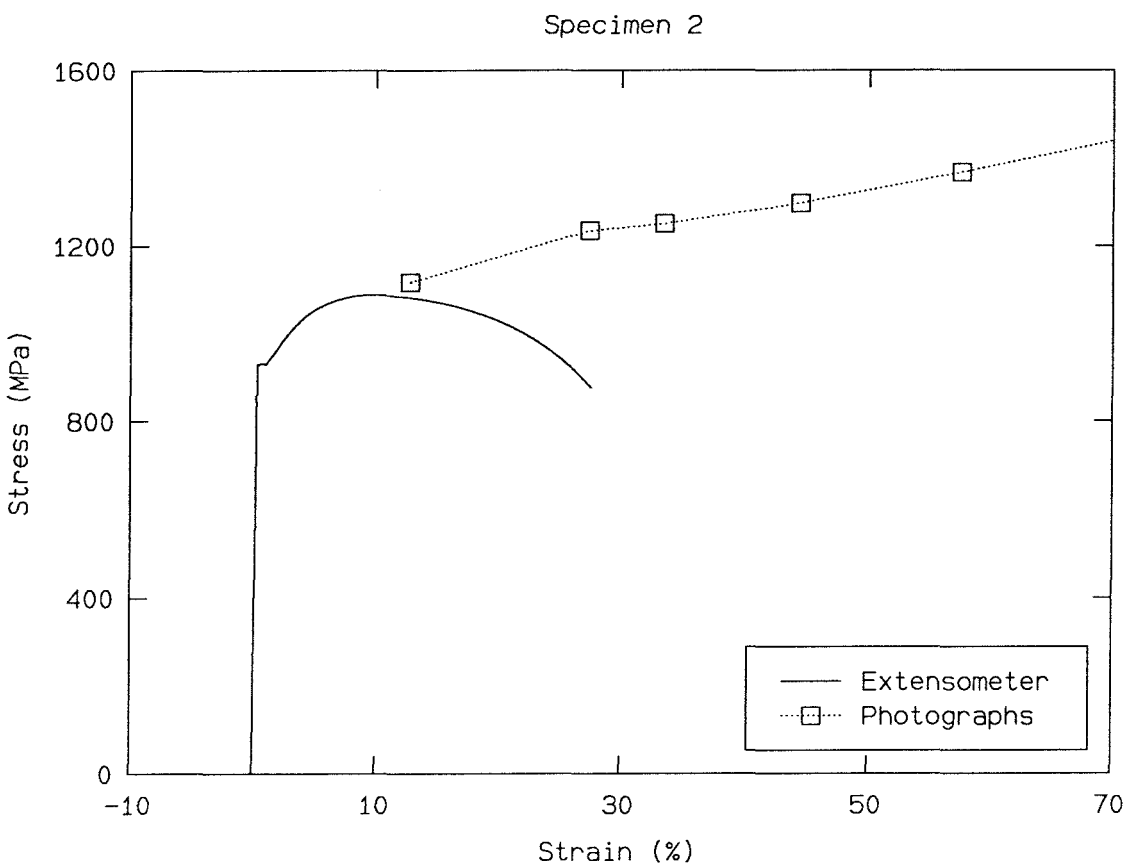


FIGURE 57. The stress–strain data for specimen 2, including the data calculated from the photographs.

The resulting stress–strain data as determined from the photographs are shown in Figures 56 through 59. It can be seen that the stress and strain do actually increase after necking, and that the stress–strain law after necking is nearly a straight line.

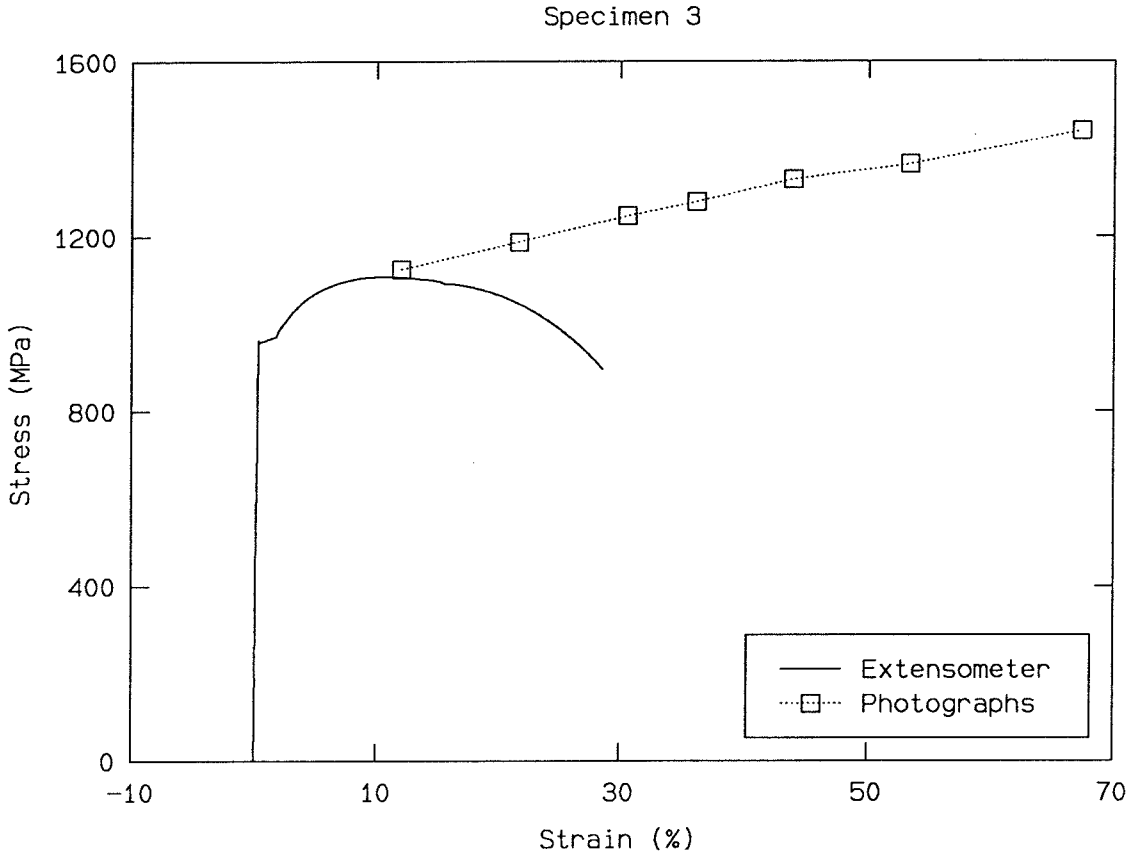


FIGURE 58. The stress–strain data for specimen 3, including the data calculated from the photographs.

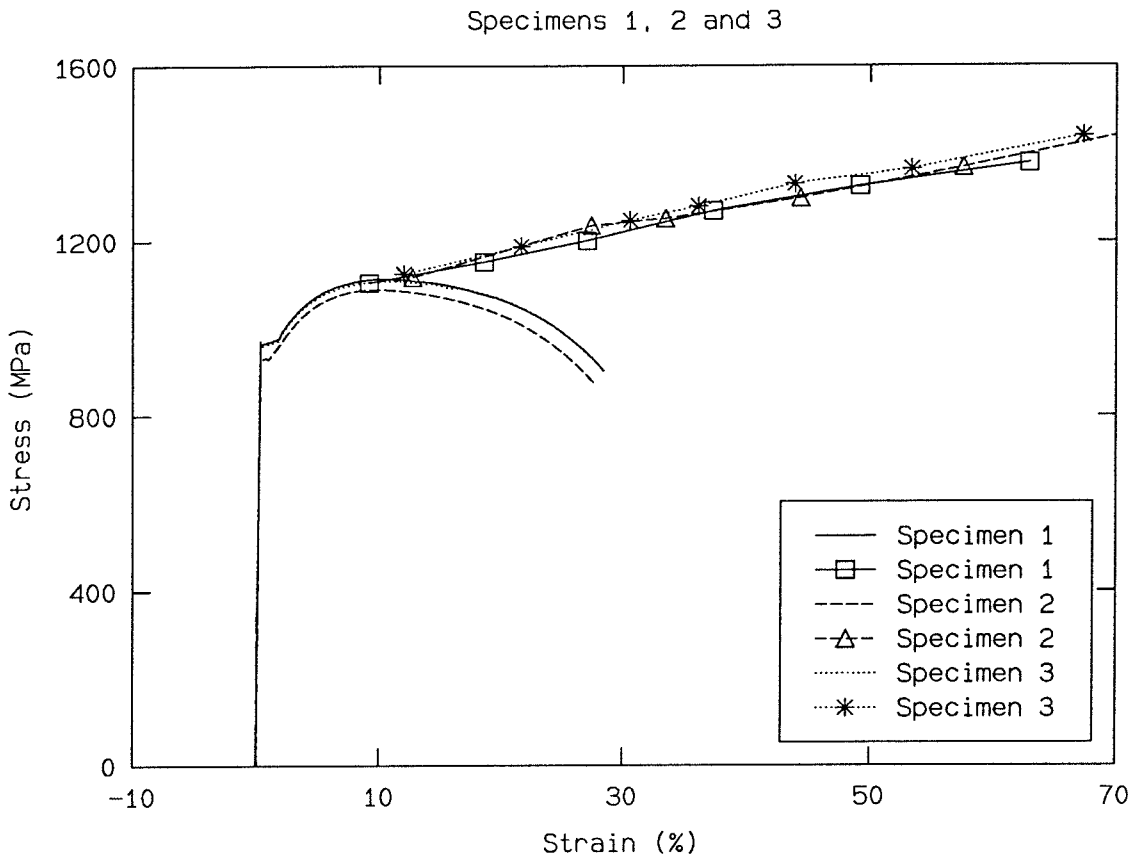


FIGURE 59. The stress–strain data for specimens 1, 2 and 3, including the data calculated from the photographs.

3.3.8 Further Correction for Stress Triaxiality

As suggested by Hertzberg [26], and by Polakowski and Ripling [58], there must be a further correction to the stress across the neck that is due to three–dimensionality of the stress distribution. In the un-necked specimen, the stress and strain states are homogeneous along the length of the test section. Since the sides of the specimen are unloaded, there are no stresses transverse to the loading axis. After necking, transverse stresses do arise, owing to the change in the geometry of the specimen. The transverse stresses must be zero at the free surfaces of the specimen, and they reach a maximum at the center of the specimen. These transverse stresses act to restrain the *axial* deformation, so that the axial stresses must also be higher in the center of the specimen than at the free surfaces. (Note that this is the opposite

of the expected elastic stress distribution for a notch, where the axial stresses are higher at the free surfaces than in the interior. The stress distribution in the neck is due to that the fact that all of the material is undergoing *plastic* flow.) The correction to the average stress, P/A, is based on the assumption that in the neck, the deformation is caused by plastic slip, which occurs near the free surfaces of the specimen. Thus, the stress required to induce further deformation is less than the average stress across the neck given by P/A.

A correction has been worked out by Bridgman [12,13] for specimens having a circular cross section by assuming that the shape of the neck can also be modeled as a circular arc. This correction is then found as a function of the ratio between the radius of the specimen at the neck, a, to the radius of curvature of the circular arc describing the shape of the neck, R.

Bridgman [13] also gives a similar correction for a two–dimensional specimen geometry (plane strain), which also depends on the ratio of the neck width to the radius of curvature of the neck profile. Unfortunately, the specimen geometry of the current study had a nearly square cross section, so it does not match exactly the conditions for either correction (axisymmetric or plane strain). Perhaps the best way to apply a correction would be a weighted average of the corrections for circular specimens and two–dimensional specimens. Since the cross section of the specimens used was nearly square, the correction for circular geometries should be somewhat better than that for the two–dimensional geometry, as the two–dimensional correction assumes that one of the thickness dimensions is infinite. The two–dimensional correction factor is slightly lower than the correction factor for circular geometries, about 1% smaller for shallow necks ($a/R = 0.1$) to about 8% smaller for deep necks ($a/R = 2.0$).

Bridgman also gives an empirical curve relating the correction factor to the strain found as $\epsilon = \ln(A_0/A)$. This empirical relation was determined for a number of materials, including steel, and since the geometry of the uniaxial specimens in this case is not circular, this empirical correction factor was used to make a preliminary estimate of the correction to the data gathered from the photographs. The results

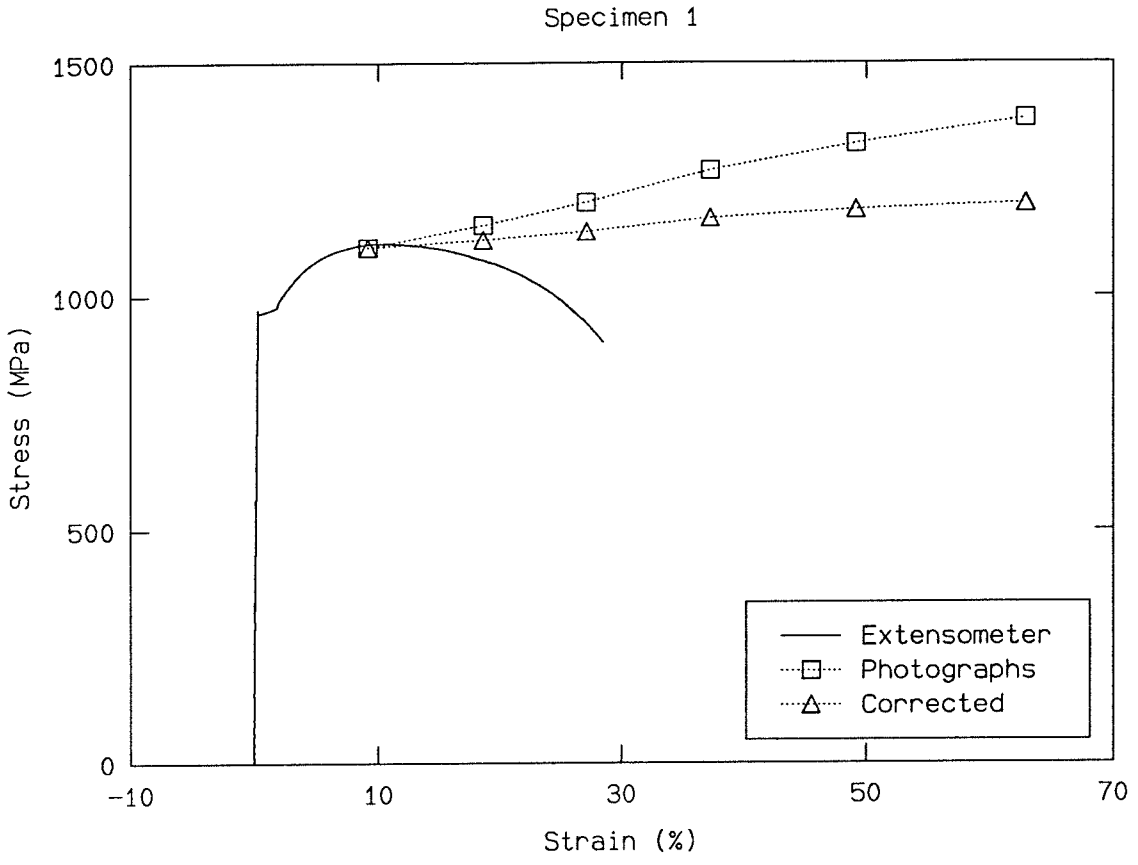


FIGURE 60. The stress–strain data for specimen 1, including the data calculated from the photographs and the empirical correction from Bridgman [12].

of using this empirical correction are shown in Figures 60 through 63.

From Figures 59 and 63, it can be seen that the stress–strain curves for all three specimens match fairly closely at high strains, on the basis of the results from the photographs.

The corrected data are shown again in Figure 64, again normalized by the yield stress. It can be seen from Figures 63 and 64 that the normalization *increases* the scatter among the data at high strain levels. It appears that the large deformation of the steel has little connection with the mechanisms that induce yielding. In that case, the large strain behavior would be better compared using another criterion. Such a criterion is suggested by Figure 47, where the data are shown on logarithmic

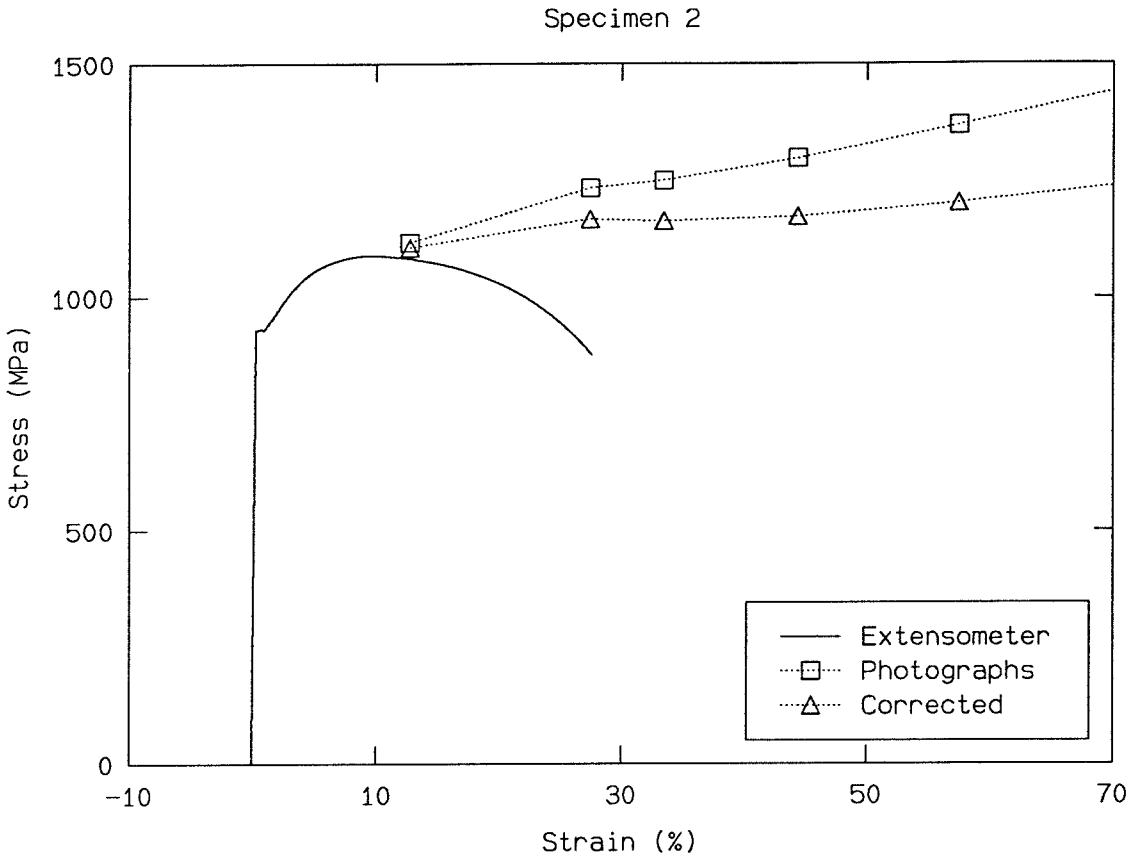


FIGURE 61. The stress–strain data for specimen 2, including the data calculated from the photographs and the empirical correction from Bridgman [12].

scales, and a fit is shown through the data after the Lüder's bands have passed through the test section. It may be better to compare the stress–strain data at large strains on the basis of such a power-law fit.

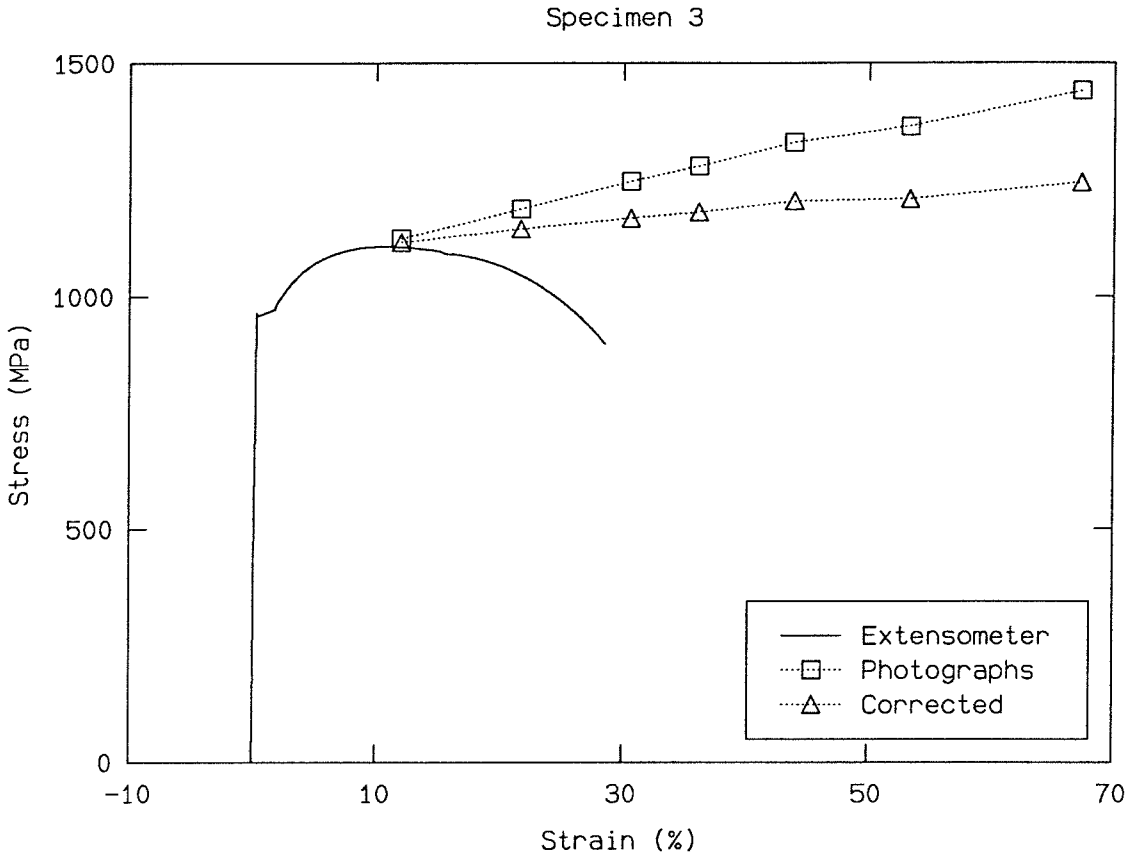


FIGURE 62. The stress–strain data for specimen 3, including the data calculated from the photographs and the empirical correction from Bridgman [12].

3.3.9 Power–Law Approximation to the Corrected Stress

In Figure 65 are shown the true and corrected stress strain data on logarithmic scales. It can be seen that neither the "true" stress–strain curves nor the empirically corrected stress–strain curves follow exactly the power–law hardening part of the curves. It can also be seen that the power–law hardening part of the curve for specimen 2 is very nearly parallel to the power–law hardening part of the curves for specimens 1 and 3. Again, it may make more sense to compare different specimens on the basis of the hardening exponent of the plastically deforming region than on the yield stress and strain at yield.

It still remains to pick an appropriate hardening exponent for input to the

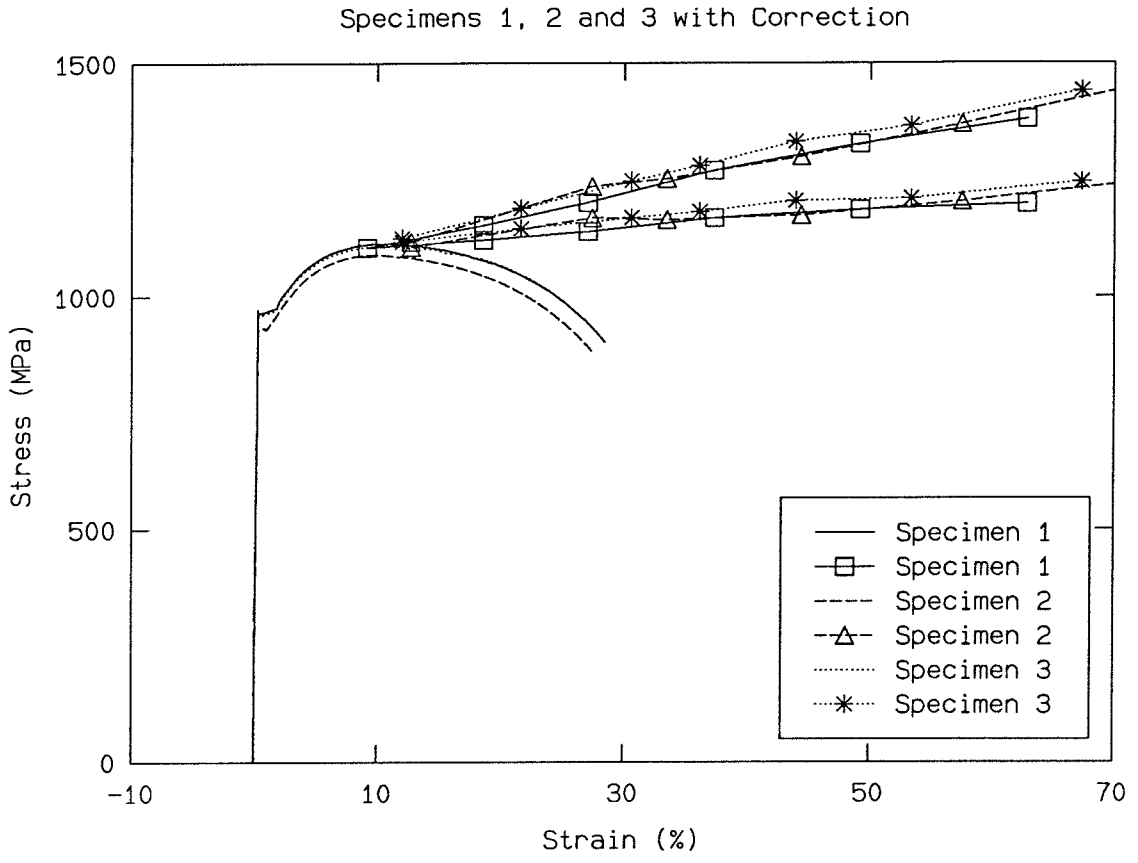


FIGURE 63. The stress–strain data for specimens 1, 2 and 3, including the data calculated from the photographs and the empirical correction from Bridgman [12].

finite–element code for comparison with the three–point–bend test. Figure 66 shows the true and corrected stress data for specimens 1 and 3, the uniaxial specimens made from the steel used in the three–point–bend test.

Figure 67 shows the true and corrected stress on a logarithmic scale with the power law determined through the least–squares fit of the data below 8%. This is the fit shown in Figures 51 and 50, with the hardening exponent of $n = 23.7$. Figure 69 shows the same data on a linear scale. It can be seen that this power law slightly underestimates the corrected stress at the higher strain levels, indicating that the hardening exponent is slightly too large.

Figure 68 shows the experimental data on a logarithmic scale with a curve rep-

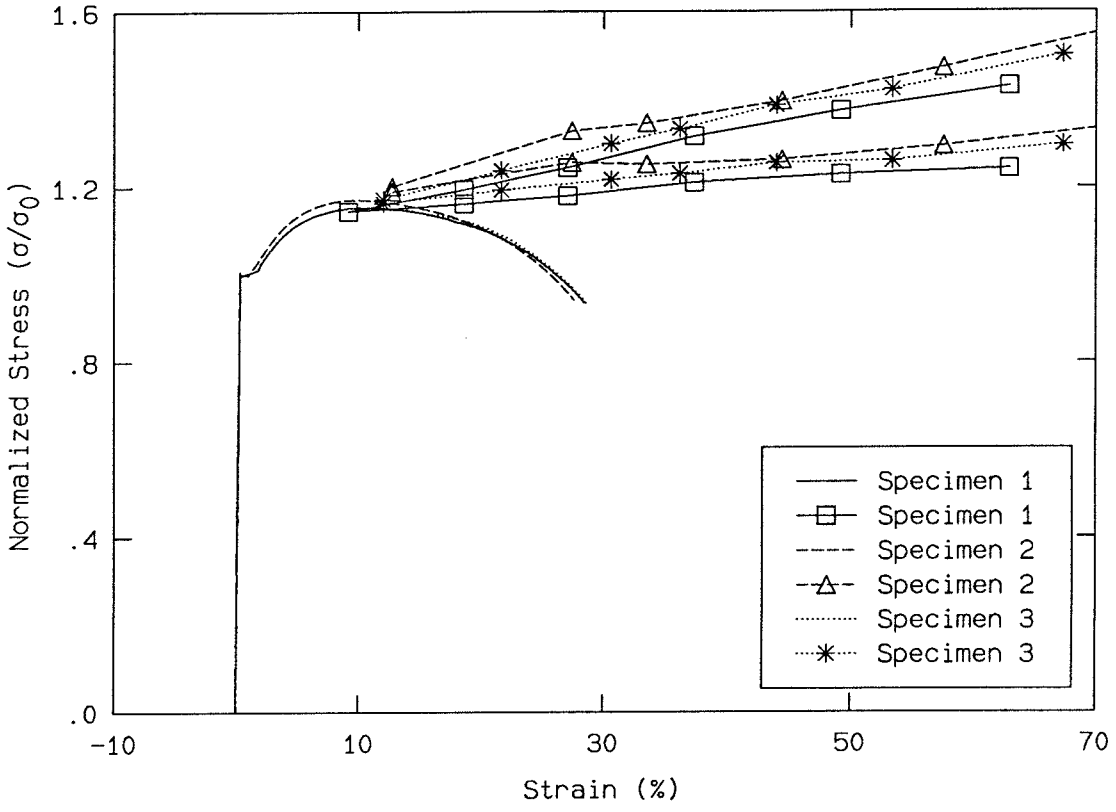


FIGURE 64. Normalized stress–strain data for specimens 1, 2 and 3, including the data calculated from the photographs and the empirical correction from Bridgman [12].

resenting a power law using a hardening exponent of $n = 20.0$, to show the variation of the power–law formulation with a small change in the hardening exponent. This power law is also shown in Figure 70 on a linear scale.

One additional factor influences the choice of the hardening exponent. The application of the power–law formulation for the stress–strain relation was intended to make a comparison between the three–dimensional, finite–element code with the two-dimensional, asymptotic HRR field. Data for the HRR field are available from Shih [77] for hardening exponents of 20 and 30, but not for exponents in between. The data in Figures 68 and 70 indicate that a choice of a hardening exponent of 20 might be best in order to make an exact comparison with the HRR field.

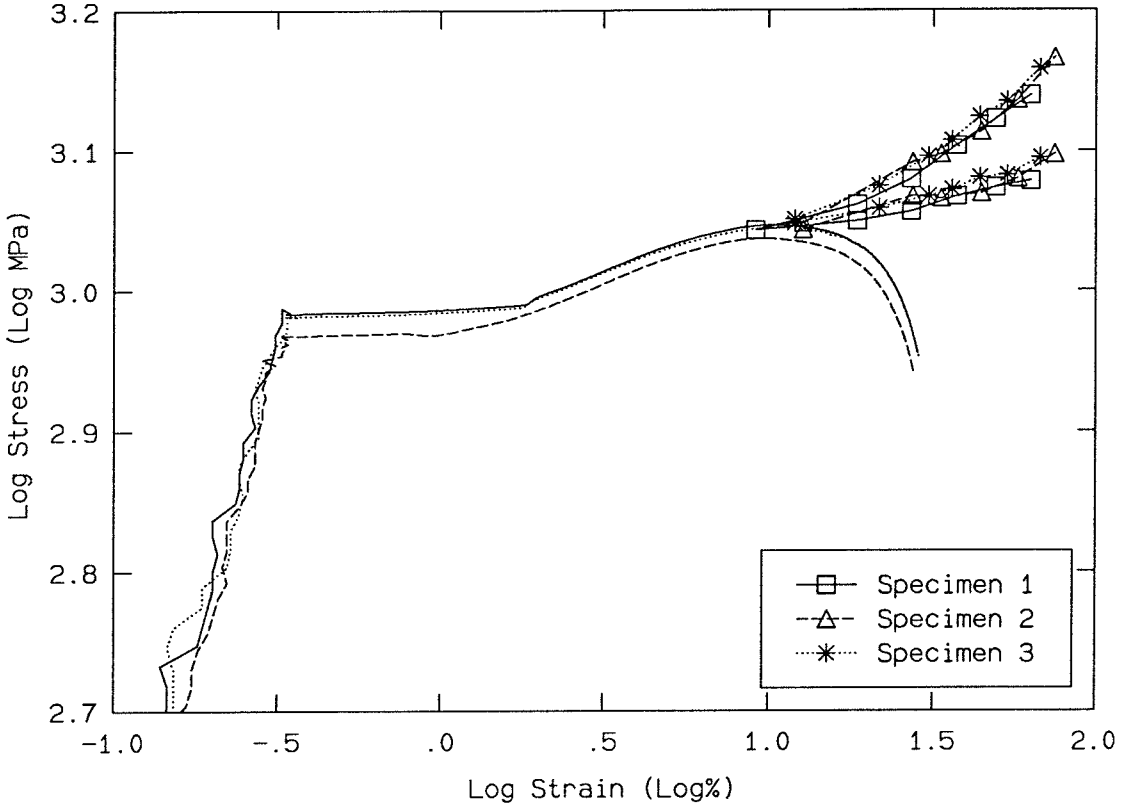


FIGURE 65. The stress–strain data for specimens 1, 2 and 3, including the data calculated from the photographs and the empirical correction from Bridgman [12], shown on logarithmic scales.

After consultation with Dr. Knauss, and with Dr. Hutchinson when he visited, a hardening exponent of 20 was chosen so as to make the closest possible comparison with the HRR field data from Shih [77].

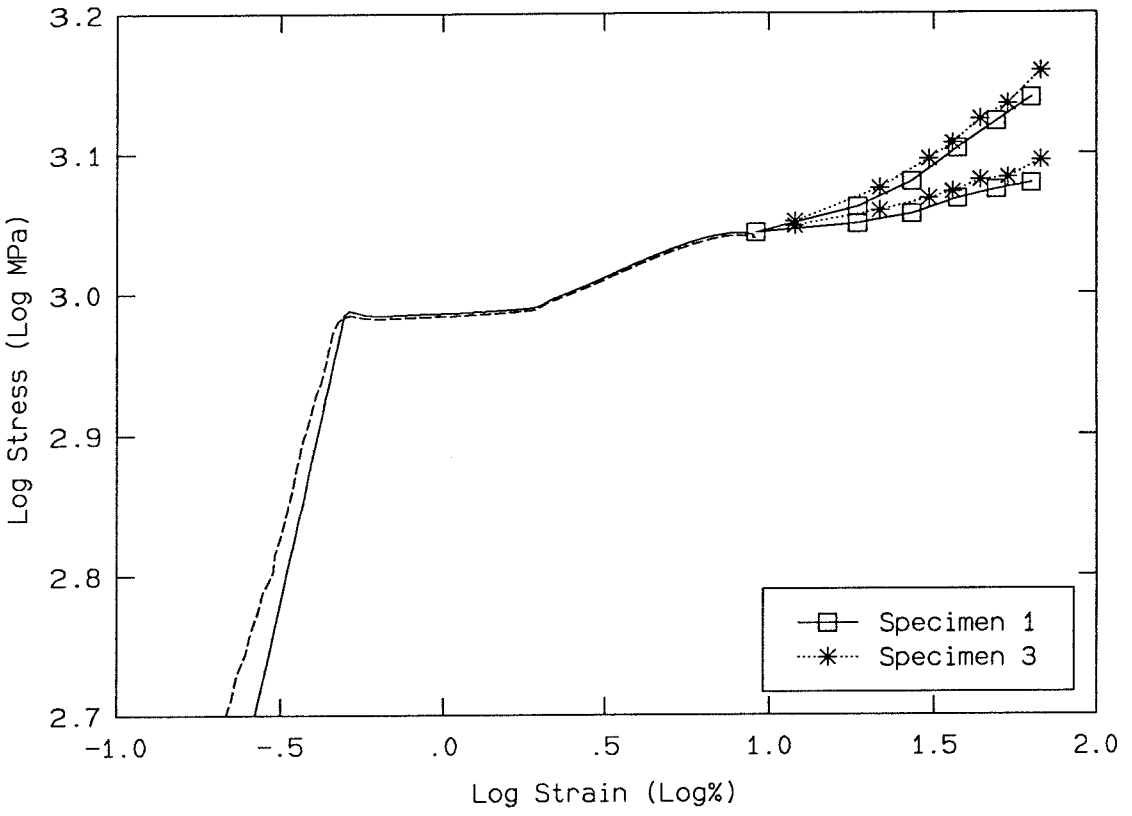


FIGURE 66. The stress–strain data for specimens 1 and 3, including the data calculated from the photographs and the empirical correction from Bridgman [12], shown on logarithmic scales.

3.3.10 Discussion

The power–law formulation used for the code does not model the uniaxial test data well in the region near the yield point, no matter what hardening exponent is used. It appears that at the high strains determined from the photographs, the power law does follow the corrected stress curve rather well. However, the use of such a power law would overestimate the work needed to induce plastic deformation, as it overestimates the stress required for plastic flow near the yield point.

It would be possible to use the data from the uniaxial test directly as input to the finite–element calculation through a tabulation of the appropriate stress for a given strain. The use of a power–law representation simplifies the calculations and

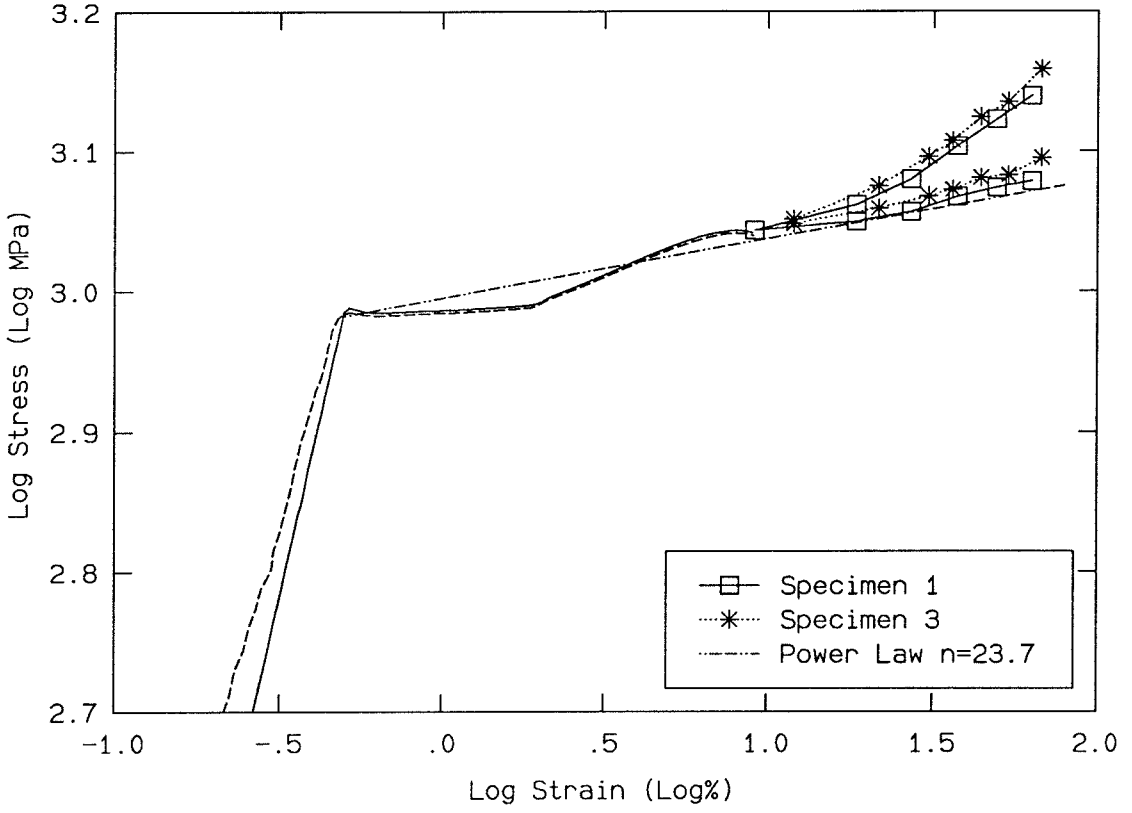


FIGURE 67. The stress–strain data for specimens 1 and 3, including the data calculated from the photographs and the empirical correction from Bridgman [12], shown on logarithmic scales. Also shown is the power law that has been least–squares fit through the yield stress and strain with a hardening exponent of $n = 23.7$.

provides a material constitutive law that should match closely with the HRR field, assuming that the HRR field has some region of dominance in this three–dimensional model.

In trying to examine the constitutive behavior at large extensions, one must be cautious about the formulation of the stress and strain tensor, and about relating the deformed configuration to the undeformed reference configuration. The finite–element model used in this study uses a linearized strain formulation, so that the deformed configuration and the reference configuration are assumed to be identical. The strains in the vicinity of the crack tip as calculated by the finite–element model

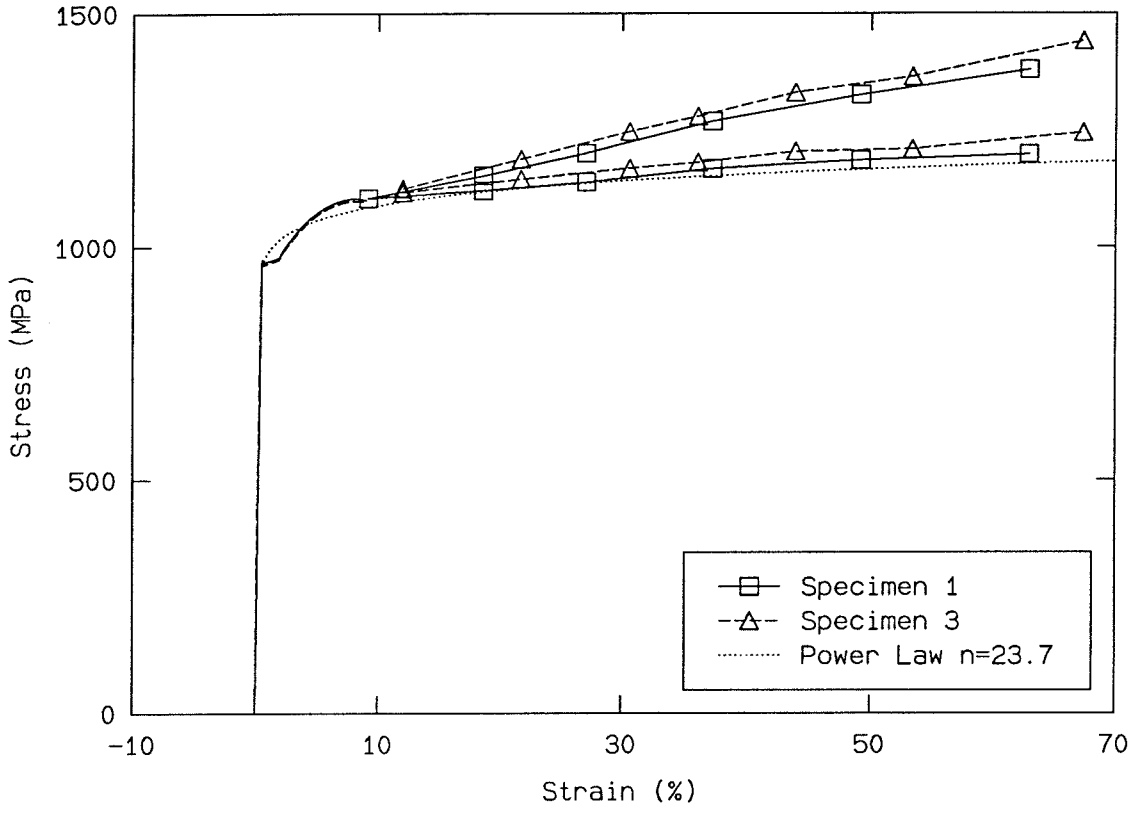


FIGURE 69. The stress–strain data for specimens 1 and 3, including the data calculated from the photographs and the empirical correction from Bridgman [12], shown on linear scales. Also shown is the power law that has been least–squares fit through the yield stress and strain with a hardening exponent of $n = 23.7$.

lie outside the region of validity of the linearized–strain theory. The measurement of the uniaxial stress and strain through the photographic record can be used in the formulation of a finite–strain, finite–element model, if desired. Any discrepancies found between the experiment and the finite–element model close to the crack tip are likely influenced by the inability of the finite–element constitutive law to model the large strains near the crack tip correctly.

The discrepancy between the numerical stress–strain model and the measured stress–strain behavior is largest at the onset of plastic deformation. As a result, the numerical model and the experiment should show the most difference at the

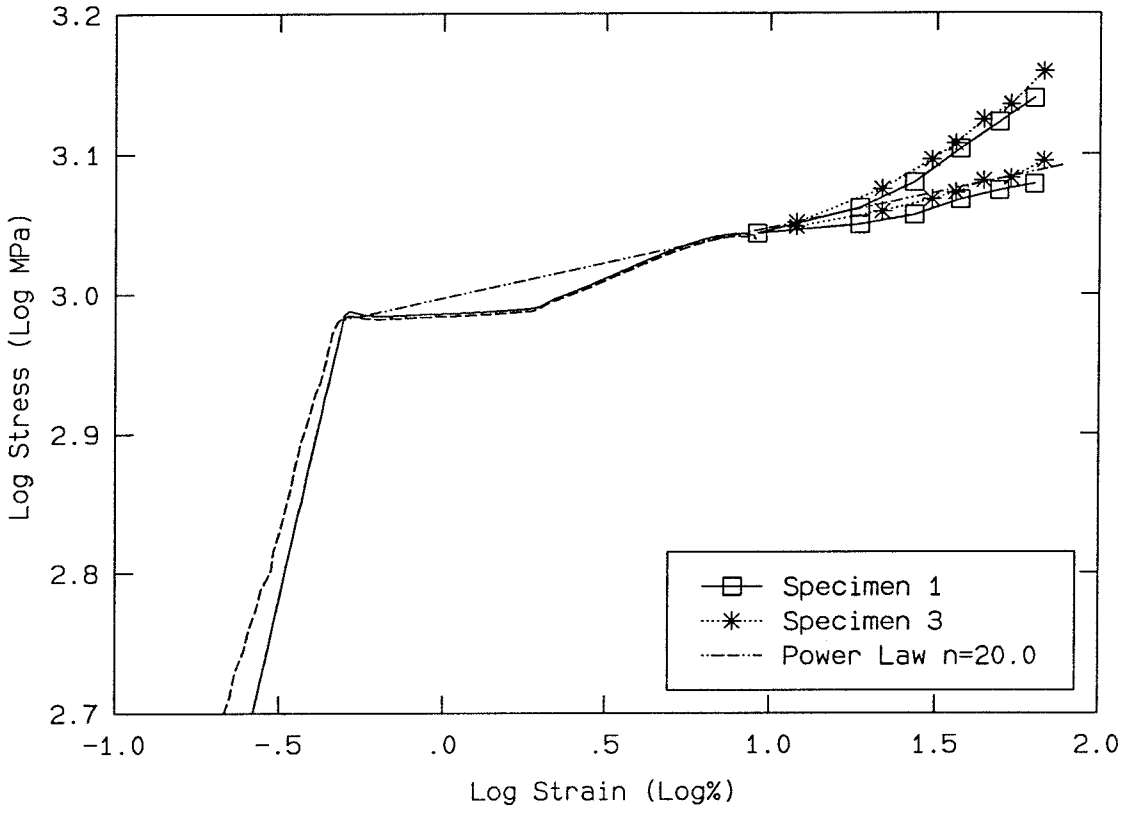


FIGURE 68. The stress–strain data for specimens 1 and 3, including the data calculated from the photographs and the empirical correction from Bridgman [12], shown on logarithmic scales. Also shown is a power law passing through the yield stress and strain with a hardening exponent of $n = 20.0$.

boundary of the plastic zone, if the discrepancy in the uniaxial law is significant in the three–dimensional field surrounding the crack tip.

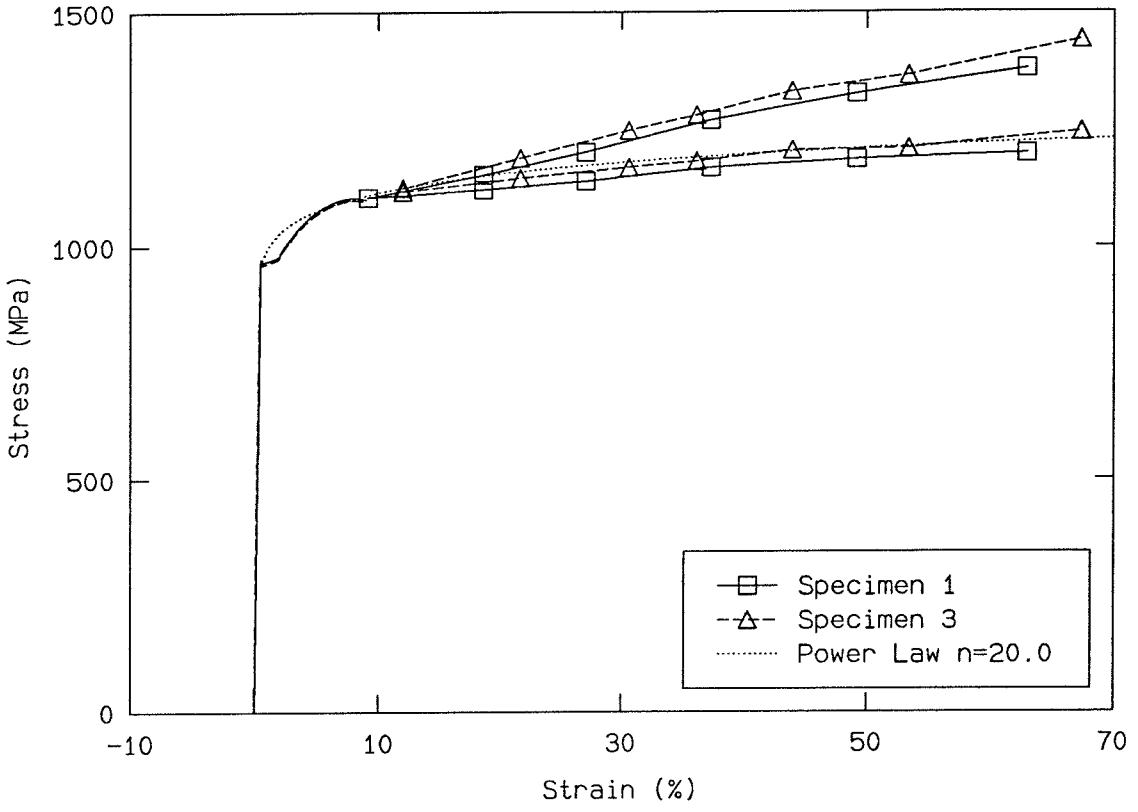


FIGURE 70. The stress–strain data for specimens 1 and 3, including the data calculated from the photographs and the empirical correction from Bridgman [12], shown on linear scales. Also shown is a power law passing through the yield stress and strain with a hardening exponent of $n = 20.0$.

3.4 Hardness Testing

The tunneling tests described in Section 4.5 were carried out on 4340 steel specimens that were left over from an earlier investigation. Since it was necessary to conserve steel for the tunneling specimens, there was not enough steel remaining to make uniaxial test specimens. Therefore, the material properties (Young's modulus, yield stress and strain) of the tunneling specimens could not be determined as accurately as the properties of the steel specimens used in the interferometric investigation. The steel was heat–treated in a manner identical to that described in Section 3.3, so as to match the properties as closely as possible. However, only six specimens were available, and it was desired to use all six to determine as closely as possible the tunneling behavior at different loads. Thus, it proved impractical to measure the material properties in a uniaxial test similar to those described in Section 3.3. As mentioned in Section 3.3.5, it was found that the uniaxial data of the current investigation could be compared to the uniaxial data of the investigation by Zehnder [91], if the constitutive behavior of the two measurements were scaled by the yield stress. This observation suggested that perhaps all of the important material properties of the tunneling specimens could be extrapolated from the properties measured in the uniaxial tests for a different batch of steel, if only the yield stress of each specimen were known.

An estimate of the yield stress of a material can be made by testing its surface hardness. The hardness is measured by pushing an indenter into the surface of a specimen with a known force and measuring the cross–sectional area of the resulting depression. As described by Ashby and Jones, [2], the ratio of the known force to the projected cross–sectional area of the indent is proportional to the yield stress of the material. The hardness of the specimens to be used for the tunneling tests were measured and compared with the hardness of the uniaxial specimens used to determine the properties of the fracture specimens in the interferometric investigation. The hardness was also compared to the behavior of 4340 steel found in the Metals Handbook [46].

The hardness was tested using a Wilson Mechanical Instrument Company, Inc.

Rockwell C Hardness and Yield Stress for 4340 Steel		
Tempering Temperature $^{\circ}C$	Hardness Rockwell C	Yield Stress MPa
205	53	1860
315	49.5	1620
425	46	1365
540	39	1160
650	31	860
705	24	740

Table 2. Variation of Rockwell C Hardness and Yield Stress with Tempering Temperature for 4340 Steel. Data from the Metals Handbook [46].

Rockwell Hardness Tester. "Rockwell" is a brand name that describes an indentation hardness test following the guidelines described in ASTM Standard 18–79 [4]. The Rockwell test proceeds as follows. The indenter is brought into contact with the specimen and loaded with a small load (98 N). The position of the indenter is recorded. An additional large load is then applied to the indenter. In this case, using the Rockwell C scale, the large load is 1373 N.* The large load is then removed, and the position of the indenter is again recorded. The increment of displacement of the indenter, e, is measured in units of 0.002 mm. The Rockwell C hardness number is then determined as $100 - e$. Actually, the Rockwell C number is read directly from a dial on the testing machine, but knowing the relation of the hardness number to the indenter depth makes it possible to translate the hardness number into a relative yield stress, if we assume that the indenter shape is self–similar.

Assuming that the indenter shape is self–similar, the cross–sectional area of the indent is proportional to the square of the indenter depth. Making this assumption, a plot of yield stress versus the quantity $(100 - \text{Rockwell C})^{-2}$ should result in a

* Different Rockwell scales use a different large load so as to be able to measure the hardness of a variety of materials. The measurable travel of the indenter is limited, so by using a smaller load, a softer material can be tested and the indenter travel kept within its instrumented range.

Rockwell C Hardness and Yield Stress for 4340 Steel		
Tempering Temperature $^{\circ}C$	Mean Hardness Rockwell C	Yield Stress MPa
540	33.7	965
540	33.2	930
540	34.0	960
540	33.1	945

Table 3. Mean Rockwell C Hardness and Yield Stress for four Uniaxial Test Specimens of 4340 Steel.

straight line.

Rockwell C values and yield stresses for 4340 steel as tabulated in the Metals Handbook [46] are shown in Table 2, and Table 3 shows the mean Rockwell C hardness values and the yield stresses measured for four uniaxial specimens of 4340 steel measured in the course of the interferometric investigation. Four hardness readings were taken for each uniaxial specimen. The uncertainty in the hardness measurement is about ± 1 Rockwell C hardness number. The values from Tables 2 and 3 are plotted in Figure 71. A plot of the yield stress versus the quantity $(100 - \text{Rockwell C})^{-2}$ is shown in Figure 72. A least–squares fit of the data from the Metals Handbook is also shown in Figure 72.

The mean Rockwell C values for three of the tunnel test specimens are shown in Table 4 with the yield stresses as determined from the least–squares fit of the Metals Handbook data. Five hardness readings were taken for each tunnel–test specimen. The uncertainty in the hardness measurement is about ± 1 Rockwell C hardness number. Only three of the tunnel–test specimens were flat enough after heat treating to be ground into suitable specimens for the hardness test. It is assumed that they are representative. From these data, it seems a good estimate of the yield stresses of the tunneling specimens would be the mean value of the yield stresses found using the line fit through the Handbook data, which is 900 MPa.

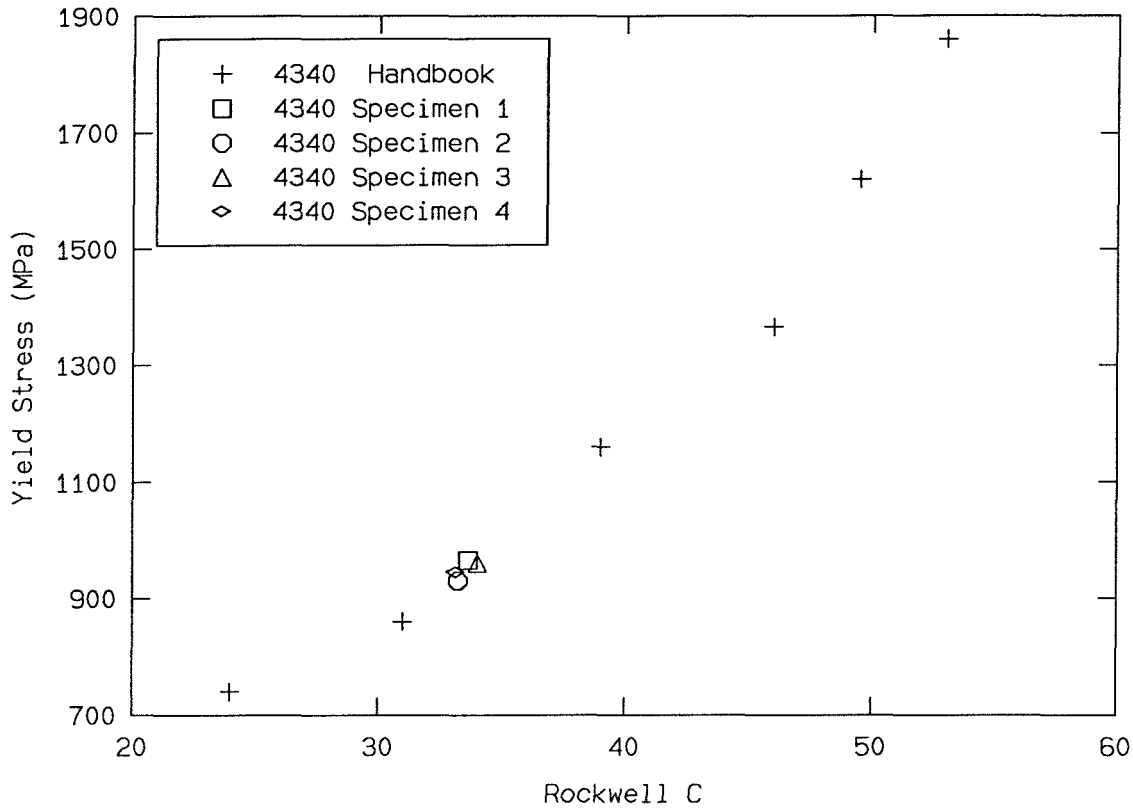


FIGURE 71. Correlation of yield stress with Rockwell C hardness value for 4340 steel.

This estimate of the yield stress is within 10% of the 960 MPa measured for the steel used in the study with the moire interferometer. It will be assumed that the steel of the tunnel–test specimens and the steel used in the interferometric measurement are similar enough that the measured tunnels are exactly what occurred in the interferometric study. Of course, the yield–stress comparison may not provide much information about the tunneling in the steel. Tunneling is the result of failure of the material in the interior of the body, which is that part of the stress–strain curve farthest from the yield point. There is no reason to suspect that the yield–stress comparison would be a good predictor of the failure behavior of a material. However, the comparison indicates that the two materials are similar, at least at yielding. It was expected that the comparison would be close. A comparison that

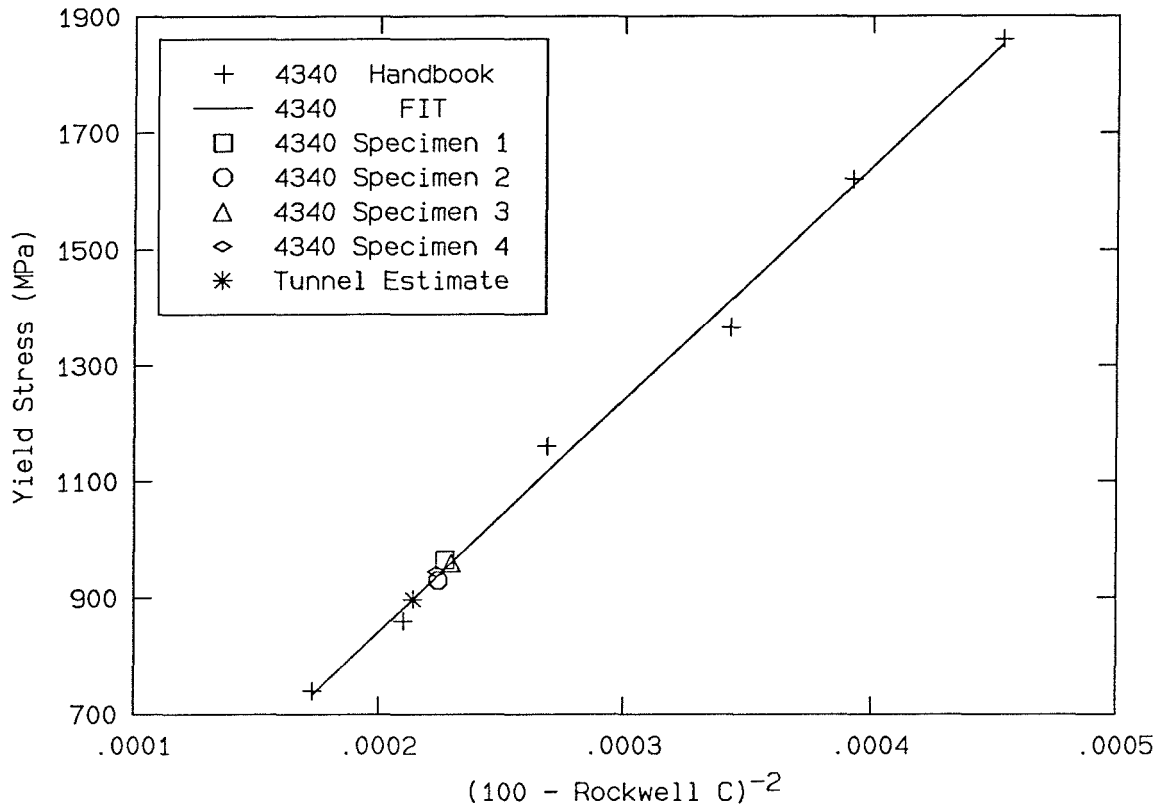


FIGURE 72. Correlation of yield stress with the quantity $(100 - \text{Rockwell C})^{-2}$ for 4340 steel. The solid line indicates a least–squares fit through the Handbook data.

was not close would have served to eliminate that specimen from the test.

On the other hand, the data from the hardness tests seem to indicate that the tunneling specimens fall outside the scatter expected from the data taken from the uniaxial specimens. However, there may be other circumstances that have an effect. The hardness test is a surface measurement, so the surface preparation has some effect on the measured value. Also, the hardness data for the uniaxial test specimens were taken after they had been pulled in two, so there may have been some strain hardening of the material (the hardness tests were taken in the corners of the specimens, where loading during the test should have been at a minimum).

Rockwell C Hardness and Yield Stress for 4340 Steel		
Tempering Temperature $^{\circ}C$	Mean Hardness Rockwell C	Yield Stress MPa
540	32.2	910 (fit)
540	31.1	885 (fit)
540	31.6	895 (fit)

Table 4. Mean Rockwell C Hardness and estimated yield stress for three tunnel–test specimens of 4340 Steel. The yield stress is estimated using the fit through the data in Figure 72.

3.5 Poisson's Ratio Experiments

One important possible source of error in comparing the experiment and the numerical model is the choice of Poisson's ratio. Poisson's ratio was chosen as 0.3 in the numerical simulation described in Narasimhan and Rosakis [51] and in Narasimhan, Rosakis and Moran [52]. Published values for 4340 steel are as low as 0.28, a discrepancy of about 7%. The out-of-plane deformation should vary linearly with Poisson's ratio, assuming that the body is well approximated using plane stress, so an error in Poisson's ratio of 7% in the numerical calculation would correspond to an error in the out-of-plane deformation of about 7%. The dependence of the in-plane deformation on Poisson's ratio is more complicated. From the two–dimensional, asymptotic, linearly elastic solution (discussed in Section 4.2.3), a deviation in Poisson's ratio of 7% can lead to an change in the calculated u_1 or u_2 displacements of from about 1% to 10%, depending on the angle from the line of the crack, and on whether plane stress or plane strain is assumed.

Comparison of the out-of-plane deformations measured using the Twyman–Green interferometer [56] with the numerical calculations using a Poisson's ratio of 0.3 indicated that this choice of Poisson's ratio is too large. From these measurements, it was estimated that Poisson's ratio should be about 0.28 to make the best fit between the experiment and the numerical model.

It was thought that the moire interferometer could be used to determine Poisson's ratio with a very simple uniaxial test in the same load frame used for the three–point–bend tests.

3.5.1 Specimen Material and Dimensions

The specimen used to measure Poisson's ratio were cut from part of a three–point–bend specimen that had been used in a preliminary test. The specimen geometry is rectangular, 149.5 mm long by 36.48 mm wide by 9.52 mm thick, as shown in Figure 73(a.). The specimen was loaded in compression along the axis parallel to its longest dimension.

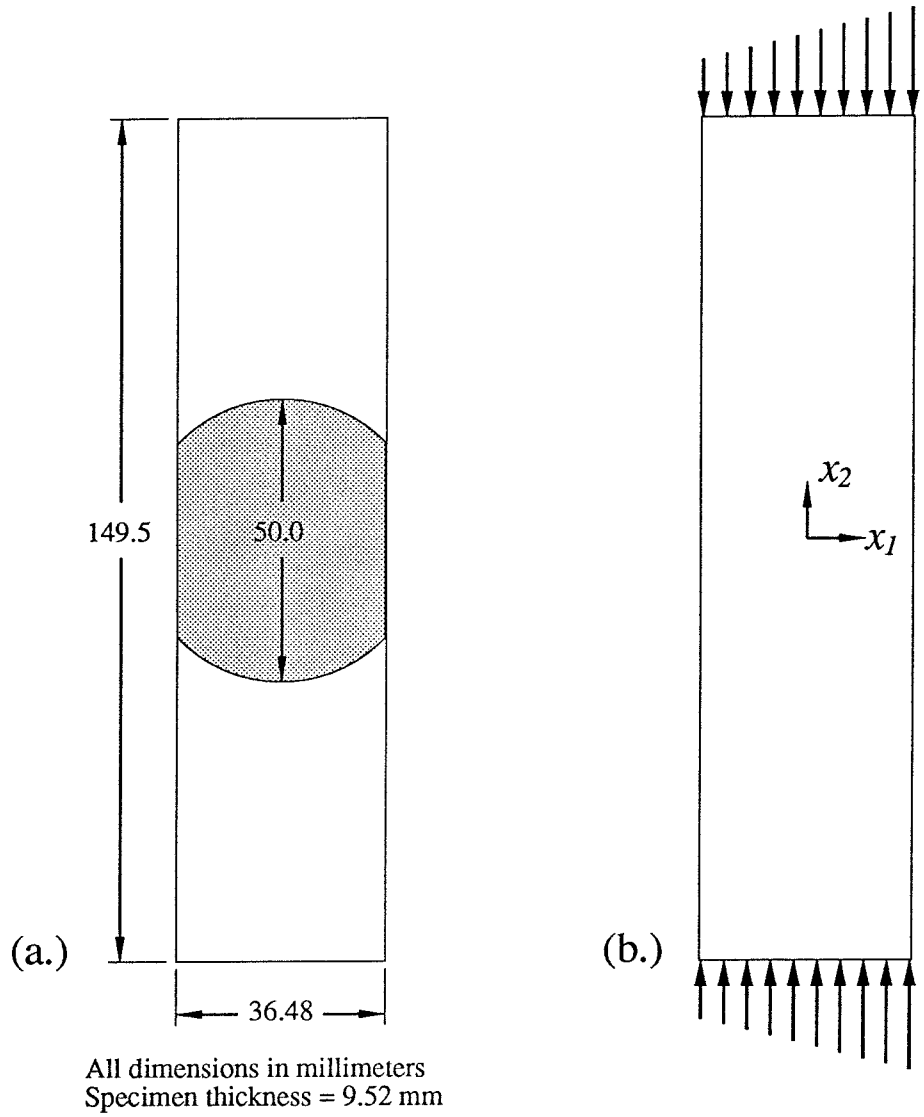


FIGURE 73. The specimen to measure Poisson's ratio. The dimensions of the specimen are shown in diagram (a.). The shaded region indicates the size of the diffraction grating applied to the specimen. Diagram (b.) indicates the assumed loading distribution and the axes used in the analysis.

A diffraction grating was attached to the center of one of the larger sides of the specimen with epoxy, as described in detail in Section 2.4.4, where the moire interferometer to measure the two in-plane displacement components is described.

3.5.2 Experiment

The experiments were carried out on a vibration–isolated optical table in order to reduce the effects of vibrations from the surroundings. The optical table is a Newport Corporation product; it measures ten feet by four feet by eighteen inches thick. The loading frame and load controller are described in more detail in Section 4.4.

From a uniaxial tension test, the yield stress has been estimated to be 960 MPa, and the strain at yield 0.5% (Section 3.3). The conditions for yield correspond to a load of about 340 kN. To ensure that the response was entirely elastic, it was intended to keep the strain below about 0.3%, corresponding to a load of 200 kN. An additional constraint was that the specimen should not buckle. The buckling load for a perfect bar of these dimensions having pinned ends is about 225 kN. (Actually, the specimen would be better described by one clamped end and one pinned end. The pinned end conditions provide a more conservative estimate of the buckling load, however.) Allowing for imperfections, the load was to have been held to about 75% of the buckling load, or 160 kN, and data taken at 8 steps separated at 20 kN increments. In fact, the oil pressure supplying the hydraulic cylinder was insufficient to achieve that load, so data were taken only up to 120 kN. This is about 0.18% strain in the axial direction, which corresponds to about 2 fringes per millimeter on the specimen for the axial displacement and about 0.6 fringe per millimeter for the transverse displacement.

In the experiment, the in-plane displacements were photographed using a Nikon FA camera body with a Nikon Micro-Nikkor 200mm f/4 lens and a Nikon 3T close-up lens (focal length 665 mm) attached to the front of the 200mm lens. This camera setup produced a magnification of about 0.6 (that is, 1 mm on the specimen corresponded to 0.6 mm on the negative). The film used was Kodak Technical Pan set at an exposure index of 120 and developed in Kodak HC-110 (dilution D) for 6 minutes. This film has an extremely fine grain size, and it *can* record fringes with wavelengths as small as 5 microns at the film plane; this corresponds to 5 microns divided by the camera magnification on the specimen.

The loading frame was designed and constructed at Caltech and has a test–section length that is adjustable by means of threaded rods. The threaded rod supports are fairly long and slender, so that they are rather susceptible to buckling. Therefore, the test had to be performed in compression, meaning that the threaded rods were loaded only in tension. The restriction that the load had to be applied in compression introduced some bending into the specimen, which had to be taken into account in reducing the data. The compression loading also introduces some rigid rotation into the specimen.

3.5.3 Data Reduction

In order to obtain an estimate of Poisson's ratio, an assumed function was fit through the displacement field data calculated from the fringes observed using the moire interferometer. The fringe data were reduced into an array of position and displacement data at discrete points using a digitizing tablet, by finding the centers of fringes on a photograph. There are two different arrays, one corresponding to the u_1 displacement, and the other corresponding to the u_2 displacement. The positions of the data points taken in the u_1 field will not in general coincide with the positions of the data points taken in the u_2 field.

The x_1 and x_2 axes will be aligned with the specimen as shown in Figure 73(b.); the x_2 axis corresponds to the direction of the applied load. It is assumed that the only nonzero stress component is σ_{22}, and that σ_{22} is a linear function of x_1 only. In particular,

$$\sigma_{22} = kx_1 + c \tag{3.5.1}$$

From this assumption, the in-plane strains can be determined as

$$\begin{aligned} \epsilon_{11} &= \frac{-\nu}{\mathrm{E}}(kx_1 + c) \\ \epsilon_{12} &= 0 \\ \epsilon_{22} &= \frac{1}{\mathrm{E}}(kx_1 + c) \end{aligned} \tag{3.5.2}$$

Integrating the first and third of equations (3.5.2), the displacements are found as

$$\begin{aligned} u_1 &= \frac{-\nu}{\mathrm{E}}\left(\frac{kx_1^2}{2} + cx_1\right) + f_1(x_2) \\ u_2 &= \frac{1}{\mathrm{E}}(kx_1 + c)x_2 + f_2(x_1) \end{aligned} \tag{3.5.3}$$

where $f_1(x_2)$ and $f_2(x_2)$ are undetermined functions of integration. From the second of equations (3.5.2), it follows that

$$\frac{\partial u_1}{\partial x_2} = -\frac{\partial u_2}{\partial x_1} \tag{3.5.4}$$

which implies that

$$f_1'(x_2) = -\frac{kx_2}{\mathrm{E}} - f_2'(x_1) \tag{3.5.5}$$

Grouping the terms involving x_2 on one side of the equation, and terms involving x_1 on the other, it is found that

$$f_1'(x_2) + \frac{kx_2}{\mathrm{E}} = -f_2'(x_1) = c_2 \tag{3.5.6}$$

where c_2 is a constant. From this equation, f_1 and f_2 are found as

$$\begin{aligned} f_1(x_2) &= -\frac{kx_2^2}{2\mathrm{E}} + c_2x_2 + c_3 \\ f_2(x_1) &= -c_2x_1 + c_4 \end{aligned} \tag{3.5.7}$$

Therefore, the displacements are given by

$$\begin{aligned} u_1 &= \frac{-\nu}{\mathrm{E}}\left(\frac{kx_1^2}{2} + cx_1\right) - \frac{kx_2^2}{2\mathrm{E}} + c_2x_2 + c_3 \\ u_2 &= \frac{1}{\mathrm{E}}(kx_1 + c)x_2 - c_2x_1 + c_4 \end{aligned} \tag{3.5.8}$$

Absorbing E into the constants k and c, equations (3.5.8) can be rewritten as

$$\begin{aligned} u_1 &= -\nu\left(\frac{k_1x_1^2}{2} + c_1x_1\right) - \frac{k_1x_2^2}{2} + c_2x_2 + c_3 \\ u_2 &= (k_1x_1 + c_1)x_2 - c_2x_1 + c_4 \end{aligned} \tag{3.5.9}$$

where $k_1 = k/\mathrm{E}$ and $c_1 = c/\mathrm{E}$.

Equation (3.5.9) contains six parameters to be fit to the displacement data. Constants c_3 and c_4 represent a rigid translation of the body, and constant c_2 represents a rigid rotation. Fitting the displacements is not a straightforward process, because the functions to be fit have a nonlinear dependence on the parameters ν, k_1 and c_1.

A good description of the procedure for fitting such a nonlinear model function is found in *Numerical Recipes* [64], along with a procedure for performing such a fit. The gist of the procedure is that it minimizes the mean-square error between the experimental data to be fit and the approximating function. The procedure had to be adapted to fit data over two dimensions, and to fit the data from the two different displacement components simultaneously.

In fact, seven parameters are fit in the model. An arbitrary rigid rotation is allowed between the two fields, so that the constant c_2 in the second of equations (3.5.9) is replaced by a constant $(c_2 + c_5)$. This additional constant, c_5, is introduced to account for two possible discrepancies that exist in the model. The first possible discrepancy is a lack of orthogonality between the gratings, which is estimated to be within ± 7 minutes of arc. The second possible discrepancy is that the alignment of the interferometers might change slightly. Each interferometer is an independent unit, so the constant c_5 also allows for any independent drift of the two interferometers.

Let the vector $\mathbf{a}$ consist of the parameters to be fit, $\mathbf{a} = \{\nu, k_1, c_1, c_2, c_3, c_4, c_5\}$. The fitting procedure consists of choosing the values in $\mathbf{a}$ in such a way that the χ^2 merit function is minimized. This function is defined as

$$\chi^2(\mathbf{a}) = \sum_{i=1}^{n_1} \left[\frac{u_1^{(i)} - \widehat{u}_1(\mathbf{x}^{(i)}, \mathbf{a})}{\sigma_i} \right]^2 + \sum_{j=1}^{n_2} \left[\frac{u_2^{(j)} - \widehat{u}_2(\mathbf{x}^{(j)}, \mathbf{a})}{\sigma_j} \right]^2 \qquad (3.5.10)$$

where $u_1^{(i)}$ is the measured value of u_1 at the i^{th} point in the u_1 displacement field photograph, taken at position $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)})$. The function $\widehat{u}_1(\mathbf{x}^{(i)}, \mathbf{a})$ is the calculated value of the displacement u_1 at position $\mathbf{x}^{(i)}$ from equation (3.5.9), using the parameters in $\mathbf{a}$. σ_i is the estimated mean-square random error in the measured value of $u_1^{(i)}$, in this case all of the σ_i were taken to be the same, about one tenth

of the displacement increment between adjacent black fringes. Similarly, $u_2^{(j)}$ is the measured value of u_2 at the j^{th} point in the u_2 displacement field photograph, taken at position $\mathbf{x}^{(j)} = (x_1^{(j)}, x_2^{(j)})$. The function $\widehat{u}_2(\mathbf{x}^{(j)}, \mathbf{a})$ is the calculated value of the displacement u_2 at position $\mathbf{x}^{(j)}$ from equation (3.5.9), using the parameters in $\mathbf{a}$. The estimated mean–square error in each measurement of $u_2^{(j)}$, σ_j, was also taken as one tenth of the displacement increment between fringes.

Since the dependence on the parameters in $\mathbf{a}$ is nonlinear, the minimizing procedure was carried out iteratively. However, fewer than five iterations were needed to achieve convergence for each of the five cases studied (indicating that the minimizing procedure is good, but not necessarily that the assumed function is a good model of the experimental data).

One estimate of the goodness of fit of the assumed function to the data is to compare the vale of the minimized χ^2 merit function to the number of data points. If the estimate of the mean–square random error, σ_i, is a good choice, each of the terms in the summations for χ^2 should be near unity, so that the value of χ^2 should be approximately equal to the number of data points, n, if that number is fairly large. Alternatively, if the random error is unknown, it can be estimated by taking each $\sigma_i = 1$ in the minimization of χ^2, and then the mean–square random error can be estimated as $\sigma_i^2 = \chi^2/n$.

3.5.4 Results

The data were fit to the assumed function at the three highest load levels, 80 kN, 100 kN and 120 kN on loading the specimen, and again at 100 kN and 80 kN on unloading. There were 301 points digitized in the u_1 field and 209 points digitized in the u_2 field at 80 kN (loading); 302 u_1 points and 262 u_2 points digitized at 100 kN (loading); 308 u_1 points and 315 u_2 points digitized at 120 kN; 311 u_1 points and 260 u_2 points digitized at 100 kN (unloading); and 313 u_1 points and 230 u_2 points digitized at 80 kN (unloading).

The results of the measurement of Poisson's ratio for the five load steps are

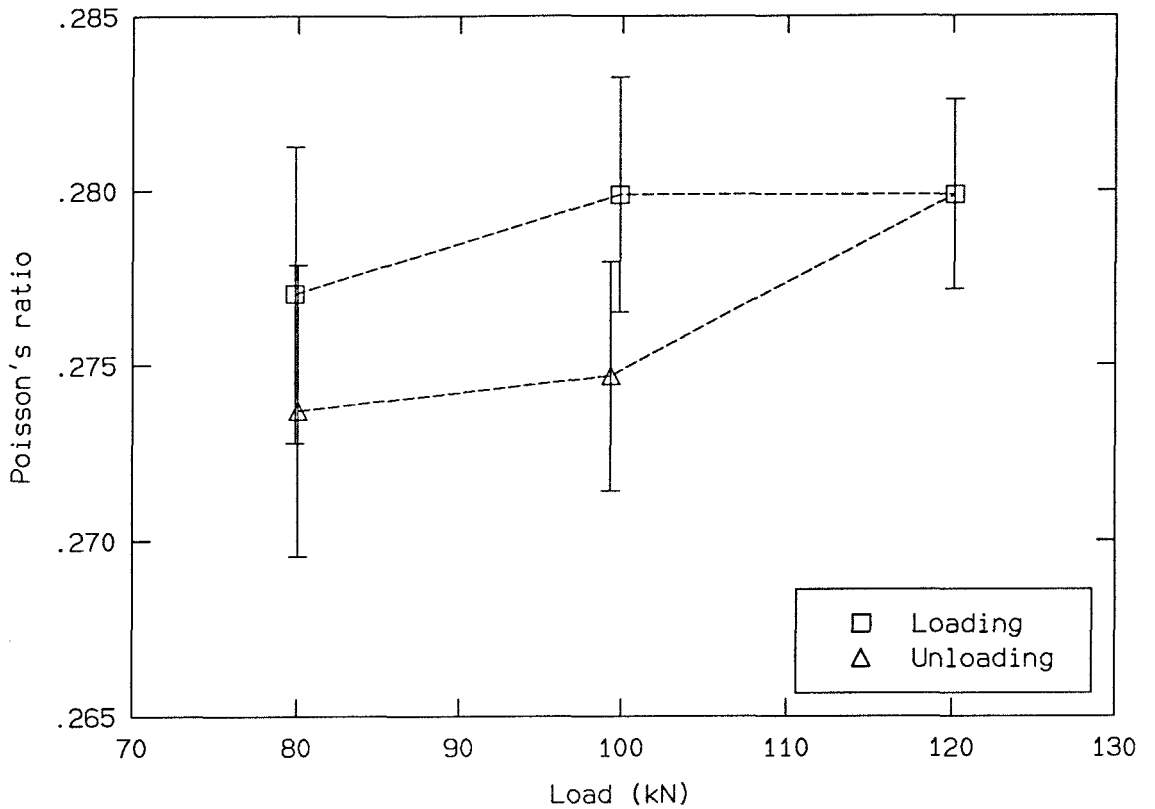


FIGURE 74. Measured values for Poisson's ratio for the three highest loads achieved in the experiment. Higher values are found during loading than during unloading. It is estimated that there is a 90% confidence that the values lie within the error bars, which are approximately ±1% of the measured values.

shown in Figure 74. The values for Poisson's ratio vary from 0.274 to 0.280, with an error estimation of about ±1% on each measurement. The variation from 0.274 to 0.280 is about 2%. The higher values for Poisson's ratio are found on loading, and the lower values on unloading. The curve indicates that there may be some hysteresis in the value for Poisson's ratio.

Without further testing, it is impossible to determine whether the variation in the values of Poisson's ratio are statistically significant, or merely random fluctuations. The information from the least–squares fit indicates either that the model fitting function is not an exceptionally good representation of the data, or that the

Load 10^3 Newtons	ν	Least–Squares Fit Parameters k_1 $10^{-4}m^{-1}$	c_1 10^{-3}	c_2 10^{-4} radian	c_5 10^{-6} radian
79.9	0.277	-1.029	-1.004	-8.677	5.261
99.9	0.280	-1.060	-1.277	-8.792	3.537
120.1	0.280	-1.113	-1.559	-8.997	6.292
99.3	0.275	-1.045	-1.284	-9.001	3.851
80.1	0.274	-1.069	-1.019	-9.035	2.588
Uncertainty	±1%	±0.5%	±0.05%	±0.1%	±20%

Table 5. Variation of parameters from least–squares fit of digitized fringe data. ν is Poisson' ratio. The assumed stress variation is $\sigma_{22}/\mathrm{E} = k_1 x_1 + c_1$. Constant c_2 represents the rigid rotation of the specimen, and constant c_5 represents the deviation from $\pi/2$ of the angle between the crossed gratings.

fringes contain more noise than anticipated. It is estimated that the random error in picking points to fit using the digitizing tablet should have a random error in position of about one tenth of the distance between black fringes. Comparison of the value of the χ^2 merit function with the number of data points in each fit indicates that the random error in the fringe position is somewhere between about 0.23 and 0.25 of the distance between black fringes. If the assumption that the error should be about one tenth of the distance between fringes is correct, the larger error estimate provided by the least–squares fit may indicate that the assumed functional representation of the displacements may not be appropriate to the data, indicating that the strain field is more complicated than assumed. However, a sizable proportion of the fringe points were digitized near the edges of the diffraction grating. The edges of the diffraction grating are subject to noise introduced in the creation process introduced by diffraction from the edges of the collimating lens. This noise may be on the order of three tenths of a fringe for the small deformation in this test.

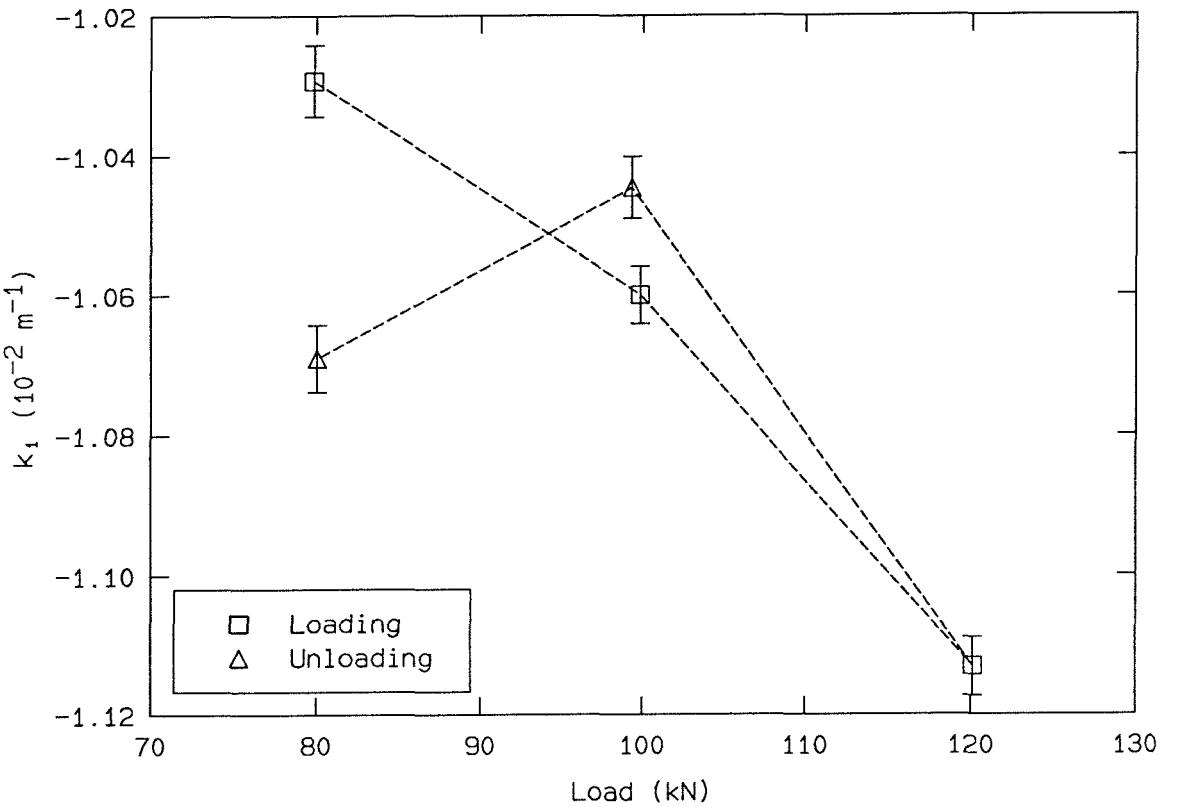


FIGURE 75. Measured values for k_1 for the three highest loads achieved in the experiment. The assumed stress distribution is $\sigma_{22}/\mathrm{E} = k_1 x_1 + c_1$. It is estimated that there is a 90% confidence that the values lie within the error bars, which are approximately $\pm 0.5\%$ of the measured values.

The estimation of a 1% error in each measurement of ν comes from a random perturbation on five of the fit parameters, ν, k_1, c_1, c_2 and c_5, using an increase of 15 in the summed, squared error as a cutoff. From *Numerical Recipes* [64], this should result in a 99 % confidence that the actual value of Poisson's ratio falls between the upper and lower values. The other parameters, c_3 and c_4, are rigid translations introduced through the choice of the fringe number. The number zero was assigned to a fringe near the origin of the coordinate frame so that the rigid offsets would be near zero as well. From the perturbation analysis, the error in each of the values of k_1 as estimated from the least–squares fit is about $\pm 0.5\%$. The error in each of the fit values for c_1 is about $\pm 0.05\%$, the error in each of the fit values for c_2 is about

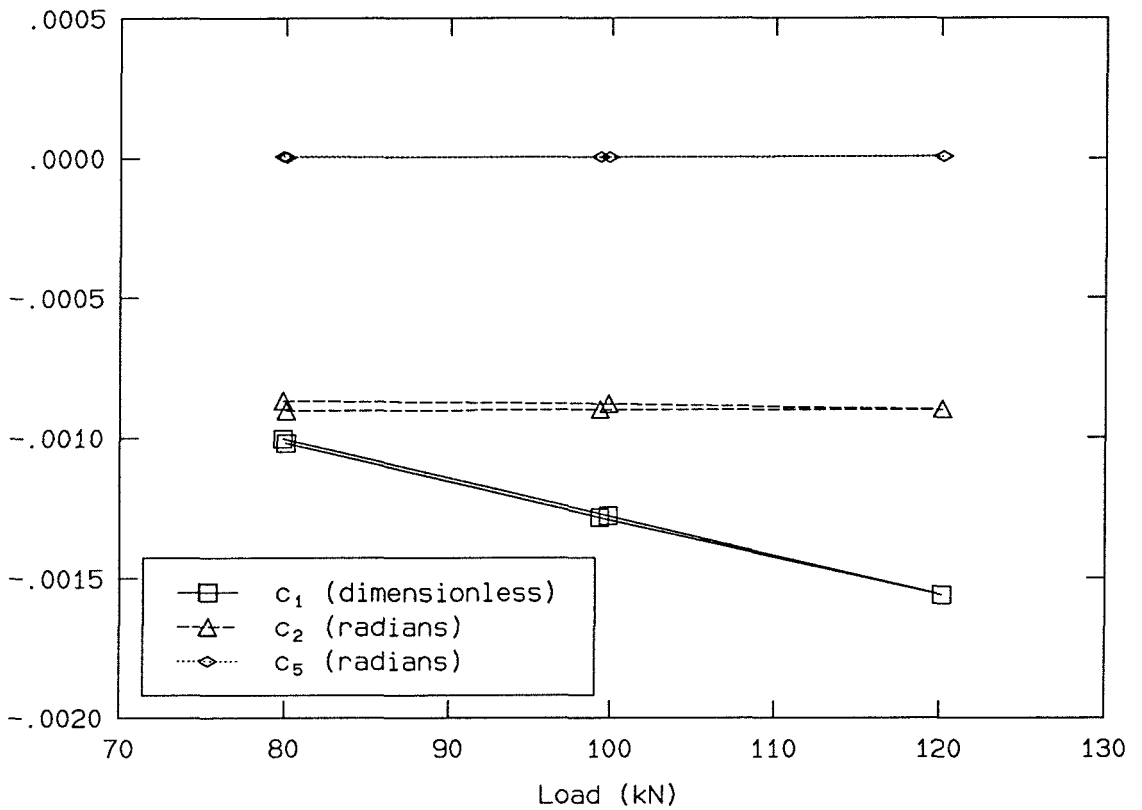


FIGURE 76. Measured values for c_1, c_2 and c_5 for the three highest loads achieved in the experiment. The assumed stress distribution is $\sigma_{22}/\mathrm{E} = k_1 x_1 + c_1$. Constant c_2 represents the rigid rotation of the specimen, and constant c_5 represents the deviation from $\pi/2$ of the angle between the crossed gratings. The error bars are not visible on the scale of the plot.

$\pm 0.1\%$, and the error in each of the fit values for c_5 is estimated to be as much as $\pm 20\%$. However, the values for c_5 are very small compared to any other of the fit parameters. The results from each data set are summarized in Table 5.

The variation of the constants k_1, c_1, c_2 and c_5 with the load are shown in Figures 75 and 76. Constants c_2 and c_5 are approximately constant with load, while c_1 varies linearly with the load. k_1 appears to be approximately linear during loading, but this is not true on unloading.

As input to the numerical model, a value of Poisson's ratio of 0.279 was chosen

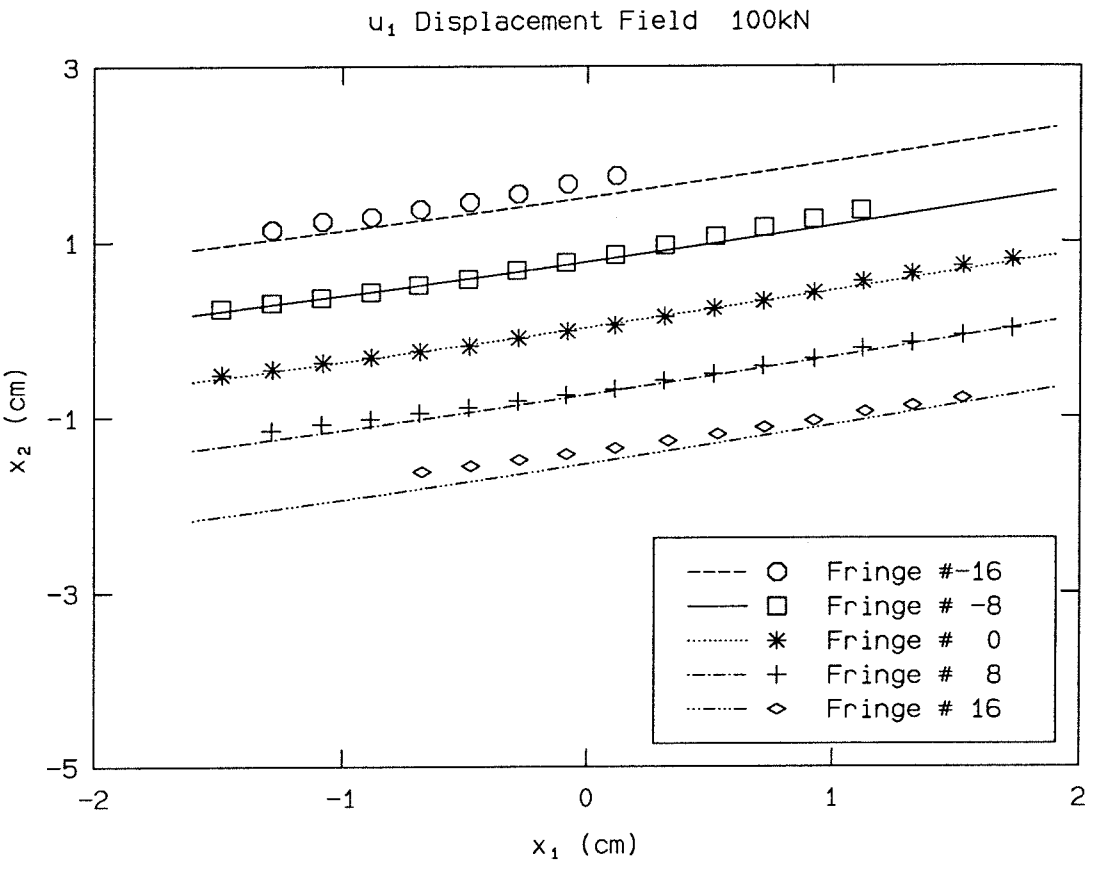


FIGURE 77. The u_1 displacement field at 100 kN, during loading. A comparison between contours of the least–squares fit function and a sample of the data taken using the digitizing tablet. The symbols represent the digitized data; the lines the fit function. The fringe numbers correspond to the number of black fringes, with fringe number zero assigned arbitrarily to lie near the origin of the coordinate frame.

as the mean of the three measurements made during loading. This is also about the value estimated from the preliminary analysis of the out-of-plane displacement data, and within 0.5% of published values of 0.28.

Figures 77 through 80 show a comparison between contours of constant displacement calculated using the functions fit at loads of 100 kN (loading) and 120 kN. The symbols are a subset of the data taken using the digitizing tablet, and the lines represent the corresponding displacement contours calculated using the least–squares fit functions. It is clear that the u_2 displacement field is much better

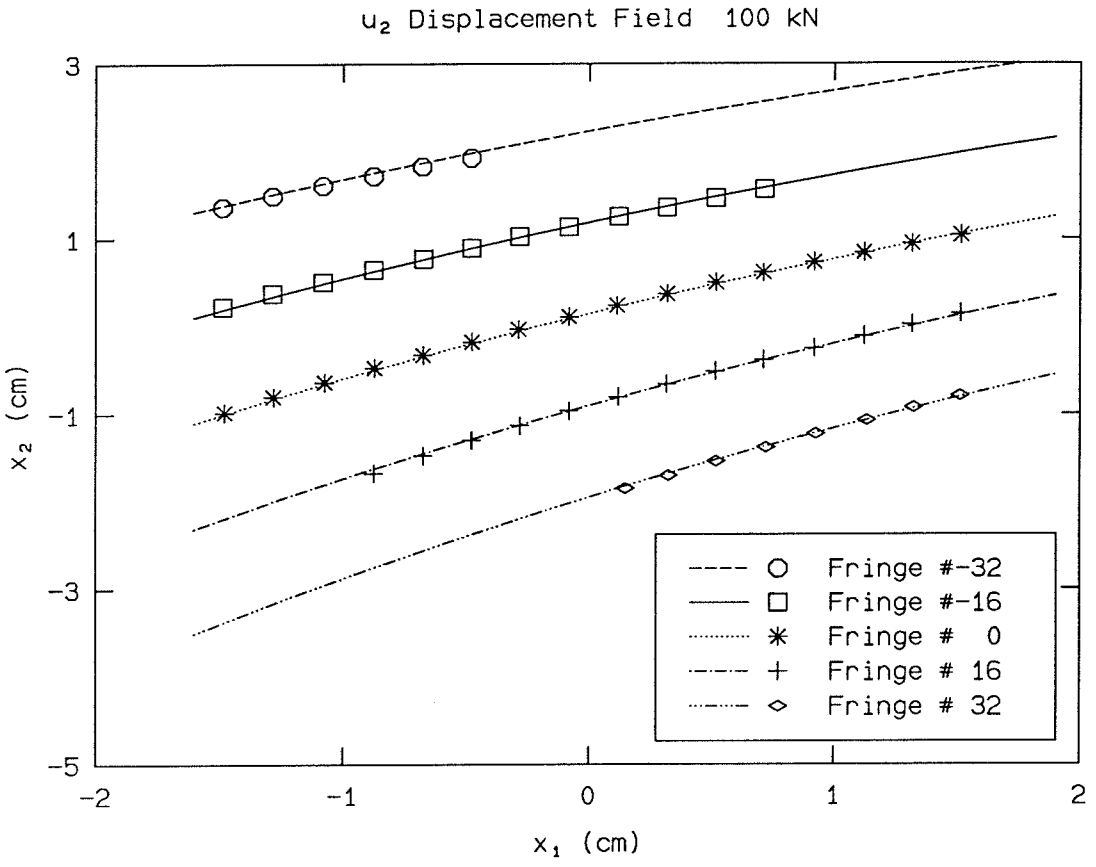


FIGURE 78. The u_2 displacement field at 100 kN, during loading. A comparison between contours of the least–squares fit function and a sample of the data taken using the digitizing tablet. The symbols represent the digitized data; the lines the fit function. The fringe numbers correspond to the number of black fringes, with fringe number zero assigned arbitrarily to lie near the origin of the coordinate frame.

represented with the assumed function of equation (3.5.9). Also, the fit is better in the center of the field of view as compared to the edges of the field.

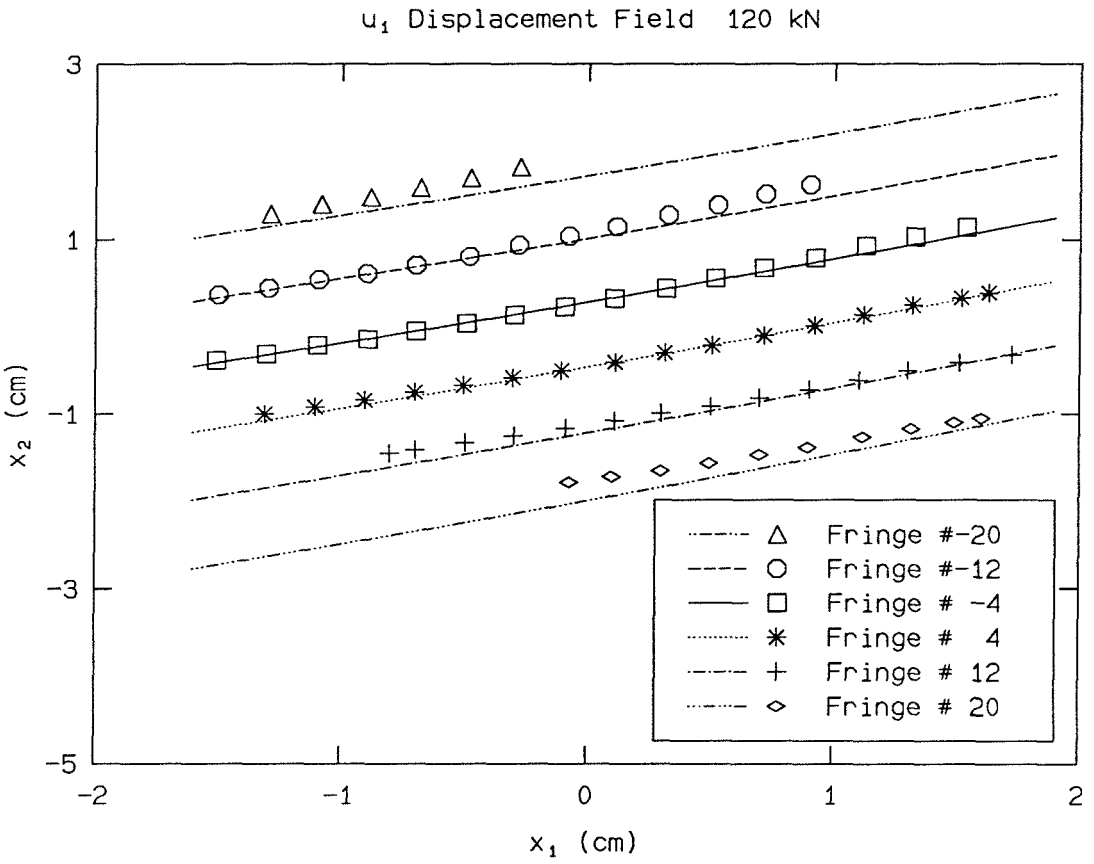


FIGURE 79. The u_1 displacement field at 120 kN. A comparison between contours of the least-squares fit function and a sample of the data taken using the digitizing tablet. The symbols represent the digitized data; the lines the fit function. The fringe numbers correspond to the number of black fringes, with fringe number zero assigned arbitrarily to lie near the origin of the coordinate frame.

3.5.5 Load Correlation

The constants determined in the fit can be used to check the calculated stress field against the applied load. The stress distribution is given by $\sigma_{22}/E = k_1 x_1 + c_1$. The constants used to define the stress distribution depend on the location of the origin of the coordinate axes with respect to the edges of the specimen. The origin of the coordinate frame was chosen arbitrarily when the data were taken with the digitizing tablet. The origin was taken at 16.81 mm from the left side of the specimen, so x_1 varies from -16.81 mm to 19.67 mm. The calculated stress

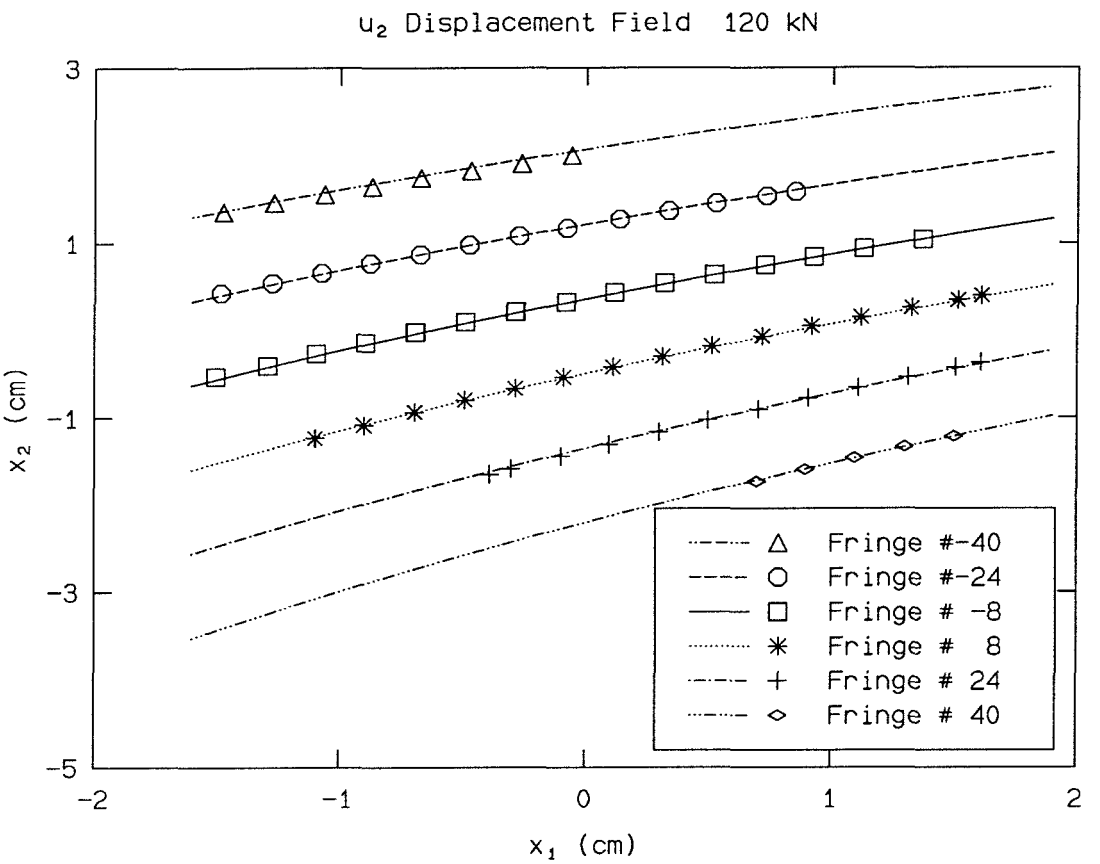


FIGURE 80. The u_2 displacement field at 120 kN. A comparison between contours of the least-squares fit function and a sample of the data taken using the digitizing tablet. The symbols represent the digitized data; the lines the fit function. The fringe numbers correspond to the number of black fringes, with fringe number zero assigned arbitrarily to lie near the origin of the coordinate frame.

distributions are shown in Figure 81. The stress at either edge of the specimen differs from the average stress by about 18% or 19% at 80 kN, about 15% at 100 kN and about 13% at 120 kN. It is clear that the bending stresses are significant in this compression loading.

The load can be calculated by assuming that the stress is constant in x_3, and by integrating the assumed stress distribution over the cross-sectional area. Using the value for Young's modulus determined from the uniaxial tension test of Section 3.3, E = 192 GPa, the calculated load is compared to the applied load (from the

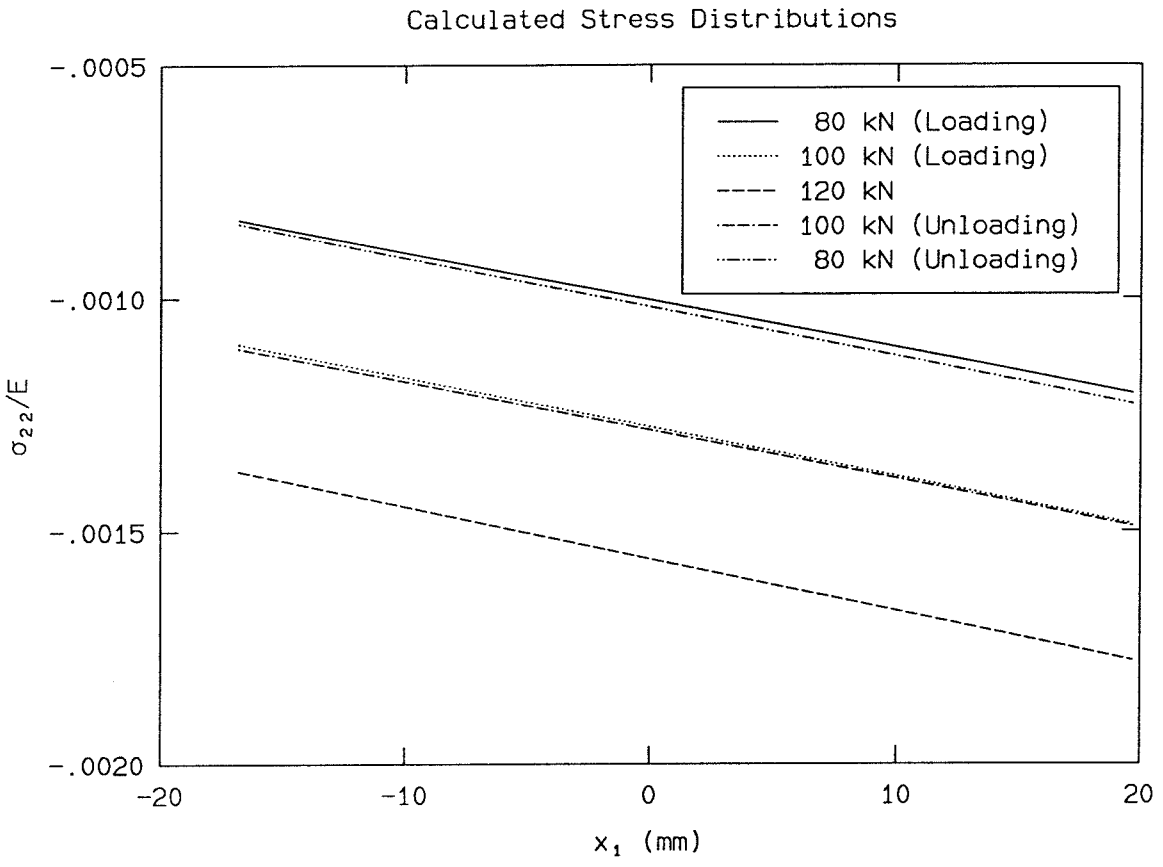


FIGURE 81. Calculated stress distributions for the three highest loads achieved in the experiment. The assumed stress distribution is $\sigma_{22}/\mathrm{E} = k_1 x_1 + c_1$. x_1 varies between -16.81 mm and 19.67 mm.

load cell output) in Figure 82.

The consistently low values calculated from the fit parameters may be an indication of bending stresses that are linear in x_3, similar to the linear stress variation in x_1. Assuming that the applied load represents the average stress on the cross section multiplied by the cross–sectional area, the stress at this surface of the specimen is about 14% to 15% low at 80 kN, about 13% to 14% low at 100 kN and about 13% low at 120 kN. These numbers are comparable to the magnitude of the bending stress variation in x_1. Alternatively, there may be a more complicated stress distribution (in both x_1 and x_3) than that assumed for the measurement.

Fortunately, the load calculation is immaterial to the calculation of Poisson's

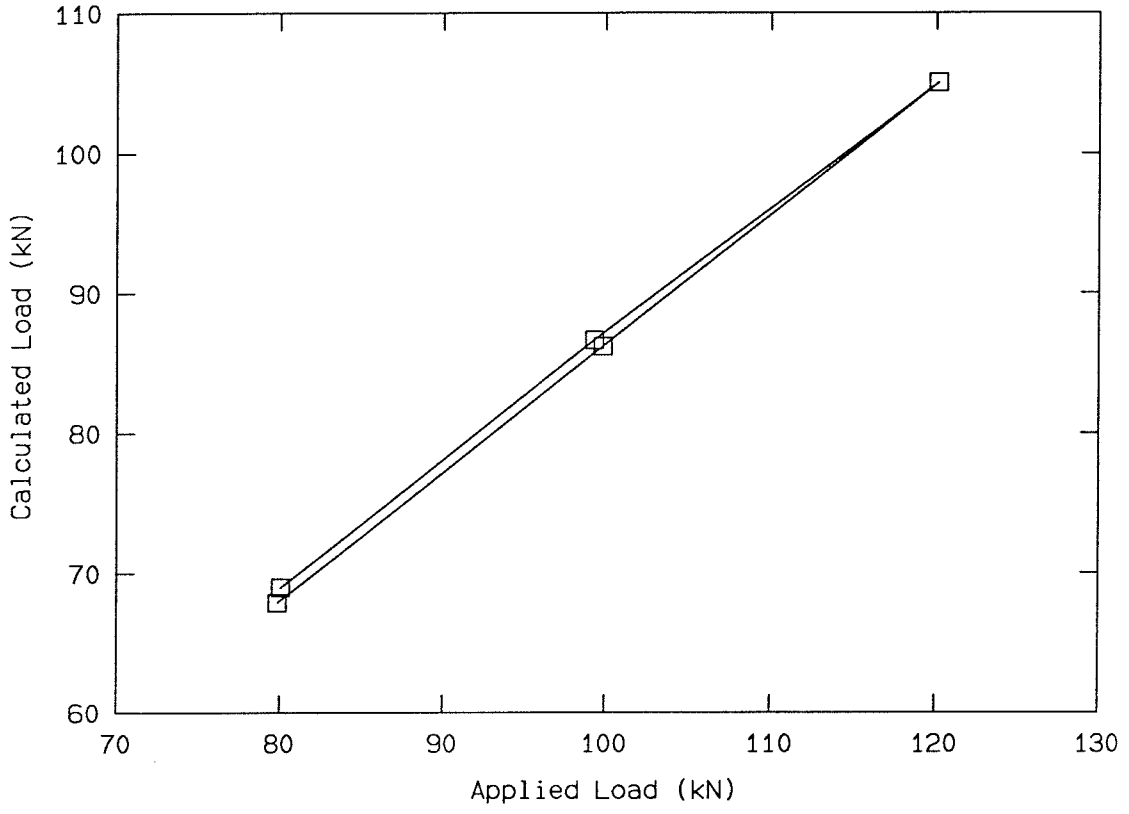


FIGURE 82. The load as calculated from the fit parameters compared to the applied load as indicated by the load cell for the three highest loads achieved in the experiment. The calculated load is approximately 85% of the applied load in each case.

ratio, as the u_1 and u_2 displacements are measured on the same surface in the same region and at the same deformation level. It would be interesting to pursue the analysis a little farther, perhaps by making moire interferometric measurements on the front and back surface at the same time to determine if there is a significant stress variation through the thickness.

However, the analysis of the data would be greatly simplified if there were no bending at all. This test should be performed in tension, if possible. However, as mentioned above, the threaded rods that make the test section of the loading frame adjustable are not capable of sustaining compressive loads, and tension loading of

the threaded rods means compressive loading of the specimen.

CHAPTER 4

Elastic and Plastic Fracture

4.1 Introduction

The thrust of this investigation is the study of fracture in a tough material. In this case, the material is a steel with wide applicability. The steel has been heat treated so as to exhibit a ductile constitutive behavior; this ductility is the mechanism by which the steel is made resistant to fracture. The wide region of large plastic deformations that develops around the tip of a crack distributes the concentrated stresses induced by the crack over a large volume of material; the plasticity serves in effect to blunt the crack and limit the stresses to levels that do not cause failure of the material.

The fracture toughness of a ductile material increases as the thickness of the specimen decreases. Plane strain, which is often considered as the limiting case of a thick specimen, constrains the shear deformation of the material, reducing the ductile distribution of stress throughout the volume. There appears to be a maximum in the dependence of the fracture toughness on the thickness of the specimen, however. As the specimen becomes thinner and thinner, eventually there is not enough material to distribute the stress effectively, and the fracture toughness begins to decrease.

The specimen geometry used in this investigation is relatively thin, and the steel has been heat treated so that it will exhibit a fairly low yield stress and high ductility. Thus, the magnitude of the plastic deformation surrounding the crack is expected to be large, and the plastic zone itself is also expected to include a sizable volume of material.

The lower fracture toughness experienced in thick specimens has led to the use of a plane–strain measure of fracture toughness as a material property for design purposes. Since the fracture toughness of a body of finite thickness will be greater than the plane–strain toughness, design of a structure using a material that satisfies a plane–strain fracture criterion should provide a factor of safety. A considerable amount of literature has been devoted to determining repeatable and reliable experimental methods of measuring fracture toughness, such as those included in the American Society for Testing and Materials (ASTM) standards.

Two of these measures of fracture toughness will be discussed in the following section, the stress–intensity factor, K, and the J–integral. The J–integral is the more recent of the two, but it will be discussed first, as it is applied in fracture of both elastic and plastic materials. The stress–intensity factor is restricted to linearly elastic materials. In each case, a critical value of the fracture toughness measure is used as a material property that may be used in designing structures. Standards such as those of the ASTM are intended to provide requirements on the specimen size and testing geometry necessary for producing reliable, repeatable tests to determine critical values of K or J.

The stress–intensity factor, K, is the multiplying factor of the dominant singular term of the two–dimensional, linearly elastic crack–tip deformation field of Williams [90]; K is the strength of the stress singularity at the tip of a sharp crack. The use of the stress–intensity factor as a fracture–toughness measure originates in that role as the strength of the singularity. The linearly elastic solution of Williams is one of the analytical crack–tip deformation fields which will be discussed in the following section, and it will be used for comparison with the experiment and the finite–element model in the next chapter. Williams's solution is in the form of an infinite series, but only the dominant singular term is retained near the tip of a sharp crack; the deformation defined by the first term in Williams's series will be referred to as the K field. An additional correction term for use in a body with a rounded notch will also be included, as the crack in the steel specimen does have a finite size with a rounded tip. The stress–intensity factor may be calculated as a function of the dimensions of the three–point–bend geometry used in the experiment; such

estimates are available from a variety of references.

The J–integral is effectively a measure of the energy stored in the region surrounding the crack tip that is available to create new surface as the crack advances. Its use is strictly valid for materials that obey a deformation plasticity constitutive law, which is equivalent to a nonlinear elasticity law, but it may be applied to materials that obey an incremental plasticity law if the loading is nearly proportional and there is no unloading. For some geometries, including the three–point–bend geometry in this experiment, the J–integral may be calculated from the work done by the applied loading. This estimate of J from the experiment will be compared with a second method of calculating J through the finite–element simulation. The value of J is also equal to the energy–release rate of a crack in the absence of dissipative material behavior, and for linearly elastic materials, the value of J may be related to the value of K.

The second analytical crack–tip deformation solution that will be compared with the experiment and the numerical model is due to Hutchinson [31] and to Rice and Rosengren [69]; it is called the HRR field. Like Williams's K field, the HRR field is a two–dimensional series solution in which all but the dominant singular term are neglected. The HRR field is applied to materials for which the stress and strain are related by a power law. The constitutive model used in the finite–element model is a power–law approximation to the constitutive law measured for the steel used in the experiment. The HRR field uses a function of the J–integral to determine the strength of the dominant singular term, much as K is the strength of the singularity for the elastic solution. The HRR field is derived within the framework of linearized–strain theory, and like the J–integral, it is strictly valid only for materials following a deformation plasticity constitutive law. As with the J–integral, its use may be extended to materials following an incremental plasticity law if the loading is nearly proportional and there is no unloading.

The finite–element model will be described in Section 4.3. This model is adapted from the model used by Narasimhan, Rosakis and Moran [52]; their assistance with this project is gratefully acknowledged. Their study included a void

growth model of ductile failure, but this feature was removed for the present study as it caused convergence difficulties at the onset of failure. The values of Young's modulus, Poisson's ratio, the yield stress, strain at yield and the hardening exponent were all calculated from the experimental measurements described in the previous chapter. The notch tip radius was changed slightly to match the specimen better, and the number of layers through the thickness was increased from seven to eleven to improve the calculation of the out-of-plane deformation in the vicinity of the notch tip. Details of the model used for the simulation are shown in Section 4.3.1.

This numerical model employs a small–strain, incremental, J_2 plasticity theory with a von Mises yield surface following an isotropic hardening law. A power–law approximation to the constitutive behavior of the experimental material is used so as to allow for comparison between the numerical model and the analytical HRR solutions tabulated by Shih [77]. The numerical solution method is described, including the BFGS iteration procedure for solutions in the nonlinear regime and a predictor–corrector method for updating the stress when plasticity is present. The constitutive law employed in the numerical code is based on the incremental plasticity constitutive law described in Section 3.2.

Internal tunneling of the crack was included artificially in the model through the release of nodes along the crack plane. The nodes released corresponded to the measured crack–tunnel profiles from the experiments described in Section 4.5. For the loads achieved in the experiment, the extent of the tunneling was not very great, although it appears that the tunnel may have some effect at distances well away from the crack tip. The results from the load steps that included the tunnel were complemented by running the simulation to the same loads without tunneling for comparison.

The local value of the J–integral at each nodal layer was calculated using a numerical method for evaluating the contour integral defining J. These local values were then averaged through the the thickness to provide an average value of J at each load step. The local value of J shows considerable variation through the

thickness of the specimen, particularly at higher loads. This evaluation of J is performed for the case including tunneling as well as for the case without tunneling for comparison, although the J–integral is not strictly applicable when the unloading caused by the tunnel occurs. The variation through the thickness of the local value of J differs greatly between the case with tunneling and the case without, but the thickness average of J is not much different between the two cases.

Section 4.4 describes the three–point–bend experiment, with figures showing the specimen geometry and the field of view seen through the camera used to record the fringe patterns of the moire interferometer. The experimental procedure is detailed. One influence on the choice of the three–point–bend geometry is that there is a method of estimating the value of J from the work done by the applied load for this geometry. The measured values of the load and load-point displacement are displayed and compared with the results of the finite–element code. The J–integral is calculated from the work done by the applied load acting through the load-point displacement. This value of J is then compared with the value of J calculated from the load and load-point displacement of the finite–element model and the J–integral calculated by the code using the contour integration scheme described above.

Finally, the experiments performed to measure the extent of crack tunneling within the body are described in Section 4.5. Six specimens were loaded to different loads and the extent of the tunnel marked through heat tinting of the steel. No tunneling was observed for two of the six specimens, and the measured profiles used as input to the finite–element code are shown for three of the remaining specimens. Again, the measured load and load-point displacement of each specimen is compared with the finite–element code, and the value of J is calculated from the work of the indenter.

A critical value of J for this material and geometry may be estimated using the initiation load of between 63500 and 68500 Newtons experimentally determined from the tunnel tests and the value of J calculated from the load and load-point displacement of the tunnel specimens. The values of J calculated from these data are $J = 195\,\mathrm{kN/m}$ at a load of 63500 Newtons and $J = 252\,\mathrm{kN/m}$ at a load of

68500 Newtons. These two values place approximate upper and lower bounds on the actual critical value of J. A second method of determining J from the tunnel tests follows the ASTM standard E 813–81 [7], in which the average crack advance, Δa, is plotted against the value of J calculated from the work done by the indenter. The intersection of a straight line fit through these data points with the line defined by $J = 2\sigma_0 \Delta a$, calculated from the estimated virtual advance of a sharp crack due to blunting. This intersection occurs at $J = 265\,\mathrm{kN/m}$. However, since this specimen begins with a notch having a finite tip radius, a better estimate might be obtained from the intersection of the line through the data points with the line given by $\Delta a = 0$. That intersection occurs at $J = 218\,\mathrm{kN/m}$. Again, these two values should probably be considered as upper and lower bounds. These estimates of a critical value of J should not be mistaken for the plane–strain critical value of J, which should be lower. This specimen does not satisfy the size requirements of the ASTM standard to produce an approved measurement of the plane–strain critical value of J.

Another estimate of the critical value for J may be determined from the finite–element model. The numerical simulation of Narasimhan, Rosakis and Moran [52] uses the void growth model to calculate a critical value of the local value of J through the thickness. The local J is always highest in the center of the specimen. In [52], the estimate of the critical local value of J from the void growth model is $J = 250\,\mathrm{kN/m}$. For the simulation run for this study without tunneling, the maximum local J at 63500 Newtons (the highest load with no tunnel) was $J = 219\,\mathrm{kN/m}$, and the maximum local J at 68500 Newtons (the lowest load with measured tunneling) was $J = 279\,\mathrm{kN/m}$. Also, it might be expected that the local value of J would remain near the critical value of J in the load step nearest to crack initiation. If that expectation is correct, the local value of J calculated from the simulation including tunneling should be close to the critical value of J at 68500 Newtons. That value of J was approximately $J = 260\,\mathrm{kN/m}$.

4.2 Analytical Approximations for Elastic and Plastic Crack–Tip Deformation Fields

Two analytical solutions for the deformation at the tip of a sharp crack will be compared with the data from the finite–element code and the data from the experiment. These analytical solutions are two-dimensional, asymptotic representations for the dominant stress singularity in the region close to the crack tip. The first solution is the eigenfunction solution for linearly elastic materials that is due to Williams [90], which is governed by the stress–intensity factor K; this solution will be referred to here as the K field. The second solution is a solution derived for power–law hardening materials obeying a *deformation* plasticity relation, which is due to Hutchinson [31] and Rice and Rosengren [69], called the HRR field.

4.2.1 The J–Integral

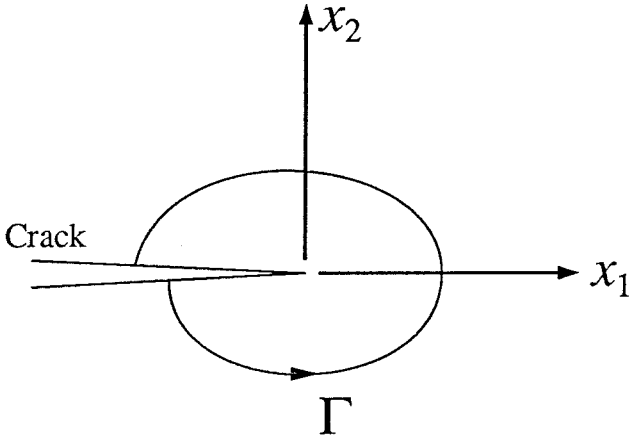


FIGURE 83. A sample contour Γ used to evaluate the J–integral. Any contour that contains only the singularity at the crack tip will yield the same value for J as long as the material is elastic and the crack surfaces are traction–free.

Both of the singular solutions (which are asymptotic to the crack tip) may be related to the applied boundary conditions of the problem through the use of the J–integral of Rice [65,66], which is a measure of the energy absorbed in the

deformation in the region surrounding the crack tip. The J–integral is derived for a two–dimensional planar body, with the integral taken along any contour Γ which encloses only the singularity at the crack tip, as shown in Figure 83. J is a path integral defined by

$$J = \int_{\Gamma} \left(W dx_2 - \mathbf{T} \cdot \frac{\partial \mathbf{u}}{\partial x_1} ds \right) \tag{4.2.1}$$

In equation (4.2.1), ds is the element of arc length along Γ, $\mathbf{T}$ is the traction vector defined by the outward normal to Γ, and $\mathbf{u}$ is the displacement vector. W is the strain–energy density defined by the integral over the loading history given by

$$W = \int_0^{\boldsymbol{\epsilon}} \sigma_{ij} d\epsilon_{ij} \tag{4.2.2}$$

Note that the J–integral is derived for materials following a deformation plasticity relation (equivalent to nonlinear elasticity), as is the HRR field. Under the assumption of linearly or nonlinearly elastic behavior, the J–integral is path–independent, so that the contour Γ may be any contour that contains no other singularities but that at the crack tip. In particular, the contour may be taken arbitrarily close to the crack tip and evaluated using the asymptotic solutions for linear elasticity or power–law hardening materials. Equally, the contour may be taken arbitrarily large, and for certain boundary conditions evaluated directly [65]. The path independence of J then determines the strength of the singular, asymptotic terms based on the applied boundary conditions.

4.2.2 The J–Integral as a Fracture Criterion

Under the restriction of nondissipative material behavior, the J–integral has been shown to be equivalent to the energy–release rate, $\mathcal{G}$, of a propagating crack [65,66]. For this class of materials, the value of J is equal to the change in the potential energy for a unit crack advance. The correspondence between J and $\mathcal{G}$ for elastic solids (linear or nonlinear) has led to the use of a critical value of J, J_c, as a fracture criterion, *which is considered to be a material property.* When the value of J reaches J_c, there is enough strain energy stored in the material around the crack tip to initiate crack propagation. A large amount of literature is devoted

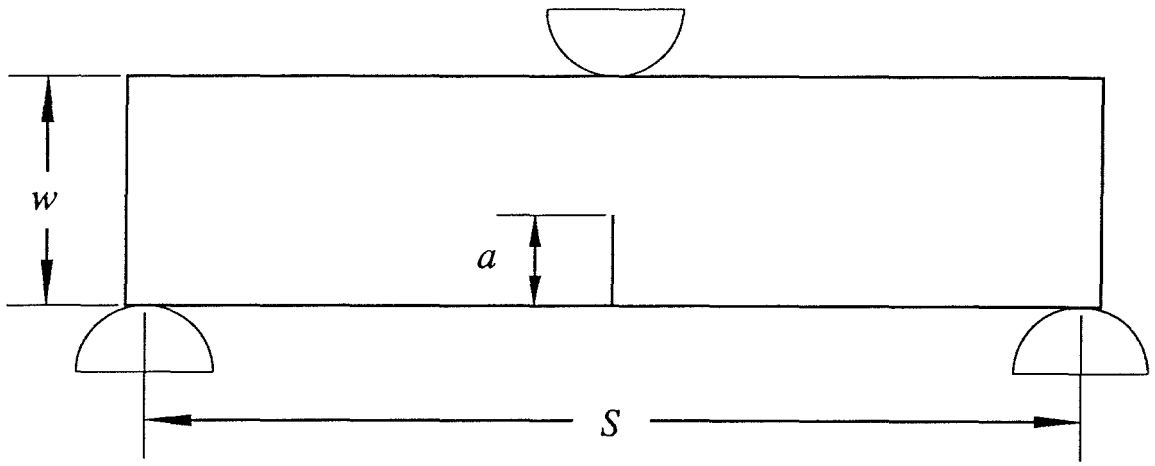


FIGURE 84. Three–point–bend specimen geometry. S is the span, w is the width. For $S/w = 4$, the J–integral can be directly calculated from the load and load–point displacement relation.

to establishing repeatable test methods for measuring J for various geometries, and in particular, to determining J_c. [1,7,10,15,20,29,39]

One influence on the choice of the specimen geometry used for the three–point–bend experiment is that a simple method of calculating the J–integral has been proposed for this geometry [15,26,68]. See Figure 84. For a three–point–bend geometry in which the width of the specimen, w, is one quarter of the span between the supports, S, the J–integral is given by

$$J = \frac{2}{(w-a)t} \int_0^{\delta} P d\delta \tag{4.2.3}$$

where a is the crack length, t is the specimen thickness, P is the applied load and δ is the displacement of the loading point. This simple relation between the value of the J–integral and the work done by the applied loading will be compared with the J–integral value estimated from a path integral employed by the numerical code. This value of the J–integral may also be compared with the magnitude of the stress–intensity factor calculated from boundary collocation methods. Additionally, the J–integral provides the scaling parameter for the magnitude of the HRR singularity, which will be used to compare with the deformation at the crack tip when there is a large plastic zone.

The J–integral calculated using equation (4.2.3) may also be used to estimate a critical value of J for this geometry and material, through the tunneling tests described in Section 4.5. However, it should be noted that the specimen geometry used in this study does not strictly follow the ASTM guidelines for the specimen size. The ASTM provides a standard [7] for determining the critical plane–strain value of the J–integral, J_c, to be used as a material property in the design of structures. Following the ASTM standard [7], an approved three–point–bend specimen should have a crack length, a, which is at least one half of the specimen width, w. The crack length of the specimen in this study is only four tenths of the specimen width. Additionally, the crack should be a sharp fatigue crack, rather than a notch with a circular tip. Perhaps most importantly, the critical value of J is intended to be measured in plane strain, as the plane–strain fracture toughness of a material is lower than the plane–stress fracture toughness, owing to the stresses through the thickness that constrain the size of plastic deformation. The plain–strain fracture toughness is thus a conservative estimate of the fracture toughness for the material in a specimen of finite thickness. In order to maintain plane–strain conditions, it is suggested by Hertzberg [26] that any of the planar dimensions of the specimen should be much larger than the zone of intense plastic flow; this zone is on the order of $2J/\sigma_0$, where σ_0 is the uniaxial yield stress of the material, as estimated by Rice [66]. Hertzberg [26] indicates that the specimen thickness and uncracked ligament should be at least twenty-five to fifty times as large as J/σ_0. The J values for this specimen range up to about 360 kN/m, and the critical value for fracture initiation is estimated to be approximately 250 kN/m (see Sections 4.3 and 4.5). The yield stress σ_0 is 960 MPa, so the planar length dimensions at 250 kN/m should be on the order of 0.65 cm to 1.3 cm, at minimum. At 360 kN/m, the planar length dimensions should be at least 0.94 cm to 1.88 cm. The specimen thickness is 1 cm, and the uncracked ligament is 4.6 cm, so this specimen seems to lie just within the suggested size scales. However, the plastic zone at the largest loads is on the order of the specimen thickness, and the ultimate failure mode is in shear, which is usually thought to correspond to a plane–stress deformation. Additionally, the plastic zones shown in Section 5.2 do not seem to indicate there are any regions that

could be considered plane strain, even in the center of the specimen, beyond a load of 35000 Newtons, which corresponds to a J value of about 50 kN/m. It appears that the high ductility (low hardening) in this material means that the specimen dimensions would have to be significantly larger to produce plane–strain conditions within the body.

One of the goals of this research is to extend the experimental investigation to deformations that include sizable plastic deformation, which is accomplished through use of a very ductile material and a relatively thin specimen geometry. The plane–strain fracture toughness is intended as a lower limit on the fracture toughness of a specimen of any thickness. A better understanding of the higher fracture toughness obtained through mechanisms such as ductility in metals or Cook-Gordon crack stopping in composites may allow for the design of structures that take advantage of the higher fracture toughness.

4.2.3 Linear Elasticity Solution

The asymptotic, two-dimensional eigenfunction solution for the stress function for a sharp crack in a linearly elastic body given by Williams [90] assumes a separable series solution of the biharmonic equation of the form $r^s f(\theta, s)$, where r and θ are as defined in Figure 85. The traction–free boundary condition along $\theta = \pm\pi$ determines the allowable values of s to be $s = \pm n/2$, where n is an integer. The traction–free boundary conditions and an assumption of symmetry with respect to the crack plane also determine the function $f(\theta, s)$, except for a multiplying factor. The symmetry assumption yields the Mode I distribution; an assumption of anti-symmetry would yield the Mode II distribution. Mode III results from assuming antiplane shear deformation. For the present discussion, only the Mode I deformation will be considered.

Requiring that the strain energy be bounded in any region surrounding the crack tip as r tends to zero limits the allowable values of s, so that the stresses and strains may have singular behavior as $r^{-1/2}$, plus higher–order terms in the series. The higher–order terms will be neglected here.

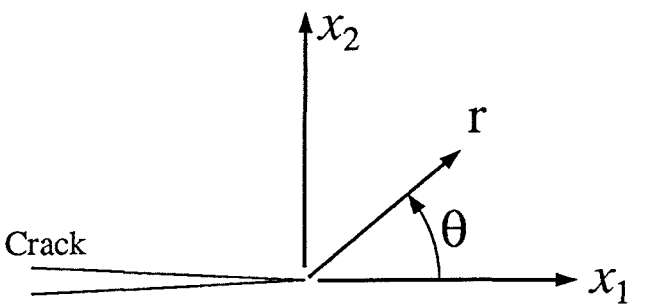


FIGURE 85. The origin is taken at the crack tip. The asymptotic fields are determined from a potential function that is assumed to be separable in r and θ.

For the Mode I crack geometry [66], the stresses are found in the right-handed $x_1x_2x_3$ Cartesian coordinate system of Figure 85 as

$$\begin{aligned}
\sigma_{11} &= \frac{K}{\sqrt{2\pi r}} \cos\frac{\theta}{2}\left(1 - \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right) \\
\sigma_{12} &= \frac{K}{\sqrt{2\pi r}} \cos\frac{\theta}{2}\left(\sin\frac{\theta}{2}\cos\frac{3\theta}{2}\right) \\
\sigma_{22} &= \frac{K}{\sqrt{2\pi r}} \cos\frac{\theta}{2}\left(1 + \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right)
\end{aligned} \tag{4.2.4}$$

and the corresponding in-plane displacements as

$$\begin{aligned}
u_1 &= \frac{K}{2G}\sqrt{\frac{r}{2\pi}} \cos\frac{\theta}{2}\left(\kappa(\nu) - \cos\theta\right) \\
u_2 &= \frac{K}{2G}\sqrt{\frac{r}{2\pi}} \sin\frac{\theta}{2}\left(\kappa(\nu) - \cos\theta\right)
\end{aligned} \tag{4.2.5}$$

where ν is Poisson's ratio and G is the elastic shear compliance. G is related to the Young's modulus, E, by

$$G = \frac{E}{2(1+\nu)} \tag{4.2.6}$$

The function of Poisson's ratio, $\kappa(\nu)$, depends on whether the solution is for plane strain or plane stress.

$$\kappa(\nu) = \begin{cases} 3 - 4\nu & \text{for plane strain} \\ \dfrac{3-\nu}{1+\nu} & \text{for plane stress} \end{cases} \tag{4.2.7}$$

The corresponding out-of-plane displacements are identically zero for the plane–strain case, and for plane stress they are proportional to the trace of the stress tensor, and are therefore found to be singular, with

$$u_3 = \frac{-K\nu t}{E\sqrt{2\pi r}} \cos\frac{\theta}{2} \tag{4.2.8}$$

where t is the specimen thickness.

The multiplying factor, K, called the *stress–intensity factor*, represents the strength of the singularity. The stress–intensity factor depends on the specific geometry of the specimen and applied boundary conditions. The stress–intensity factor is a linear function of the applied load. Stress–intensity factors for various geometries have been collected, as by Tada, Paris and Irwin [83], or Rooke and Cartwright [70].

4.2.4 Added Correction for Finite Notch Tip

The K field solution is derived for a mathematically sharp crack, having a tip of zero radius. Tada, Paris and Irwin [83] also provide a correction for the stress field when the sharp crack is replaced with a notch having a finite tip radius. The geometry is as shown in Figure 86. The notch is shown as having a circular tip with a radius of curvature ρ. In this case, the origin of the coordinate frame is taken at $\rho/2$ ahead of the center of curvature of the notch tip. The correction consists of a stress term of the order $r^{-3/2}$, which is intended to make the surface of the notch traction–free. These stresses are added to those in equation (4.2.4), so that the total stress is given by

$$\begin{aligned}
\sigma_{11} &= \frac{-K}{\sqrt{2\pi r}}\frac{\rho}{2r}\cos\frac{3\theta}{2} + \frac{K}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\left(1-\sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right) \\
\sigma_{12} &= \frac{-K}{\sqrt{2\pi r}}\frac{\rho}{2r}\sin\frac{3\theta}{2} + \frac{K}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\left(\sin\frac{\theta}{2}\cos\frac{3\theta}{2}\right) \\
\sigma_{22} &= \frac{K}{\sqrt{2\pi r}}\frac{\rho}{2r}\cos\frac{3\theta}{2} + \frac{K}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\left(1+\sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right)
\end{aligned} \tag{4.2.9}$$

where ρ is the radius of curvature of the notch as shown in Figure 86. Note that these additional terms have no effect on the out-of-plane deformation field, as the trace of the additional stress terms, $(\sigma_{11}+\sigma_{22})$, is zero. The in-plane displacements associated with this stress field are given by

$$\begin{aligned} u_1 &= \frac{K}{2G\sqrt{2\pi}} \cos\frac{\theta}{2}\left[\frac{\rho}{\sqrt{r}} + \sqrt{r}\big(\kappa(\nu) - \cos\theta\big)\right] + C \\ u_2 &= \frac{K}{2G\sqrt{2\pi}} \sin\frac{\theta}{2}\left[\frac{\rho}{\sqrt{r}} + \sqrt{r}\big(\kappa(\nu) - \cos\theta\big)\right] \end{aligned} \tag{4.2.10}$$

where C is an arbitrary constant of integration. The u_2 displacements are constrained to be zero along the crack line because of symmetry, so there is no corresponding rigid translation for u_2. Since the trace of the additional stress terms is zero, the in-plane displacement field associated with the added correction is independent of whether the deformation is assumed to be plane stress or plane strain. One convenient choice for C is $C = 0$, which would be most useful if the boundary conditions applied were displacements corresponding to the terms of the K field for a sharp crack, although the applied boundary conditions would have to be centered at the appropriate position of the origin, which would be different for different notch tip radii. A second choice for C, convenient for this study, is to select C so that u_1 is zero at the notch edge, at the location of the first nodal coordinate $(r = \rho/2, \theta = 0)$. In this case, C is given by

$$C = \frac{-K}{2G\sqrt{2\pi}}\Big(\kappa(\nu)+1\Big)\sqrt{\frac{\rho}{2}} \tag{4.2.11}$$

This additional correction for a finite notch tip radius has two appealing properties, the first of which is that the magnitude of the $r^{-3/2}$ stress term is also governed by the same stress–intensity factor as the $r^{-1/2}$ stress term from the solution for the sharp crack. Thus, a stress–intensity factor calculated from the boundary conditions applied to a body containing a sharp crack is also applicable to the same body in which the crack has been replaced by a rounded notch. The second appealing property is that the stresses and displacements of the added correction possess a stronger singularity than the sharp crack stresses and displacements, so that the correction may be ignored if r is much larger than ρ.

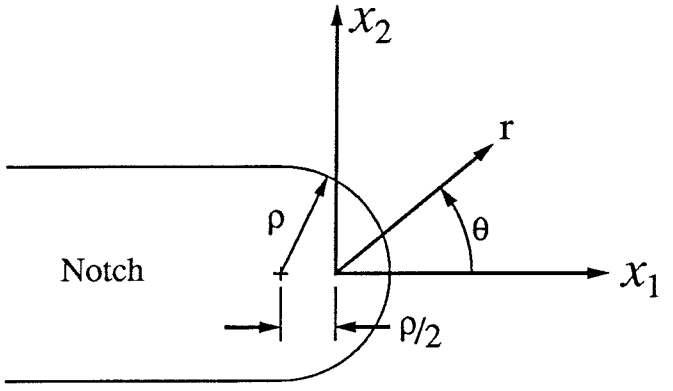


FIGURE 86. The local geometry at a circular notch tip from Tada, Paris and Irwin [83]. The radius of curvature of the notch is ρ, and the origin of the r, θ coordinate system should be $\rho/2$ ahead of the center of curvature to use the K field solution with correction.

4.2.5 The Stress–Intensity Factor as a Fracture Criterion

A critical value of the stress–intensity factor, K_c, has also been used as a material dependent fracture criterion, with an equivalent ASTM standard for measuring K_c [5]. Since the stress–intensity factor is derived within the framework of linear elasticity, however, its applicability as a fracture criterion must be limited to elastic materials, or to materials in which the size of the plastic zone around the crack tip is contained in a small region (compared with some specimen dimension). The latter condition is described as *small–scale yielding*. By contrast, the J–integral criterion is intended to be used with specimens that experience plastic deformation on a larger scale.

As mentioned earlier, the specimen geometry of this study is not an ASTM–approved geometry for determining J_c, the critical value of the J–integral necessary to initiate fracture propagation, because the dimensions of the specimen are not large enough to ensure that plane–strain conditions prevail, or to keep the plasticity from having a significant effect on the stress distribution along the uncracked ligament when the ultimate failure loads are approached. In fact, the dimensions of an ASTM–approved geometry to measure K_c must be on the order of twenty times as large as the dimensions of an ASTM–approved geometry to measure J_c,

as incicated by Hertzberg [26]. The larger dimensions of the specimen to determine K_c are necessary to ensure that any plastic deformation is well contained and may be considered to be small–scale yielding. The ductility of the material used in this study results in a plastic zone that has a size on the same order as any other length dimension of the specimen geometry.

The stress–intensity factor K and the J–integral are directly related when the material response is linearly elastic, or when the deformation may be described as small–scale yielding and the contour Γ is taken outside the plastic region. For plane stress, it is found that

$$J = \frac{K^2}{\mathrm{E}} \tag{4.2.12}$$

while for plane strain

$$J = \frac{(1-\nu^2)K^2}{\mathrm{E}} \tag{4.2.13}$$

where E is again Young's modulus and ν is Poisson's ratio. From the relation that $J = \mathcal{G}$, it follows that the energy–release rate of a propagating crack, $\mathcal{G}$, is also related to the stress–intensity factor, K.

The stress–intensity factor predates the J–integral, and there exists even more literature on repeatable methods of determining K and K_c than exists for J and J_c. One traditional method for calculating stress–intensity factors has been through boundary–collocation techniques [22]. In boundary collocation, Williams's series solution for the stress function is matched with the stress function corresponding to the applied boundary conditions at m selected points. This procedure yields values for the first $2m$ coefficients of the series, with the remaining terms neglected. In the paper by Gross and Srawley [22], m is taken as 15, 18, 21 and 24 points successively, to ensure that the procedure converged to the correct value for the stress–intensity factor, K.

Stress–intensity factors for the three–point–bend geometry used in this study are available from a variety of sources [5,14,22,26,70,83]. An average of the values found in the literature has been chosen as

$$K = (27.95\ \mathrm{m}^{-1/2})\left(\frac{P}{t}\right) \tag{4.2.14}$$

where P is the applied load, and t is the thickness of the specimen. The variation in the value of K as calculated from the various sources in the literature is about $\pm 1\,\%$.

4.2.6 Remarks

There are some obvious difficulties in approximating the deformation of the three–point–bend specimen of the current study with the K field solution. The biggest discrepancy is the finite thickness of the specimen. The K field is a two–dimensional solution, either for plane stress or plane strain. The finite thickness of the specimen introduces three–dimensional effects not accounted for in the K field solution. Corrections could be made to the K field plane–stress solution, but Swedlow [82] suggests that such corrections will not be successful, owing to the singularity of the K field. Part of this study is intended to examine the importance of the three–dimensional nature of the deformation near the notch tip as it affects the in-plane displacement fields on the surface of the body.

A second difficulty with the classical K field is that the stress concentration at the notch tip results in plastic deformation. However, in small–scale yielding, the inclusion of a plastic zone surrounding the notch tip can be treated in a way that is similar to the correction for a finite tip radius. In this case the plastic zone is felt as a small increase in the notch tip radius, and the K field is expected to remain valid in a region outside the plastic zone, in the same way that the more singular term that is due to the finite notch tip radius dies out away from the notch tip. Plastic deformation is important in that it provides the mechanism for the high toughness of ductile metals. The HRR field is one attempt to model the plastic deformation at the notch tip; however, the HRR field is expected to be valid only in a small region inside the plastic zone. Investigation of the evolution of the plastic deformation around the notch is also part of this study, as is the extent of the range over which the HRR field is valid, if any.

A third difficulty associated with the K field is the question of the size scale at which the singular term of the K field no longer dominates the nature of the

deformation surrounding the notch tip. One way to investigate this question is to include higher–order terms from the Williams's series solution. However, if the three–dimensional effects are large, or the plastic zone is large, there may be no region in which the deformation is well described by Williams's series.

The finite notch tip correction is derived from the solution for the deformation at the tip of a long, thin ellipse. Since no mathematically sharp crack exists in a physical body, owing to the atomic nature of real solids, perhaps the best way to approach a three–dimensional problem would be to add a three–dimensional correction to the solutions for an elliptical or hyperbolic notch worked out by Neuber [53]. These functions possess no singularities, although the elliptic or hyperbolic geometry is somewhat difficult to tackle. See Pfaff [56].

If nothing else, the K field solution provides a starting point for studying the deformation fields. However, the K field is derived for two-dimensional deformation, either plane strain or plane stress. One point of interest is the effect of the three-dimensionality of the specimen on the deformations predicted using the K field. If it is assumed that the two cases of plane strain and plane stress represent limiting cases of infinitely thick (plane strain) and infinitesimally thin (plane stress) three–dimensional bodies, the two–dimensional solutions may be thought of as bounds on the deformations expected in a finite three–dimensional body. The out-of-plane deformation is zero in plane strain, but singular in plane stress, so there is an infinite change between the limit cases for the out-of-plane deformation. In either plane strain or plane stress, though, the in-plane deformation is bounded. One might therefore expect the in-plane deformation to be better characterized by the K field than the out-of-plane deformation.

The variation of the in-plane deformation fields between plane strain and plane stress then depends on the change in the value of $\kappa(\nu)$ between plane strain and plane stress. For a Poisson's ratio of $\nu = 0.3$, the plane–strain value of $\kappa(\nu)$ is 1.80, while the plane–stress value of $\kappa(\nu)$ is 2.08, a difference of about 15%. However, the effect of the change in $\kappa(\nu)$ is also a function of the value of θ, as may be seen in equation (4.2.10). Along $\theta = 0$, the plane–strain value of the quantity $\left[\kappa(\nu) - \cos\theta\right]$

is 0.80, while for plane stress the value of $[\kappa(\nu) - \cos\theta]$ is 1.08, a difference of about 30%. Along $\theta = \pi$, the plane–strain value of the quantity $[\kappa(\nu) - \cos\theta]$ is 2.80, while for plane stress the value of $[\kappa(\nu) - \cos\theta]$ is 3.08, a difference of just 10%. Thus, while the variation in $\kappa(\nu)$ between plane strain and plane stress may be 15%, the variation in the limit cases of the displacement fields as predicted by the K field solution may be as little as 10% or as much as 30%.

4.2.7 Power–Law Hardening Materials — HRR Field

An asymptotic solution for the case of a crack in a two-dimensional body composed of a material that follows a power–law constitutive relation during plastic deformation was derived by Hutchinson [31] and by Rice and Rosengren [69], and has become known as the HRR field. This solution was derived within the framework of deformation plasticity, in which the stresses and strains can be directly related, unlike incremental plasticity, in which the stress and strain rates are related. Since deformation plasticity is equivalent to nonlinear elasticity, the stresses and strains may be derived from a potential function. The HRR field solution is also found by assuming that the potential follows a spatial distribution of the form $r^s f(s,\theta)$, and by restricting the solution to include only the dominant singular term, which is again determined by assuming that the strain energy is bounded in any region surrounding the crack tip. Solution of the equilibrium equation using the potential function results in a fourth–order differential equation for $f(s,\theta)$. Imposing traction–free boundary conditions along $\theta = \pm\pi$ yields a unique solution if s is chosen. This method of solution is analogous to the Williams K field solution in linearly elastic materials, which is also a separable solution in r and θ, and the HRR solution reduces to the K field for a material with a hardening exponent of unity, corresponding to linearly elastic behavior.

The material is assumed to follow a power law governed by the hardening exponent n, so that the strains are a function of the stresses to the power n. By bounding the strain–energy density in any region about the crack tip, the stresses are found to have a singularity of $r^{-1/(n+1)}$, and the strains are found to have

a singularity of $r^{-n/(n+1)}$ as r approaches zero. As in the K field, the in-plane displacements are bounded, behaving as $r^{1/(n+1)}$ as r approaches zero. The out-of-plane displacement is singular as r approaches zero, however. The differential equation for the θ variation has been solved numerically and tabulated for plane stress and plane strain for various values of n by Shih [77]. For elastic materials, with $n = 1$, the K field singularity of $r^{-1/2}$ is recovered. For power–law materials, the stresses have a milder singularity than the stresses found using linear elasticity, but the strains have a stronger singularity than the strains in the linearly elastic case.

Again, this solution is asymptotic to the crack tip, and is related to the applied boundary conditions through the J–integral. In this case, the linearly elastic field quantities are assumed to be small compared with the plastic deformation, and the elastic deformation is neglected.

Shih [77] assumes a uniaxial constitutive law in which the plastic strain is related to the applied stress by

$$\frac{\epsilon_{11}^p}{\epsilon_0} = \alpha\left(\frac{\sigma_{11}}{\sigma_0}\right)^n \tag{4.2.15}$$

where σ_0 is the yield stress, $\epsilon_0 = \sigma_0/E$ is the strain at yield, and α is a material constant. n is the hardening exponent, and all elastic strains are neglected. This relation is generalized to multiaxial stress using J_2 deformation plasticity as

$$\frac{\epsilon_{ij}^p}{\epsilon_0} = \frac{3}{2}\alpha\left(\frac{\overline{\sigma}}{\sigma_0}\right)^{n-1}\frac{S_{ij}}{\sigma_0} \tag{4.2.16}$$

where S_{ij} is the deviatoric stress defined as $S_{ij} = \sigma_{ij} - (1/3)\sigma_{kk}\delta_{ij}$, and $\overline{\sigma}$ is the current yield stress given by

$$\overline{\sigma}^2 = \frac{3}{2}S_{ij}S_{ij} \tag{4.2.17}$$

The incremental plasticity law used in the finite–element code is described in Section 3.2. Deformation plasticity and incremental plasticity coincide when the loading path is restricted to be proportional. In proportional loading, the stress tensor is a time–independent tensor multiplied by a positive, monotonically increasing scalar function of time. When proportional loading holds, the incremental

plasticity law takes the form

$$\frac{\epsilon_{ij}^p}{\epsilon_0} = \frac{3}{2}\left[\left(\frac{\overline{\sigma}}{\sigma_0}\right)^{n-1} - 1\right]\frac{S_{ij}}{\sigma_0} \tag{4.2.18}$$

Since the elastic strains are neglected in the HRR formulation, the second term in equation (4.2.18) is assumed to be negligible compared to the first, so that the plastic strain may be approximated as

$$\frac{\epsilon_{ij}^p}{\epsilon_0} = \frac{3}{2}\left(\frac{\overline{\sigma}}{\sigma_0}\right)^{n-1}\frac{S_{ij}}{\sigma_0} \tag{4.2.19}$$

Compare equation (4.2.19) to equation (4.2.16). Note that for this study, the material parameter α is taken as unity.

The elastic strain may be written in a similar form as

$$\frac{\epsilon_{ij}^e}{\epsilon_0} = (1+\nu)\frac{S_{ij}}{\sigma_0} + \frac{(1-2\nu)}{3}\frac{\sigma_{kk}}{\sigma_0}\delta_{ij} \tag{4.2.20}$$

The stress, strain and displacement variations are given by Shih [77] in the following form:

$$\sigma_{ij} = \sigma_0\left[\frac{J}{\alpha\sigma_0\epsilon_0 I_n r}\right]^{\frac{1}{n+1}}\tilde{\sigma}_{ij}(\theta, n) \tag{4.2.21}$$

$$\epsilon_{ij} = \alpha\epsilon_0\left[\frac{J}{\alpha\sigma_0\epsilon_0 I_n r}\right]^{\frac{n}{n+1}}\tilde{\epsilon}_{ij}(\theta, n) \tag{4.2.22}$$

$$u_i = \alpha\epsilon_0 r\left[\frac{J}{\alpha\sigma_0\epsilon_0 I_n r}\right]^{\frac{n}{n+1}}\tilde{u}_i(\theta, n) \tag{4.2.23}$$

where $\tilde{\sigma}_{ij}(\theta, n)$, $\tilde{\epsilon}_{ij}(\theta, n)$ and $\tilde{u}_i(\theta, n)$ are tabulated for θ in two–degree increments. The tabulated variations are available for either plane strain or plane stress, and for n ranging from 2 to 100. The parameter I_n is a normalizing factor that depends on whether the deformation is plane strain or plane stress.

For the present study, the hardening exponent n was chosen to match the experimentally measured uniaxial stress–strain curve with $n = 20$, and $\alpha = 1$. For $n = 20$, assuming plane strain, $I_n = 4.21$. For $n = 20$, assuming plane stress, $I_n = 2.74$.

4.2.8 Remarks on the HRR Field

Approximating the plastic deformation surrounding the crack with the HRR field presents difficulties similar to approximating the elastic deformation with the K field. As with the K field, the HRR field is derived for a two–dimensional body, while the deformation near the crack tip is strongly three–dimensional in character. Making a three–dimensional correction will be difficult for the HRR field, again because of the the singular nature of the stress and strain functions.

Plasticity at the crack tip is the basis for the HRR approximation; however, there is no guarantee that the plastic zone will ever become large enough so that the HRR approximation has any region of dominance. Additionally, the HRR field is derived using a deformation plasticity law, which does not in general match the incremental plasticity law used in the finite–element code. The incremental theory is usually considered to be more correct, as it models the physical body more realistically. Deformation and incremental theories coincide only when the loading path is proportional, although the two theories approach the same result when the loading develops in a definite direction [34]. The HRR field itself is a case of proportional loading, as is the K field. However, the HRR field and the K field have a different spatial dependence, so if the HRR field develops from the K field, the loading must pass through some transition from one to the other that is not proportional. This transition will take place at the edge of the plastic zone.

One interesting remark in Shih's work [77] is that the region in which the HRR field dominates for large–scale yielding conditions is strongly dependent on whether the uncracked ligament is subjected primarily to tension or bending. This is important, because the three–point–bend specimen geometry has a strong component of bending across the uncracked ligament. Actually, Shih [78] indicates that the HRR field will have a larger domain of validity if the uncracked ligament is subject to bending than if it is subjected to tension alone. (Their study considered a ligament that was completely plastic.) In part, this conclusion depends on the criterion for comparing the finite–element model to the HRR field. The HRR field was closer to the finite–element model with significant bending in a small region near the crack

tip, but it began to deviate from the numerical solution very quickly outside that small region. The deviation between the HRR field and the numerical model for tension only was larger in the same small region, but it seemed to reach a constant level outside that small region. It is likely that the K field (and associated higher–order terms) are also strongly dependent on the stress distribution across the uncracked ligament. The addition of higher–order terms in the HRR field to characterize the plastic deformation in a larger region would run into further difficulty because the assumption that the elastic strains are negligible would lose validity.

Hutchinson [verbal communication] suggests that the HRR field may be improved by adding the elastic strains that correspond to the HRR stresses. This addition may be particularly important in the case of the hydrostatic stress, since the plastic strains are independent of the hydrostatic stress.

In a recent paper, Zhang and Ravi–Chandar [93] have investigated the HRR field numerically, using several different constitutive laws and numerical models. The boundary conditions used apply a K field displacement around a circular geometry. The u_2 displacements calculated from this study match fairly well with the HRR field, but the u_1 displacements seem not to agree well at all. Hutchinson [verbal communication] suggests that the plots of log u_1 against log r are incorrect because the HRR field within the applied K field may have a constant offset of the location of the origin in the x_1 direction, similar to the effect of the finite notch tip in the K field.

Presumably, a correction could be made to the HRR field to account for the effects of a finite notch tip radius, as was made for the K field. It might be best to assume a form for the notch that is a long thin ellipse, and to add a stress term having a stronger singularity in r than the HRR field, which leaves the notch surface traction–free. This correction would be governed by the same multiplying function of J that governs the strength of the HRR field singularity. This added variation would also be a valid solution for the numerically determined stress potential.

4.3 Finite–Element Analysis

The finite–element model employed in the present work is adapted from the model used by Narasimhan, Rosakis and Moran [52] in a study of void growth as a fracture initiation and propagation mechanism. The void growth model has been removed from the code for the present work, as it caused convergence difficulties when the void volume fraction became large. The constitutive relation and mesh have been changed to match more closely the experimentally measured values of the experimental specimen used in the current study. Tunneling of the crack in the interior of the body has been included through the release of the boundary constraints on the crack plane. The choice of nodes to be released was determined through a measurement of the profile of the crack tunnel at several loads, as described in Section 4.5.

The finite–element code is a modified version of FEAP (Finite Element Analysis Program), which originated with R. L. Taylor at University of California, Berkeley, and M. Ortiz of Brown University.

4.3.1 Model Description

The three–point–bend experiment of Figure 87 was modeled using a three dimensional finite–element code with a power–hardening, elastic–plastic constitutive relation. Because of the symmetry of geometry and loading about the plane of the crack and about the midplane of the specimen, only one quarter of the specimen must be modeled.

The mesh geometry of the simulation is shown in Figures 88 through 90. The entire model consisted of 3200 eight–noded block elements arranged in 10 layers and having 3971 nodes and 11913 degrees of freedom. The physical specimen is a plate 30 cm long by 7.6 cm high by 1 cm thick. A notch of 3 cm in length is cut into one of the long sides of the plate. In order to match the physical specimen as closely as possible, the notch is modeled in the numerical simulation as having a circular tip of radius 0.015 cm. The notch in the steel specimen is cut with an electric

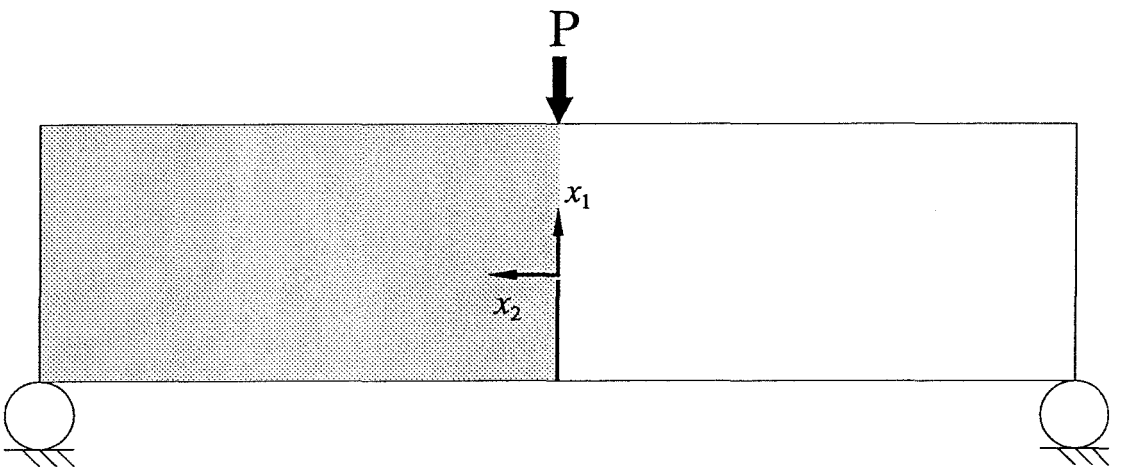


FIGURE 87. Diagram of the geometry of the three–point–bend experiment. Because of symmetry, only the shaded side of the specimen needs to be modeled. Also, symmetry about the midplane of the specimen means only one quarter of the entire specimen must be modeled.

discharge using a wire of 0.02 cm diameter, resulting in a notch of 0.03 cm width, and a rounded tip of approximately 0.015 cm radius. Detail of the mesh near the notch tip is shown in Figure 89. Ten layers of elements are used through the half thickness of the specimen. The elements vary through the thickness in a parabolic distribution, with the elements being smaller near the free surface to capture the corner singularity. The variation of the element size through the thickness is shown in Figure 90. In the region surrounding the notch tip, the elements are arranged in wedges of 15° each. The elastic modulus, yield stress and power–hardening exponent were calculated from the uniaxial tension test.

The physical parameters of the present model are intended to match as closely as possible the parameters of the steel specimen used in the interferometric investigation.

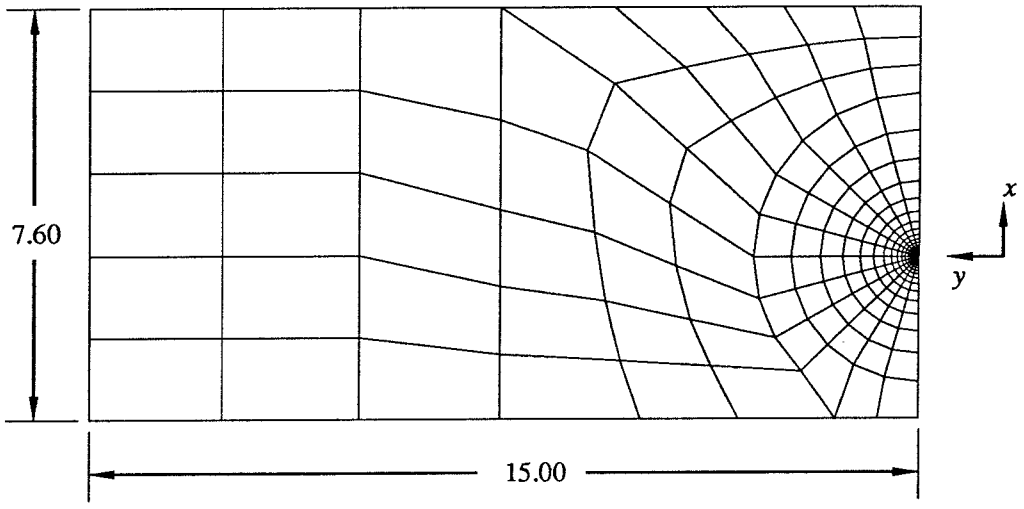


FIGURE 88. Diagram of the mesh in the finite–element model of the three–point–bend specimen. The notch is at the lower right, and is modeled as having a circular tip of radius 0.015 cm. The elements are concentrated near the notch tip. All dimensions are in centimeters.

4.3.2 Model Constitutive Relations

A brief discussion of the elastic and plastic constitutive relation used for this study is found in Section 3.2. The numerical model assumes material homogeneity and isotropy in both the elastic and plastic constitutive relations, and the numerical simulations are carried out using a small–strain, incremental plasticity theory. The stresses are computed and the yield surface updated using a tangential predictor – radial return method with subincrementation. The stress update procedure is described in Narasimhan and Rosakis [50,51]. The $\overline{\mathrm{B}}$ method of Hughes [30] modified by Nakamura, Shih and Freund [48] was used to relieve artificial mesh locking caused by plastic incompressibility. As recommended by Nakamura, Shih and Freund [48], a stabilization parameter of 0.05 was used in the calculations.

Both models used a power–law hardening relation between the stress and strain in the plastic regime. The initial yield stress of the first model was about 10% higher than that of the second model. The plastic behavior of the two models was quite similar, as the hardening exponent of the first model was 22 as compared to 23.7 in the second model. An exponent of 1 corresponds to linear elasticity, while an

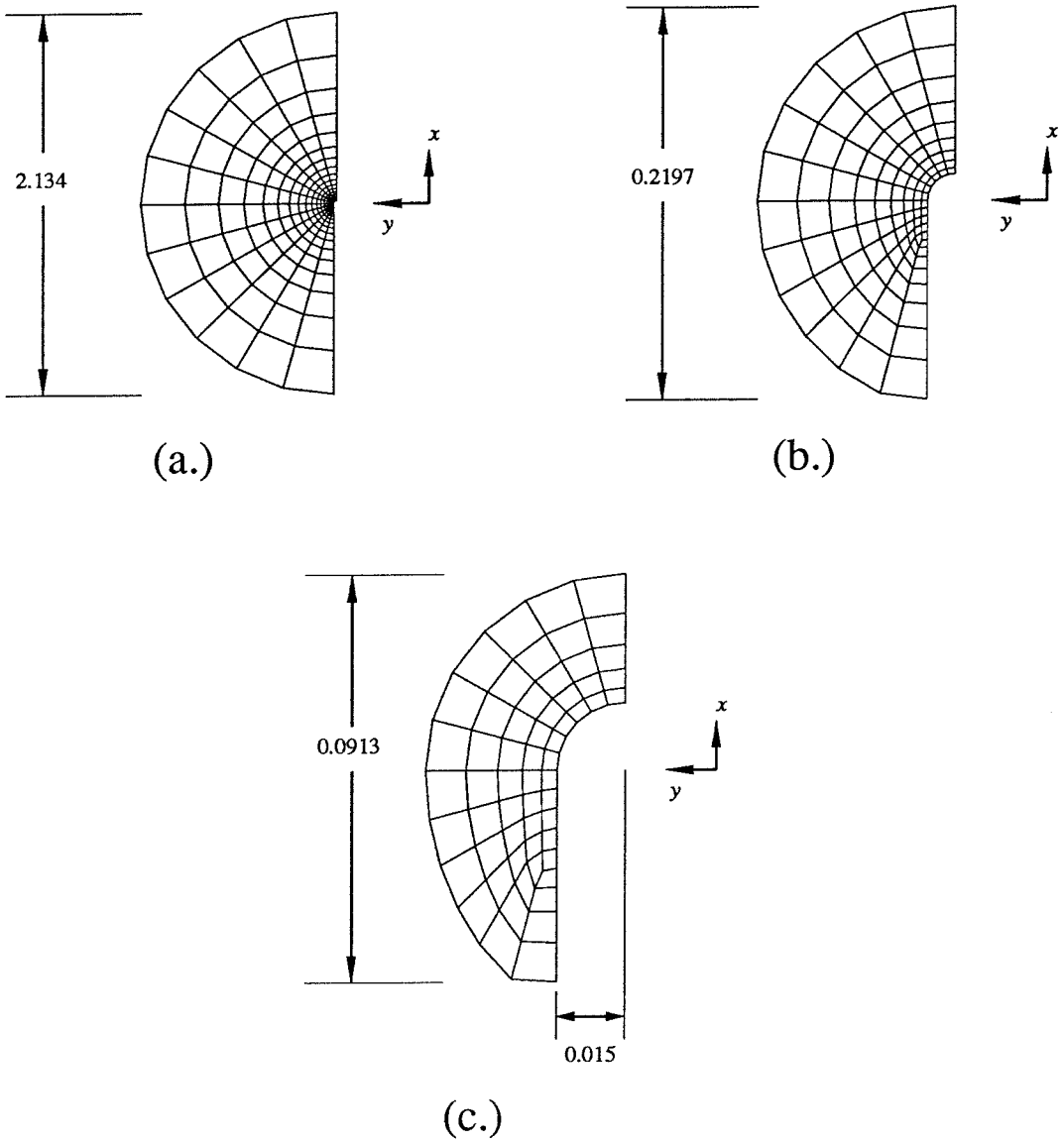


FIGURE 89. Detail of the mesh near the notch tip for the finite–element model of the three–point–bend specimen. The notch crack is modeled as having a circular tip of radius 0.015 cm, as shown in diagram (c.). All dimensions are in centimeters.

exponent of infinity corresponds to an elastic–perfectly plastic material.

Although the constitutive relations used in the numerical models contain no explicit time dependence (no rate–dependent effects, no dynamic effects and no

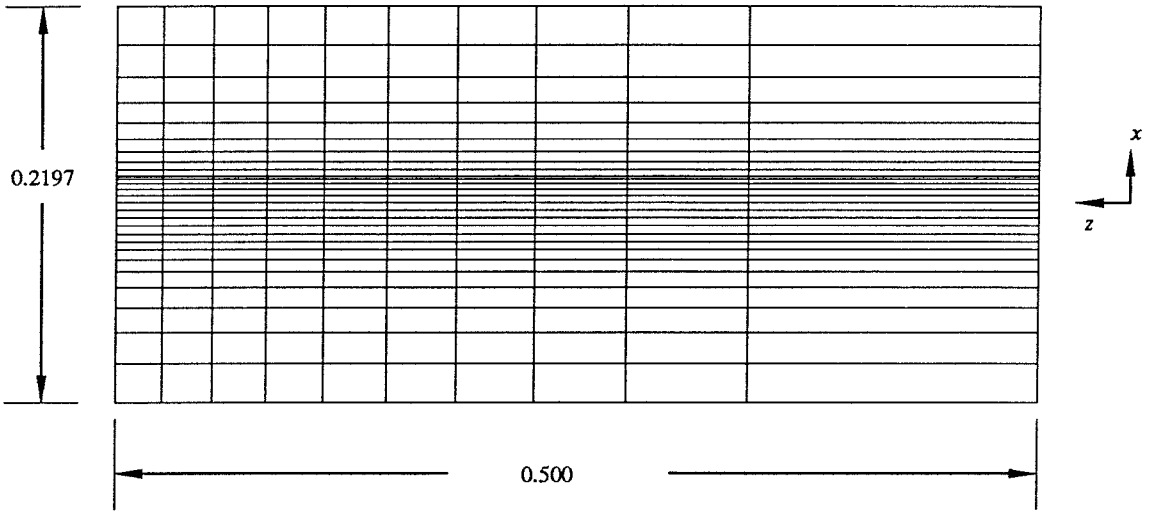


FIGURE 90. Detail of the element variation through the thickness of the finite–element model of the three–point–bend specimen. The notch tip is in the center of the diagram, running horizontally. The scale corresponds to diagram (b.) in Figure 89. Because of symmetry of the geometry and loading, only half of the specimen must be modeled. The free surface is at the left, the midplane at the right. All dimensions are in centimeters.

time–dependent effects such as creep), the solution of the finite–element equations is performed in a time domain to allow for the inclusion of such explicit time dependence. In this model the load history alone is explicitly a function of time, and time steps are used to implement load steps. The load history is monotonically increasing and is specified to be linear with time in this study.

4.3.3 Solution Approach

An iterative solution method is used to solve for the increments of the displacements, strains and stresses corresponding to the increment of the applied load at each time step.

It is assumed that an equilibrium solution for time t has been achieved. At time t, the displacements are held in the displacement matrix $\mathbf{U}(t)$. The strains at time t are calculated from the displacements through the $\overline{\mathbf{B}}$ matrix as

$$\boldsymbol{\epsilon}(t) = \overline{\mathbf{B}}\mathbf{U}(t) \tag{4.3.1}$$

The $\overline{\mathbf{B}}$ matrix is used in place of the normal $\mathbf{B}$ matrix to eliminate numerical error caused by mesh locking caused by plastic incompressibility. The normal $\mathbf{B}$ matrix is calculated from the derivatives of the shape functions for the element. The $\overline{\mathbf{B}}$ matrix is derived from the $\mathbf{B}$ matrix by dividing the $\mathbf{B}$ matrix into a deviatoric part and a volumetric part, $\mathbf{B} = \mathbf{B}_{dev} + \mathbf{B}_{vol}$. The volumetric part of the $\mathbf{B}$ matrix is replaced by a volumetric matrix calculated at a reduced number of quadrature points using reduced–order shape functions, $\overline{\mathbf{B}} = \mathbf{B}_{dev} + \overline{\mathbf{B}}_{vol}$. In elastic analysis, the $\overline{\mathbf{B}}$ method yields an identical stiffness matrix when compared to selective reduced integration in which the compliance matrix, $\mathbf{C}$, is separated into deviatoric and volumetric parts, and the volumetric part is numerically integrated at the reduced quadrature points. A stabilization parameter of 0.05 was used to create a $\overline{\mathbf{B}}$ matrix which lies between the usual $\mathbf{B}$ matrix and the full $\overline{\mathbf{B}}$ matrix, $\overline{\mathbf{B}} = \mathbf{B}_{dev} + \overline{\mathbf{B}}_{vol} + 0.05(\mathbf{B}_{vol} - \overline{\mathbf{B}}_{vol})$. This stabilization parameter is intended to eliminate spurious zero–energy modes that may occur because of the reduced integration, while still providing a strain–displacement matrix that does not significantly overestimate the stiffness of the element in the event of the nearly incompressible behavior in fully plastic deformations.

Since small strains are assumed in the model, the $\overline{\mathbf{B}}$ matrix remains constant throughout the loading history. The equilibrium stresses at time t are held in the matrix $\boldsymbol{\sigma}(t)$, and the nodal force vector $\mathbf{P}(t)$ is calculated from the stresses through the $\overline{\mathbf{B}}$ matrix in a numerically calculated integral over the volume of the model

$$\mathbf{P}(t) = \int_{Volume} \overline{\mathbf{B}}\, \boldsymbol{\sigma}(t) \quad dV \tag{4.3.2}$$

The applied nodal force vector $\mathbf{F}(t)$ is specified by the load history. Gross equilibrium of the model requires that

$$\mathbf{F}(t) = \mathbf{P}(t) \tag{4.3.3}$$

for all time t. The tangent stiffness matrix $\mathbf{K}(t)$ at time t is calculated by numerical integration over the volume

$$\mathbf{K}(t) = \int_{Volume} \overline{\mathbf{B}}^T \mathbf{C}^*(t)\, \overline{\mathbf{B}} \quad dV \tag{4.3.4}$$

where $\mathbf{C}^*(t)$ is the history–dependent compliance matrix, derived from the history–dependent compliance four–tensor. The superscript T indicates the transpose of the matrix. As described in the section on elastic and plastic constitutive relations, $\mathbf{C}^*(t)$ depends on the accumulated plastic strain as well as on the current stress and strain states. For purely elastic behavior, $\mathbf{C}^*$ reduces to the elastic compliance $\mathbf{C}$.

The time is incremented to $t+\Delta t$, where Δt is a suitably chosen time increment. At time $t + \Delta t$, the externally applied forces $\mathbf{F}(t + \Delta t)$ must still balance with the forces that are due to the element stresses $\mathbf{P}(t + \Delta t)$,

$$\mathbf{F}(t + \Delta t) = \mathbf{P}(t + \Delta t) \tag{4.3.5}$$

The forces that are due to the element stresses at time $t + \Delta t$ are decomposed into the known forces at time t, $\mathbf{P}(t)$, and the increment at time $t + \Delta t$, $\Delta\mathbf{P}$. The increment of the nodal forces that are due to the element stresses is approximated through use of the tangent–stiffness matrix of time t, $\mathbf{K}(t)$, multiplying the displacement increment at time $t + \Delta t$, $\Delta\mathbf{U}$.

$$\Delta\mathbf{P} \approx \mathbf{K}(t)\Delta\mathbf{U} \tag{4.3.6}$$

Thus, the equation that must be solved for the displacement increment $\Delta\mathbf{U}$ at time $t + \Delta t$ is

$$\begin{aligned} \mathbf{K}(t)\Delta\mathbf{U} &= \mathbf{F}(t + \Delta t) - \mathbf{P}(t) \\ &= \Delta\mathbf{F} \end{aligned} \tag{4.3.7}$$

The stiffness matrix is inverted and the matrix equation is solved using Gaussian elimination, leading to an approximation for the displacement increment $\Delta\mathbf{U}$.

Equation (4.3.7) is a first approximation to the equilibrium solution at time $t+\Delta t$. In general, the actual increment of the nodal forces corresponding to the displacement increments will not satisfy the equilibrium condition of equation (4.3.5). In order to refine the solution, an iteration procedure is used.

4.3.4 BFGS Iteration

The solution for the displacement increment at time $t + \Delta t$ in equation (4.3.7) is used as a first approximation in an iterative scheme to improve the solution. This iteration method is called the BFGS method (after Broyden, Fletcher, Goldfarb and Shanno), as described in Bathe [9], and it was suggested for use in finite elements by Matthies and Strang [43]. This method is labelled a quasi-Newton approach because it updates the stiffness matrix (or actually its inverse) in a limited way at each iteration to improve speed of convergence. A full Newton method updates the stiffness matrix completely at each iteration, while a modified Newton method does not update the stiffness matrix at all during iteration.

The matrix equation to be solved is still as shown in equation (4.3.7); however, it is now done in an iterative procedure. At the i^{th} iteration,

$$\begin{aligned}\Big(\mathbf{K}(t+\Delta t)\Big)^{(i-1)}\Delta\mathbf{U}^{(i)} &= \mathbf{F}(t+\Delta t) - \Big(\mathbf{P}(t+\Delta t)\Big)^{(i-1)} \\ &= \Delta\mathbf{F}^{(i-1)}\end{aligned} \tag{4.3.8}$$

where $(\mathbf{K}(t+\Delta t))^{(0)} = \mathbf{K}(t)$ and $(\mathbf{P}(t+\Delta t))^{(0)} = \mathbf{P}(t)$ are from the equilibrium solution at time t. $\Delta\mathbf{F}^{(i-1)}$ is the out-of-balance force vector, which becomes zero (or very small) when equilibrium is satisfied.

As mentioned, the BFGS method effectively acts to update the inverse of the stiffness matrix at each iteration. Actually, the stiffness matrix is not inverted, but is factored. The BFGS method uses vector multiplications and the factors of the original stiffness matrix to solve for the displacement increment at each iteration. The effect of the vector multiplications is identical to updating the inverse of the stiffness matrix. Rewriting equation (4.3.8) to use the inverse form of the stiffness matrix,

$$\begin{aligned}\Delta\mathbf{U}^{(i)} &= \Big(\mathbf{K}^{-1}(t+\Delta t)\Big)^{(i-1)}\Big[\mathbf{F}(t+\Delta t) - \Big(\mathbf{P}(t+\Delta t)\Big)^{(i-1)}\Big] \\ &= \Big(\mathbf{K}^{-1}(t+\Delta t)\Big)^{(i-1)}\Delta\mathbf{F}^{(i-1)}\end{aligned} \tag{4.3.9}$$

The total displacement is then approximated by

$$\mathbf{U}^{(i)}(t+\Delta t) = \mathbf{U}^{(i-1)}(t+\Delta t) + \beta\Delta\mathbf{U}^{(i)} \tag{4.3.10}$$

with $\mathbf{U}^{(0)}(t+\Delta t) = \mathbf{U}(t)$. The scalar parameter β is varied until the component of the new out-of-balance load vector parallel to the vector $\Delta\mathbf{U}^{(i)}$ is small compared to some tolerance. This is accomplished by varying β until the calculated nodal force vector, $\left(\mathbf{P}(t+\Delta t)\right)^{(i)}$, which corresponds to β, is such that the inner product

$$\left(\Delta\mathbf{U}^{(i)}\right)^T\left[\mathbf{F}(t+\Delta t) - \left(\mathbf{P}(t+\Delta t)\right)^{(i)}\right] = \left(\Delta\mathbf{U}^{(i)}\right)^T\Delta\mathbf{F}^{(i)} \tag{4.3.11}$$

is less than some tolerance. The superscript T denotes the transpose of the vector.

The new out-of-balance force vector depends on the parameter β through the constitutive law. Increasing β increases the displacement increment and therefore the strain increment. The stress increment is a nonlinear function of the strain increment, and the new nodal forces are calculated from the stresses as in equation (4.3.2), which is also put in iterative form

$$\left(\mathbf{P}(t+\Delta t)\right)^{(i)} = \int_{Volume} \overline{\mathbf{B}}\left(\boldsymbol{\sigma}(t+\Delta t)\right)^{(i)}\, dV \tag{4.3.12}$$

The procedure for calculating the stresses corresponding to the displacement increment $\beta\Delta\mathbf{U}$ are described in the next section.

As described in Matthies and Strang [43], the BFGS method is an attempt to make an efficient iterative search for the solution $\Delta\mathbf{F} = \mathbf{0}$. The variation of β produces a search direction which is nearly orthogonal to the out-of-balance force $\Delta\mathbf{F}^{(i)}$. In particular, the updating of the inverse of the stiffness matrix is made inexpensive to compute.

Two difference vectors, $\boldsymbol{\delta}^{(i)}$ and $\boldsymbol{\gamma}^{(i)}$, are formed to represent the change in the displacements and out-of-balance forces between increments. These vectors are

$$\boldsymbol{\delta}^{(i)} = \left(\mathbf{U}(t+\Delta t)\right)^{(i)} - \left(\mathbf{U}(t+\Delta t)\right)^{(i-1)} \tag{4.3.13}$$

$$\boldsymbol{\gamma}^{(i)} = \Delta\mathbf{F}^{(i-1)} - \Delta\mathbf{F}^{(i)} \tag{4.3.14}$$

The updated inverse of the stiffness matrix, $(\mathbf{K}^{-1}(t+\Delta t))^{(i)}$, will be made to satisfy the quasi-Newton equation

$$\boldsymbol{\delta}^{(i)} = \left(\mathbf{K}^{-1}(t+\Delta t)\right)^{(i)}\boldsymbol{\gamma}^{(i)} \tag{4.3.15}$$

In the BFGS method, the update of the inverse of the stiffness matrix is accomplished through a matrix multiplication of the form

$$(\mathbf{K}^{-1}(t+\Delta t))^{(i)} = (\mathbf{A}^T)^{(i)}\Big(\mathbf{K}^{-1}(t+\Delta t)\Big)^{(i-1)}\mathbf{A}^{(i)} \tag{4.3.16}$$

where $\mathbf{A}^{(i)}$ is a square matrix calculated from $\boldsymbol{\delta}^{(i)}$ and $\boldsymbol{\gamma}^{(i)}$. $\mathbf{A}^{(i)}$ is defined by

$$\mathbf{A}^{(i)} = \mathbf{I} + \mathbf{v}^{(i)}(\mathbf{w}^{(i)})^T \tag{4.3.17}$$

where $\mathbf{I}$ is the identity matrix. The vectors $\mathbf{v}^{(i)}$ and $\mathbf{w}^{(i)}$ are found as

$$\mathbf{v}^{(i)} = -\left[\frac{(\boldsymbol{\delta}^{(i)})^T\boldsymbol{\gamma}^{(i)}}{(\boldsymbol{\delta}^{(i)})^T\Big(\mathbf{K}(t+\Delta t)\Big)^{(i-1)}\boldsymbol{\delta}^{(i)}}\right]^{1/2}\Big(\mathbf{K}(t+\Delta t)\Big)^{(i-1)}\boldsymbol{\delta}^{(i)} - \boldsymbol{\gamma}^{(i)} \tag{4.3.18}$$

$$\mathbf{w}^{(i)} = \frac{\boldsymbol{\delta}^{(i)}}{(\boldsymbol{\delta}^{(i)})^T\boldsymbol{\gamma}^{(i)}} \tag{4.3.19}$$

The term in brackets in equation (4.3.18) is the condition number of the update, $c^{(i)}$, which is checked before the update of the stiffness matrix inverse is made. The condition number at the i^{th} iteration is

$$c^{(i)} = \left[\frac{(\boldsymbol{\delta}^{(i)})^T\boldsymbol{\gamma}^{(i)}}{(\boldsymbol{\delta}^{(i)})^T\Big(\mathbf{K}(t+\Delta t)\Big)^{(i-1)}\boldsymbol{\delta}^{(i)}}\right]^{1/2} \tag{4.3.20}$$

If the condition number becomes very large, the update in equation (4.3.16) would be numerically dangerous, so a preset tolerance is used to prevent updating if the condition number exceeds the tolerance.

With the newly calculated matrices $(\mathbf{K}^{-1}(t+\Delta t))^{(i)}$ and $(\mathbf{P}(t+\Delta t))^{(i)}$, the process can begin again at equation (4.3.9) to evaluate the displacement increment at iteration $i+1$.

The iteration for the displacement and force increments continues until the Euclidean norm of the out-of-balance force vector $\Delta\mathbf{F}^{(i)}$ is smaller than the out-of-balance force at the start of the iteration by a preset tolerance.

$$\|\Delta\mathbf{F}^{(i)}\| \leq \text{TOL}\ \|\Delta\mathbf{F}^{(1)}\| \tag{4.3.21}$$

The displacement at time $t + \Delta t$, $\mathbf{U}(t + \Delta t)$, is then given by

$$\mathbf{U}(t + \Delta t) = \mathbf{U}(t) + \sum_{i=1}^{n} \Delta \mathbf{U}^{(i)} \tag{4.3.22}$$

where n is the number of iterations to convergence.

4.3.5 Stress Update

The stresses in the elements are computed from the calculated strain increment using a tangential predictor – radial return method. In order to prevent artificial unloading during the iteration procedure, the stresses and strains at each iteration are calculated from the known equilibrium solution from the previous time step and the total displacement increment for the current time step. At the i^{th} iteration for time step $t + \Delta t$, the total displacement increment, $\Delta \mathbf{U}^{total}$, is the sum of the displacement increments for iterations 1 through i for time step $t + \Delta t$. The strain increment that is due to that displacement increment is given by

$$\Delta \boldsymbol{\epsilon} = \overline{\mathbf{B}}\ \Delta \mathbf{U}^{total} \tag{4.3.23}$$

In calculating the stress increment for the current iteration, this total strain increment is used along with the previous equilibrium solution to prevent the possibility that the stress increment at one iteration will lie outside the current yield surface, while the next stress increment will be a correction in the opposite direction, causing artificial elastic unloading that is due to finite step sizes.

The procedure for updating the stress and yield surface requires only the constant elastic compliance matrix, C, and the current parameter defining the yield surface, $\overline{\sigma}$. First, an elastic stress increment is determined from the strain increment through the elastic compliance tensor C.

$$\Delta \boldsymbol{\sigma}^{e} = \mathbf{C}\ \Delta \boldsymbol{\epsilon} \tag{4.3.24}$$

A trial stress state $\boldsymbol{\sigma}^{A}$ is calculated from the stress increment and the stress state from the known equilibrium solution at time t, $\boldsymbol{\sigma}^{0}$.

$$\boldsymbol{\sigma}^{A} = \boldsymbol{\sigma}^{0} + \Delta \boldsymbol{\sigma}^{e} \tag{4.3.25}$$

The stress state $\boldsymbol{\sigma}^0$ is assumed to lie inside the yield surface at time t. The yield surface at time t is defined by the function $f(\boldsymbol{\sigma}, \boldsymbol{\epsilon}) = 3\mathrm{J}_2(\boldsymbol{\sigma}) - \overline{\sigma}^2$, where $\overline{\sigma}$ is a function of the accumulated equivalent plastic strain, and is a measure of the current yield stress. At time t, $\overline{\sigma} = \overline{\sigma}^0$. If the trial stress state $\boldsymbol{\sigma}^A$ is such that

$$3\mathrm{J}_2(\boldsymbol{\sigma}^A) \leq (\overline{\sigma}^0)^2 \tag{4.3.26}$$

the stress path has stayed in the elastic regime, and no further calculation of the stress state is necessary. Otherwise, the yield surface has been crossed by the trial stress solution, and the yield surface and stress state must be updated together. A stress state between the equilibrium stress of time t, $\boldsymbol{\sigma}^0$ and the trial stress state $\boldsymbol{\sigma}^A$ is found, which lies on the yield stress of time t. This contact stress state, $\boldsymbol{\sigma}^C$, is defined by

$$\boldsymbol{\sigma}^C = \boldsymbol{\sigma}^0 + q\boldsymbol{\sigma}^A \tag{4.3.27}$$

and

$$3\mathrm{J}_2(\boldsymbol{\sigma}^C) = (\overline{\sigma}^0)^2 \tag{4.3.28}$$

For the isotropic hardening material satisfying the Huber–Von Mises yield condition, this results in a quadratic equation for the scalar parameter q. Note that the path from $\boldsymbol{\sigma}^0$ to $\boldsymbol{\sigma}^C$ is completely elastic.

A corrected trial stress $\boldsymbol{\sigma}^B$ is determined from the contact stress $\boldsymbol{\sigma}^C$ using the assumption that the plastic strain rate is always normal to the yield surface. If the plastic strain rate is directed along a normal to the Huber–Von Mises yield surface, it will be given by

$$\dot{\boldsymbol{\epsilon}}^p = \dot{\lambda}\mathbf{S} \tag{4.3.29}$$

where $\mathbf{S}$ is the deviatoric stress and the invariant $\mathrm{J}_2 = \frac{1}{2}S_{ij}S_{ij}$. An approximate plastic strain increment is given by $\Delta\lambda\mathbf{S}$, where $\mathbf{S}$ and $\Delta\lambda$ are calculated at the contact stress $\boldsymbol{\sigma}^C$. This plastic strain increment is subtracted from the calculated strain increment from the stress state $\boldsymbol{\sigma}^C$ to the stress state $\boldsymbol{\sigma}^A$, approximated by $(1-q)\Delta\boldsymbol{\epsilon}$, to estimate the elastic strain increment during plastic deformation.

$$\Delta\boldsymbol{\epsilon}^e \approx (1-q)\Delta\boldsymbol{\epsilon} - \Delta\lambda\mathbf{S} \tag{4.3.30}$$

The stress increment corresponding to that elastic strain increment is

$$\Delta\boldsymbol{\sigma} = \mathbf{C}\big((1-q)\Delta\boldsymbol{\epsilon} - \Delta\lambda\mathbf{S}\big) \tag{4.3.31}$$

The corrected trial stress $\boldsymbol{\sigma}^B$ is the contact stress $\boldsymbol{\sigma}^C$ plus the calculated stress increment.

$$\begin{aligned}\boldsymbol{\sigma}^B &= \boldsymbol{\sigma}^C + \mathbf{C}\big((1-q)\Delta\boldsymbol{\epsilon} - \Delta\lambda\mathbf{S}\big)\\ &= \boldsymbol{\sigma}^C + \big((\boldsymbol{\sigma}^A - \boldsymbol{\sigma}^C) - \Delta\lambda\mathbf{C}\,\mathbf{S}\big)\\ &= \boldsymbol{\sigma}^A - \Delta\lambda\mathbf{C}\,\mathbf{S}\end{aligned} \tag{4.3.32}$$

The yield surface is updated using the hardening law H, in this case a power law. The accumulated equivalent plastic strain increment $\Delta\overline{\epsilon}^p$ is given by

$$\begin{aligned}\Delta\overline{\epsilon}^p &= \big(\tfrac{2}{3}\Delta\epsilon^p_{ij}\Delta\epsilon^p_{ij}\big)^{\frac{1}{2}}\\ &= \Delta\lambda\big(\tfrac{2}{3}S_{ij}S_{ij}\big)^{\frac{1}{2}}\\ &= \Delta\lambda\big(\tfrac{2}{3}2\mathrm{J}_2\big)^{\frac{1}{2}}\\ &= \tfrac{2}{3}\Delta\lambda\overline{\sigma}^0\end{aligned} \tag{4.3.33}$$

H is defined as the rate of change of the function $\overline{\sigma}$ with respect to the equivalent accumulated plastic strain $\overline{\epsilon}^p$, which is determined from the power law described in the section on constitutive behavior. In order to update the yield surface, H will be determined at $\overline{\sigma}^0$.

$$H(\overline{\sigma}^0) = \left.\frac{d\overline{\sigma}}{d\overline{\epsilon}^p}\right|_{\overline{\sigma}=\overline{\sigma}^0} \tag{4.3.34}$$

The final value of the yield stress, $\overline{\sigma}^F$, is then calculated as

$$\overline{\sigma}^F = \overline{\sigma}^0 + H(\overline{\sigma}^0)\Delta\overline{\epsilon}^p \tag{4.3.35}$$

In general, the corrected trial stress $\boldsymbol{\sigma}^B$ will not lie on the updated surface. A final update of the stress state is scaled from $\boldsymbol{\sigma}^B$, forcing it to lie on the updated yield surface

$$\boldsymbol{\sigma}^F = \left(\frac{\overline{\sigma}^F}{\sqrt{3\mathrm{J}_2(\boldsymbol{\sigma}^B)}}\right)\boldsymbol{\sigma}^B \tag{4.3.36}$$

The stress path from $\boldsymbol{\sigma}^C$ to $\boldsymbol{\sigma}^F$ is an elastic–plastic path. To minimize the effect of finite increments, the excess stress increment $\boldsymbol{\sigma}^F - \boldsymbol{\sigma}^C$ is divided into m subincrements and the corrected trial stress $\boldsymbol{\sigma}^B$, the yield surface update, and the final correction to $\boldsymbol{\sigma}^F$ are calculated m times.

4.3.6 Inclusion of Crack Tunneling

There are several possible sources of discrepancy between the experiment and the finite–element code. One of these discrepancies that could be addressed was the tunneling of the crack within the specimen used in the experiment. The tunneling of the crack within the experimental body was measured using a multispecimen approach, as outlined in Section 4.5. The measured shape of the crack tunnels at several loads are also shown in Section 4.5.1. Before these tests were made, it was thought that the crack might have tunneled very deeply into the interior of the body before ultimate failure (as much as four or five millimeters), or that crack tunneling might actually have begun at a low load, meaning that the finite–element model was not modeling the physical body used in the experiment over a wide range of load. Fortunately, that seems not to be the case. Tunneling was estimated to initiate at about 63500 Newtons, and the highest load achieved in the experiment was 73500 Newtons, with one load step in between at 68500 Newtons. The depth of the tunnel was about 0.8 millimeter in the center of the specimen at a load of 73500 Newtons, as compared to the notch tip radius of 0.15 mm and a plate thickness of 10 mm. The plastic zone is also about 10 mm in size at the highest loads.

The advance of the crack was incorporated in the finite–element model through the release of the displacement constraint on the nodes corresponding to the shape of the tunneled crack. No tunneling was found up to a load of 63500 Newtons, and tunnel profiles were measured for loads of 68500 Newtons and 73500 Newtons, with 73500 Newtons being the highest load at which data were taken from the experiment. The experimentally determined tunnel profiles are shown in Section 4.5.1. It was necessary to remove constraints on twenty-two nodes to approximate the shape of the tunnel at 68500 Newtons, and to remove constraints on a further thirteen nodes to approximate the tunnel shape at 73500 Newtons. The nodes were released in four groups. Ten nodes were released on loading from 63500 Newtons to 66000 Newtons; twelve more nodes were released between 66000 and 68500 Newtons; six more nodes were released between 68500 and 71000 Newtons; and the final seven nodes were released between 71000 and 73500 Newtons. The loading in each case took place in five increments of 500 Newtons each, each group of nodes was released

before the first 500 Newton increment was applied.

The sudden removal of the constraint on the nodes caused a significant increase in the number of iterations required for convergence for the first load increment. For example, in the first step after releasing nodes, from 63500 Newtons to 64000 Newtons, twenty-nine iterations were required for convergence, while the average number of iterations needed for steps up to 63500 Newtons was about five.

Data are available from the numerical model with tunneling at 66000, 68500, 71000 and 73500 Newtons. Additionally, the code was run up to 73500 Newtons *without* the tunnel in an effort to be able to differentiate the effects of the tunnel from other effects, such as plasticity. Data from the model without tunneling are also available at 68500 and 73500 Newtons.

4.3.7 J–Integral Calculation

For a crack or notch in a three–dimensional body, the J–integral of equation (4.2.1) is defined as a surface integral over a cylindrical contour that surrounds the entire notch tip. The scheme used for evaluating the three–dimensional J–integral implemented in the numerical code was to calculate a local, two–dimensional value of the J–integral, $\hat{J}(x_3)$, at each nodal layer, and then a thickness average of the two–dimensional values was used as the value of J for the three–dimensional body. This procedure made it possible to examine the variation in the local value of J, $\hat{J}(x_3)$, at different locations through the thickness of the specimen, as discussed in the papers by Narasimhan and Rosakis [51] and Narasimhan, Rosakis and Moran [52].

The method for calculating the local two–dimensional value of J follows the domain integral approach of Li, Shih and Needleman [40]. See also the papers by Nakamura, Shih and Freund [47,48] and by Shih, Moran and Nakamura [80], for more detail. With this approach, the contour integral of equation (4.2.1) is transformed into an area integral through the use of a suitable, smooth weighting function. Referring to Figure 91, the contour integral of equation (4.2.1) may be

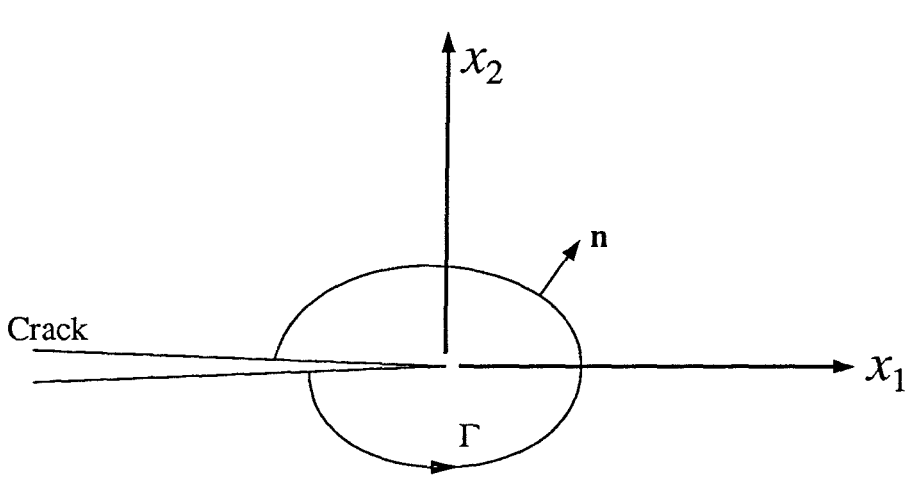


FIGURE 91. A sample contour Γ used to evaluate the J–integral. $\mathbf{n}$ is the unit outward normal. Any contour that contains only the singularity at the crack tip will yield the same value for J as long as the material is elastic and the crack surfaces are traction–free.

rewritten slightly as

$$J = \int_{\Gamma} \left(W\delta_{1j} - \sigma_{ij} \frac{\partial u_i}{\partial x_1} \right) n_j ds \tag{4.3.37}$$

where the contour Γ is traversed as shown in the figure, and $\mathbf{n}$ is the unit normal as shown. W is the strain–energy density, ds is the element of arc length and δ_{ij} is the Kroneker delta. A second contour Γ_2 is introduced, as shown in Figure 92, along with a sufficiently smooth weighting function q_1. The weighting function is such that $q_1 = 1$ on Γ and $q_1 = 0$ on Γ_2. Let C be the closed contour consisting of the curves Γ_2, $C+$, Γ and $C-$. C is traversed as shown in Figure 92. It should be noted that the direction travelled along the original contour Γ has been reversed in evaluating the path integral. Let A be the area enclosed by the closed contour C; $\mathbf{m}$ is the unit outward normal for area A. Note that along Γ, the direction of unit normal $\mathbf{m}$ is also opposite to the original normal $\mathbf{n}$.

The J–integral of equation (4.3.37) may then be written as

$$J = \oint_C \left(\sigma_{ij} \frac{\partial u_i}{\partial x_1} - W\delta_{1j} \right) m_j q_1 ds - \int_{C+} \sigma_{2j} \frac{\partial u_i}{\partial x_1} q_1 ds - \int_{C-} \sigma_{2j} \frac{\partial u_i}{\partial x_1} q_1 ds \tag{4.3.38}$$

Since the crack faces are traction–free, the second and third terms of equation (4.3.38) are zero. Applying the divergence theorem to the first term of equation

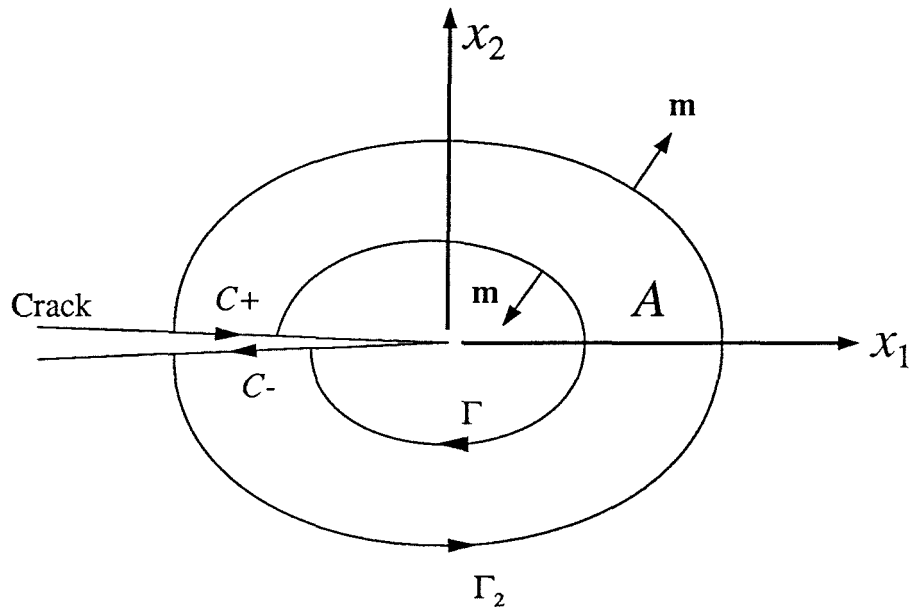


FIGURE 92. A second contour, Γ_2, is introduced to change the contour integral into an area integral. Let C be the closed contour composed of the four curves Γ_2, $C+$, Γ and $C-$, traversed in the sense of the arrows. Let A be the area enclosed by C. Note that Γ is traversed in the opposite sense from Figure 91, and that the outward unit normal to A, $\mathbf{m}$, is also in the opposite direction to $\mathbf{n}$ along Γ.

(4.3.38), the two–dimensional J is found to be

$$J = \int_A \left(\sigma_{ij} \frac{\partial u_i}{\partial x_1} - W \delta_{1j} \right) \frac{\partial q_1}{\partial x_j} dA \qquad (4.3.39)$$

As J is path–independent, the area integral of equation (4.3.39) is area–independent, provided the function q_1 satisfies the requirements stated above, namely, that q_1 is equal to unity on Γ and is zero on Γ_2. Equation (4.3.39) is a domain integral that may be evaluated using the same integration procedures as are employed to evaluate the stiffness matrix in the finite–element code, using the same shape functions over the elements that lie within the area A.

In the implementation of this scheme in the finite–element model, the contour Γ was taken to the tip of the crack. At each of the eleven nodal layers through the thickness, the local value was calculated for five different choices of the contour Γ_2, and an average of the resulting five J–integral values was used as the local $\hat{J}(x_3)$ at

each layer. The five contours were four circular contours centered at the origin of the semicircle of the notch tip, having radii of length 0.3 cm, 0.5 cm, 1.0 cm and 2.0 cm, and a square contour with sides of length 6.0 cm, also centered at the origin of the semicircle of the notch tip. (The plate thickness is 1.0 cm, and the notch has a semicircular tip of radius 0.015 cm.) The origin of the coordinate system is at the center of curvature of the notch. For the circular contour, the weighting function q_1 was taken as a linear function in r, a cone shape, within the contour, and zero outside the contour. In this case, $q_1(r) = 1 - r/R$ for $r < R$ and $q_1(r) = 0$ for $r \geq R$, which has the required property that $q_1 = 1$ at $r = 0$, and $q_1 = 0$ at $r = R$, with R being the external radius of each of the various circular contours. For the square contour, the weighting function was taken as a bilinear function of x_1 and x_2 with $q_1(x_1, x_2) = (1 - x_1/L_1)(1 - x_2/L_2)$ for $x_1 < L_1$ and $x_2 < L_2$, and $q_1(x_1, x_2) = 0$ for $x_1 \geq L_1$ or $x_2 \geq L_2$. The rectangular contour thus has dimensions $2L_1$ by $2L_2$.

The values of $\hat{J}(x_3)$ vary through the thickness, as discussed in detail in the paper by Narasimhan and Rosakis [51] and in the paper by Narasimhan, Rosakis and Moran [52]. The largest value of $\hat{J}(x_3)$ is found at the specimen midplane, and the lowest value at the free surface, except for the cases when the crack tunneling is included. The variation of the value of $\hat{J}$ normalized by the average value of J is shown in Figure 93 for loads from 4.0 kN to 73.5 kN. These are data from calculations done with no tunneling included.

The variation of $\hat{J}$ itself through the thickness is shown in Figure 94 for loads from 18.5 kN to 73.5 kN. Again, no tunneling was included in the model for these calculations.

Of course, the tunneling of the crack invalidates the J–integral calculation, because of the unloading that occurs near the crack tip. The unloading occurs following a linearly elastic constitutive law, so that the assumption used in deriving the J–integral, namely, a deformation plasticity constitutive law equivalent to nonlinear elasticity, is no longer applicable. Thus, after tunneling, this calculation procedure can no longer be strictly related to the energy–release rate at the crack tip. It is interesting, however, to compare the results of the calculation for the

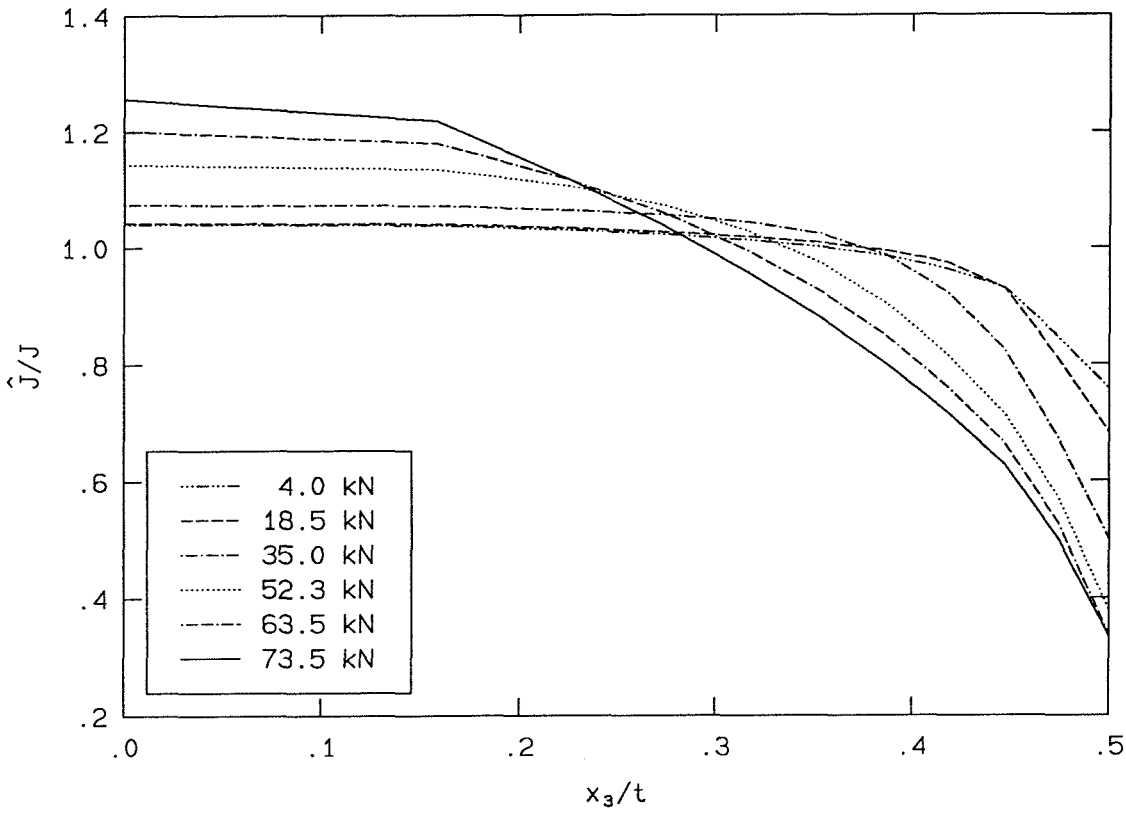


FIGURE 93. The variation of $\hat{J}$ through the thickness normalized by the average value of J at each load. No tunneling was included for these calculations.

two cases, one including tunneling and one without tunneling. Figure 95 shows the results of the domain integral calculation for $\hat{J}(x_3)$ at three different load steps for the case including tunneling compared to the case with no tunneling. One curve represents the calculation at 63.5 kN, which is the highest load for which there was no tunneling, and there are two curves each at 68.5 kN and 73.5 kN, one curve for the case with tunneling, and one without.

The values calculated at each of the five contours did differ slightly, most likely because of the dissipation involved by using the incremental plasticity formulation and a deviation from strictly proportional loading, although some part of the discrepancy may be caused by the discretization. Use of a deformation theory plasticity

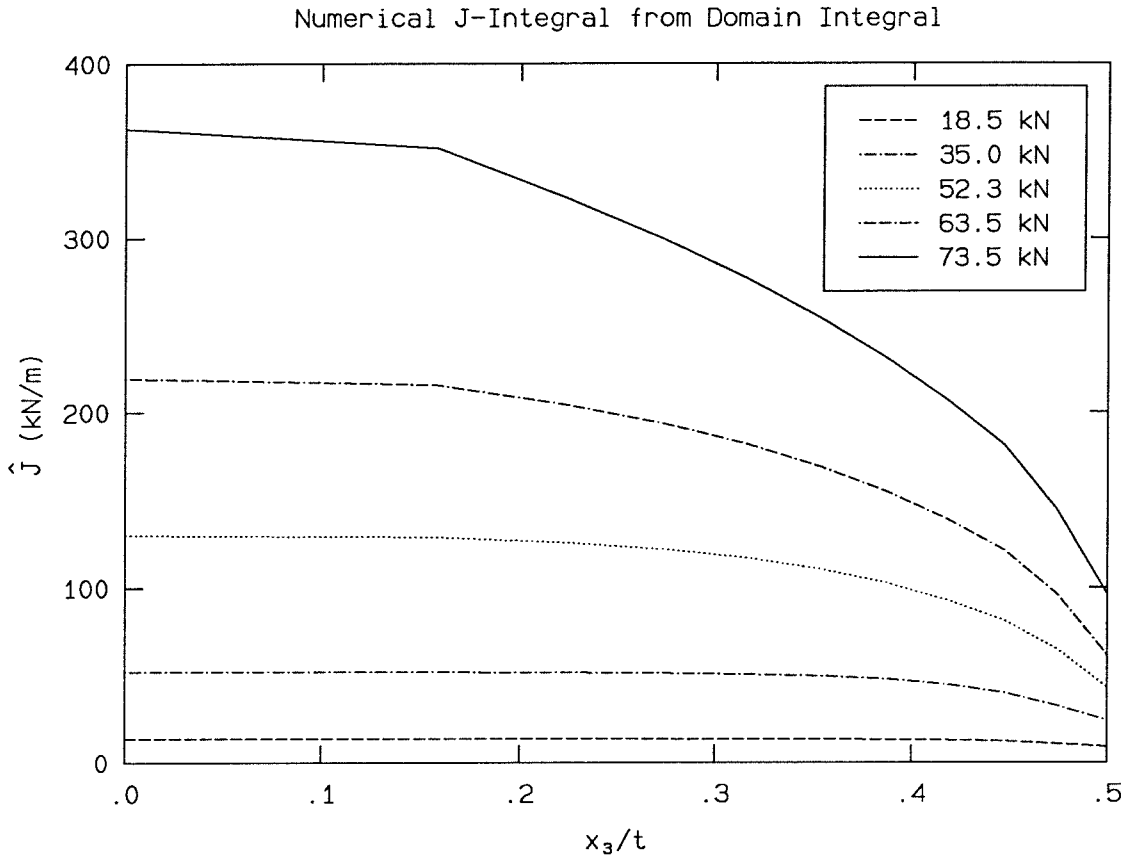


FIGURE 94. The variation of $\hat{J}$ through the thickness. No tunneling was included for these calculations.

formulation, equivalent to a nonlinear elasticity theory, should have preserved the path independence of the local J values. As expected, the agreement among the five contours is best at lower loads, before any plastic deformation has begun. At higher loads, the values at the different contours agree better at the midplane of the specimen than at the free surface. The circular contour of radius 0.3 cm always gave the smallest value for the local J, and the square contour always gave the largest value. At a load of 100 N, the values at the five different contours vary by less than 4% at the midplane, and by less than 6% at the free surface. At 63500 N, the the values at the five contours differ by less than 6% at the midplane, and by about 11% at the free surface. At 73500 N, in the case without tunneling, the values at the five different contours vary by about 5% at the midplane, but by about 19% at

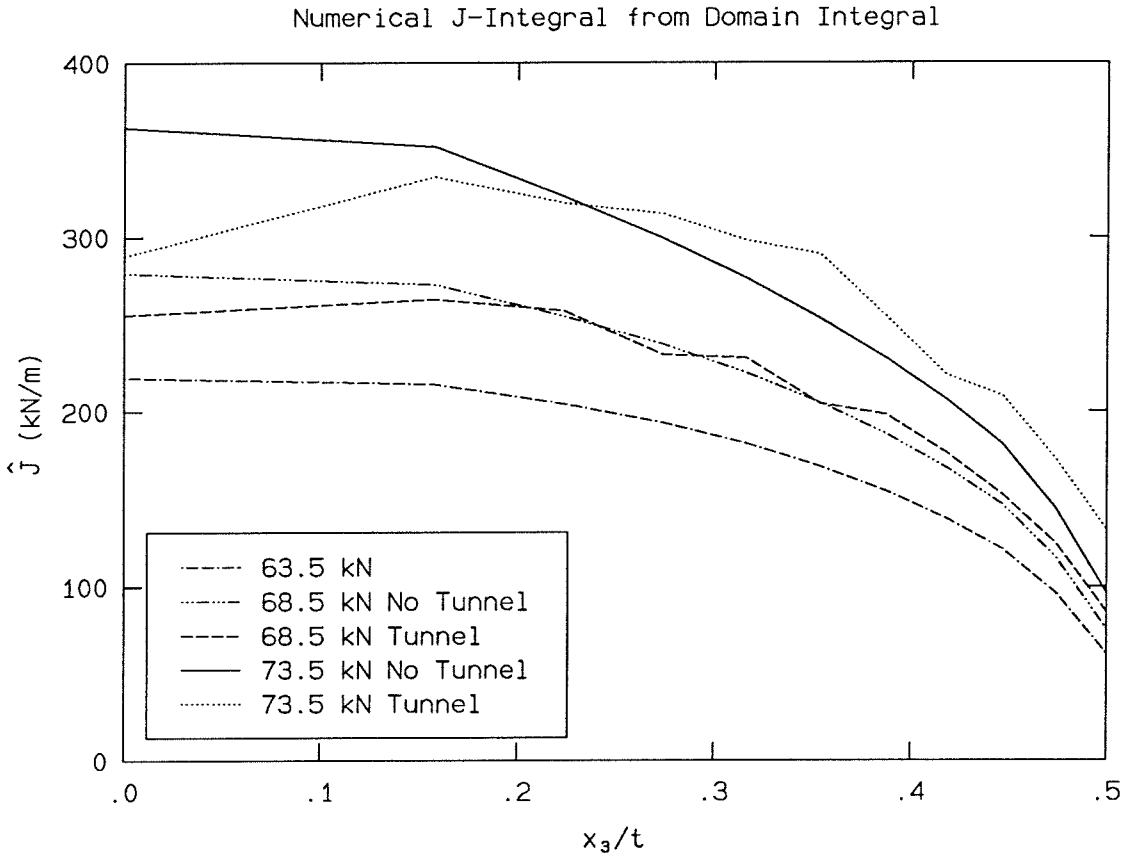


FIGURE 95. The variation of $\hat{J}$ through the thickness. 63.5 kN is the highest load for which no tunneling was observed in the experiment. 73.5 kN is the highest load reached in the interferometric experiment.

the free surface. In the case where tunneling is included, at 73500 N, the values of the five different contours vary by about 24% at the midplane (where the tunnel is deepest); the variation is about 16% halfway to the free surface, and about 28% at the free surface.

The thickness average of the two-dimensional J values are not as strongly affected by the tunneling, however. At 68500 N, the three-dimensional J value is calculated as 259 kN/m without the tunnel and 262 kN/m with the tunnel, a difference of about 1%. At 73500 N, the three-dimensional J value is calculated as 328 kN/m without the tunnel and 344 kN/m with the tunnel, a difference of about 5 %. The thickness-averaged J values are shown as a function of load in Figure

96. Also shown are the plane–stress and plane–strain relations between J and the stress–intensity factor K for elastic materials.

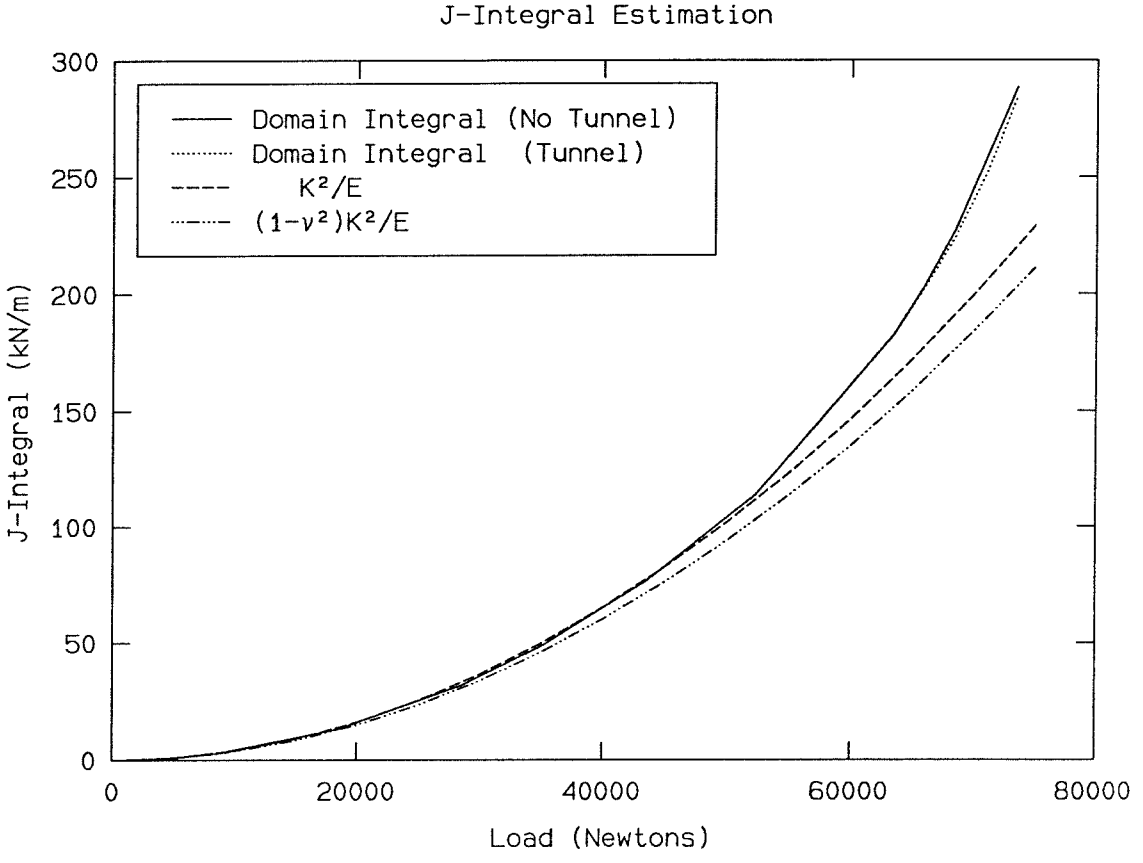


FIGURE 96. The thickness–averaged value for J as a function of load. The effect of the tunnel on the thickness average of the domain–integral calculation is not very large. The relation between J and the stress–intensity factor K for elastic materials is also shown. For elastic materials under plane–stress conditions, $J = K^2/\mathrm{E}$, under plane strain, $J = (1 - \nu^2)K^2/\mathrm{E}$.

4.3.8 Discussion

The BFGS method is an attempt to increase the speed of the convergence of the iterative solution in the regime of nonlinear deformation. In general, the most time–consuming portion of the solution is the factoring of the tangent stiffness matrix for solution of the equation $\mathbf{K}\ \Delta\mathbf{U} = \Delta\mathbf{F}$. This must be performed only once per time step using the BFGS method, with additional matrix updates effected by simple multiplications. In contrast, the full Newton method must update and solve the entire matrix equation at each iteration. The modified Newton method must also update and solve the stiffness matrix equation at each time step, but retains the same stiffness matrix throughout all iterations. The modified Newton method might seem the best choice, but the slowness of convergence in large plastic deformations offsets the gain in not updating the matrix. If the stiffness matrix were kept constant for all deformations, as in the elastic case, it would have to be factored only once, but as the deformation moved away from its original state, the convergence would be extremely slow because the search for a solution would take place along poorly chosen directions. However, if the number of iterations becomes large, the task of updating the matrix itself involves a large number of multiplications. In that case, the factors of the original stiffness matrix for that time step may be used alone, and the BFGS sequence restarted.

This finite–element model contains nonlinearity only in the constitutive relation, not in the strain–displacement relation. The small–strain assumption will become invalid in a region around the crack tip that grows with the applied load.

There are some interesting results in the paper by Narasimhan, Rosakis and Moran [52]. Comparing Figures 6 and 7 of that report, which show the contours of equivalent stress and void volume fraction for plane strain, with Figures 12 and 13 of that report, showing the contours of equivalent stress and void volume fraction for the three–dimensional model, there is a striking difference in the distribution of these quantities between the plane–strain and three–dimensional models. Also, the plot of the hydrostatic stress in Figure 11 of that report seems to match closely with the tunnel profiles measured in the current study. (Obviously, the largest

component of the hydrostatic stress is σ_{22}, which is connected to the flat tunneling within the specimen.)

The tunneling investigation used for the current study might be considered a complement to the numerical simulation of Narasimhan, Rosakis and Moran [52], in that it would provide a check on the assumptions used in predicting failure caused by void formation. The experiments indicate that initiation should occur between 63500 Newtons and 68500 Newtons. The finite–element model of [52] predicts tunneling initiation at a critical value of the local J value, $\hat{J} = 250$kN/m. From Figure 95, using the bounds on the tunneling initiation load indicated by the experiment, the value of $\hat{J}$ calculated using the finite–element model in the current study agrees fairly well with the prediction made by Narasimhan, Rosakis and Moran [52]. The maximum value of $\hat{J}$ at 63.5 kN is calculated to be 219 kN/m along the centerline of the specimen at $x_3/t = 0$. The maximum value of $\hat{J}$ at 68.5 kN (with no tunnel) is calculated to be 279 kN/m, also at $x_3/t = 0$. The average of these values at 63.5 and 68.5 kN without tunneling is 249 kN/m. Also, the value of $\hat{J}$ at 68.5 kN (including tunneling) is calculated to be 255 kN/m at $x_3/t = 0$, 264 kN/m at $x_3/t = 0.16$, and 258 kN/m at $x_3/t = 0.22$. If the effects of the local unloading are not too great, these values may be thought of as an approximation to the critical value of $\hat{J}$. In either case, the values are close to the estimate of 250 kN/m from the numerical model of Reference [52].

Since the specimen geometry actually includes notch with a finite radius rather than a sharp crack, it may not be correct to perform the J–integral calculation as described, in which the inner contour Γ is taken to the origin, which lies inside the notch. A better approach would be to use a small but finite–sized Γ outside the notch. Provision for the use of such a contour was actually implemented in the code, but this was discovered by the author only after the fact. Also, the position of the crack tip was held constant for all calculations. The tip should probably be translated forward with the tunnel, although the maximum extent of the tunnel was only about 0.8 mm, which was within all five of the contours used.

4.4 Three–Point–Bend Experiments

This section contains a brief description of the interferometric experiment itself, including details of the specimen geometry and the loading apparatus, and the field of view of the optical arrangement is shown. A calculation of the value of the J–integral is performed using the measured load and load-point displacement record, and the calculated value of J is compared with the value of J calculated from the numerical model.

4.4.1 Specimen Material and Dimensions

The three point bend experiments were performed on aircraft–quality, vacuum–degassed 4340 steel plates 1 cm thick and 30.5 cm long by 7.6 cm high with a single edge crack 3 cm in length as shown in Figure 97. The loading configuration is shown in Figure 98.

As mentioned in Section 3.3, all of the test specimens were cut from the same piece of 4340 steel, a plate four inches wide by one-half inch thick by eight feet long. The steel was heat–treated to give a ductile response, first being "normalized" at 871 C with air cooling to remove any effects of rolling or cutting, then heated to 815 C and oil–quenched to produce a Martensitic transformation and finally tempered at a temperature of 538 C for one hour and air–cooled. Two uniaxial test specimens were also cut from the same piece of steel and heat–treated identically with the three–point–bend specimens.

The crack was introduced by first making a saw cut of approximately 2 cm in length into one edge of the specimen. An electrical discharge was then used to cut an additional 1.1 cm into the specimen. The electrical discharge makes a notch of about 0.3 mm in width. The edges of each specimen were then ground flat and parallel so that a crack having a total length of 3 cm remained. The large sides of the specimen were also ground as flat and as parallel as possible resulting in a thickness of 1 cm. One side of the fracture specimen was lapped flat and polished to a mirror finish in order to measure the out-of-plane displacements with a Twyman-Green

interferometer. The other side was also lapped flat, and sanded with fine sandpaper to provide a good bond with the diffraction grating. The diffraction grating was attached with epoxy as described in detail in Section 2.4.4, in the chapter where the moire interferometer to measure the two in-plane displacement components is discussed. It is assumed that the specimen deformation is symmetric through the thickness, so that the three measured displacement components can be treated as if they had been measured on the same surface.

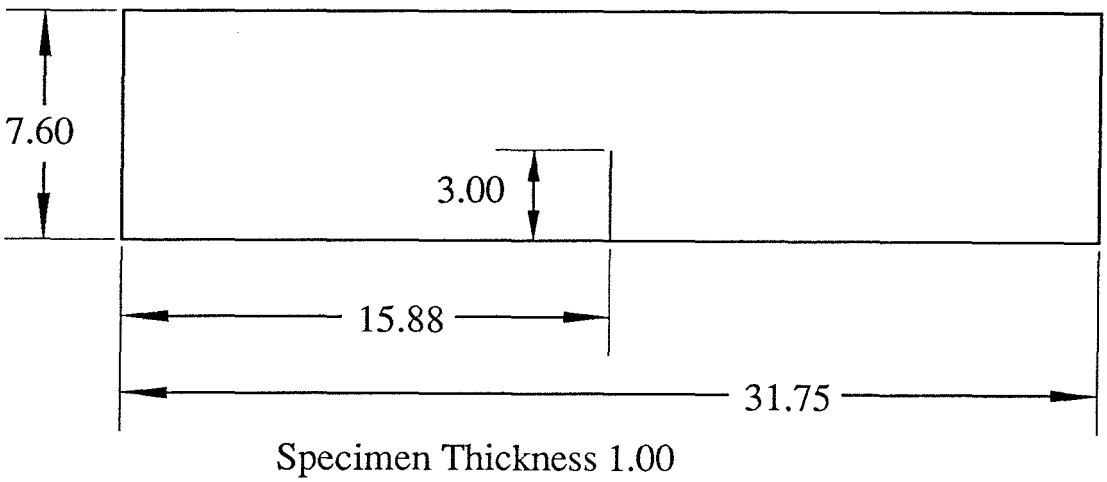


FIGURE 97. Three–point–bend specimen geometry. Material is a ductile heat treatment of 4340 steel.

Data from the numerical simulation are available at nine load steps: 100 N, 4000 N, 8500 N, 18500 N, 35000 N, 52300 N, 63500 N, 68500 N, and 73500 N. The code could be restarted from any one of these steps to produce additional output, if necessary.

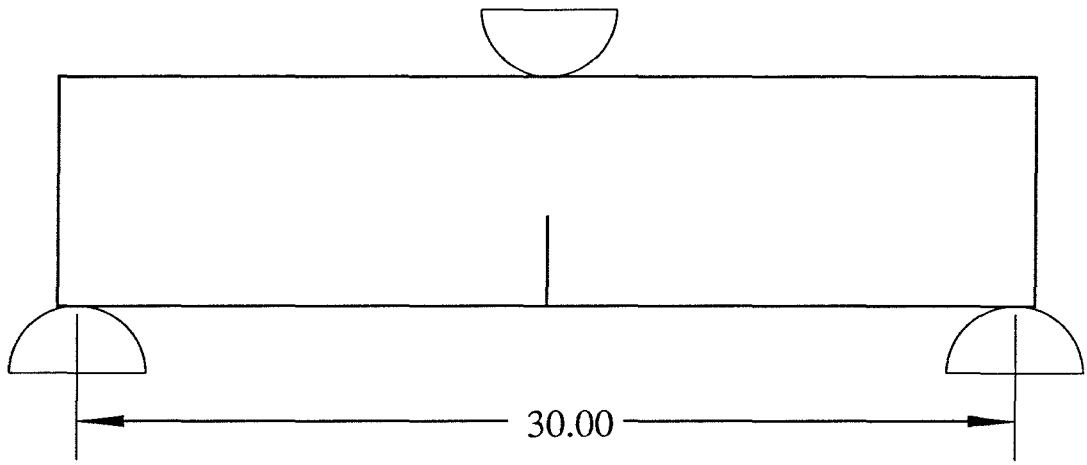


FIGURE 98. Loading configuration for the three–point–bend specimen, showing the cylindrical supports and cylindrical indenter opposite the crack. (In the experiment, the specimen was inverted, with the crack in the top edge.)

4.4.2 Experiment

The experiments were carried out on a vibration–isolated optical table in order to reduce the effects of vibrations from the surroundings. The optical table is a Newport Corporation product, measuring ten feet by four feet by eighteen inches thick. The loading frame was designed and constructed at Caltech and has a test section length that is adjustable by means of threaded rods. The load itself is introduced through a hydraulic ram, which is supplied from a Miller 559-Hydraulic Power Unit 66 that is rated for 11.6 gallons per minute at a working pressure of 2670 psi. The pressure is supplied through a Moog Series 62 Flow Control Servovalve, which is controlled by a Shore Western Mfg., Inc. SC1100-1 DC Servo Controller.

The servo controller balances the load feedback against an input command voltage to provide control of the desired applied load. The load was measured and fed back to the servo controller with the same Sensotec 400,000 Newton capacity load cell used in the uniaxial tests. As seen from the figure, the central point of the three–point–bend arrangement is held fixed, transmitting the load to the load

cell, while the two outer loading points are attached to an I beam that is moved by the hydraulic ram. Although this setup makes it slightly more difficult to align the apparatus without any load, it does have the advantage that the crack tip remains nearly stationary under loading. This is important because it means that the optics do not need gross adjustments during a test.

During a test, the servo controller is supplied with a command voltage corresponding to a load where information is available from one of the finite–element simulations. That load is held fixed while the optics are adjusted to try to remove any effects of rigid rotations that are most easily seen as loss of symmetry in the fringe patterns. For small deviations, this adjustment is orthogonal to the adjustment for the spacing of the virtual reference grating formed by the crossed, incoming laser beams, so the virtual reference–grating spacing and the displacement increment between contours remain constant. The two in-plane displacement fringe patterns are then photographed, as is the out-of-plane displacement pattern on the opposite side of the specimen.

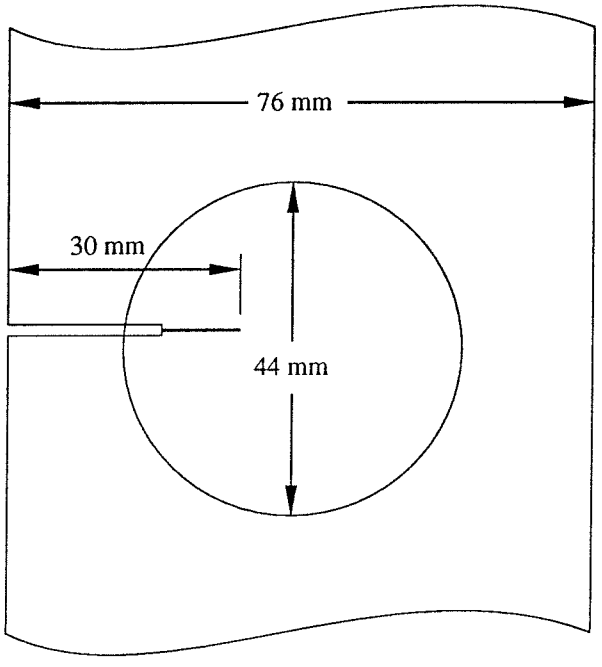


FIGURE 99. A section of the specimen showing the location of the applied diffraction grating with respect to the crack tip. The center of the circle is 7 mm ahead of the crack tip and 2.5 mm below the crack tip.

The location of the applied diffraction grating with respect to the crack tip is shown in Figure 99. The 44 mm circle represents the approximate field that is unaffected by diffraction from the edges of the lens holder. The center of the diffraction grating is slightly offset from the crack tip. The offset ahead of the crack is intentional, so that the saddle points in the displacement field fringes may be seen. These are predicted by the finite–element code to lie between 2 and 3 cm ahead of the crack tip. The offset in the other direction is due to sliding of the grating before the epoxy has cured. The liquid epoxy is quite a good lubricant. The edge of the holographic plate is held parallel to the edge of the specimen using a jig, but translation parallel to the edge of the specimen could not be completely eliminated.

In the experiment, the in-plane displacements were photographed using a Nikon FA camera body with a Nikon Micro-Nikkor 200 mm f/4 lens and a Nikon 3T close-up lens (focal length 665 mm) attached to the front of the 200 mm lens. This camera setup produced a magnification of about 0.75 (that is, 1 mm on the specimen corresponded to 0.75 mm on the negative). After the fringes were photographed this way, a Nikon TC-200 Teleconverter was placed between the camera and the 200 mm lens. The teleconverter doubled the magnification of the image without disturbing the focus of the lens. Six pictures of each displacement component were taken at each load step, three using the teleconverter and three without. The three pictures were each taken using a different length of exposure, in an attempt to bracket the optimum exposure and also to provide some redundancy in the event of a mishap in the development. (Since the image is made up of essentially collimated light, the exposure cannot be controlled by the aperture without losing some of the information around the edges.) The pictures taken without the teleconverter were intended to capture the fringe information away from the crack, but the magnification was not high enough to record the smallest fringe spacing which the optical system was capable of transmitting to the film. With the teleconverter in the path, the field of view is reduced, but the image is magnified enough so that the limitations that are due to the film resolution are less restrictive than the limits that are due to the finite aperture of the camera. The film used was Kodak Technical Pan set at an exposure

index of 120 and developed in Kodak HC-110 (dilution D) for 6 minutes. This film has an extremely fine grain size, and it can record fringes with wavelengths as small as 5 microns at the film plane; this size corresponds to 5 microns divided by the camera magnification for objects on the specimen itself.

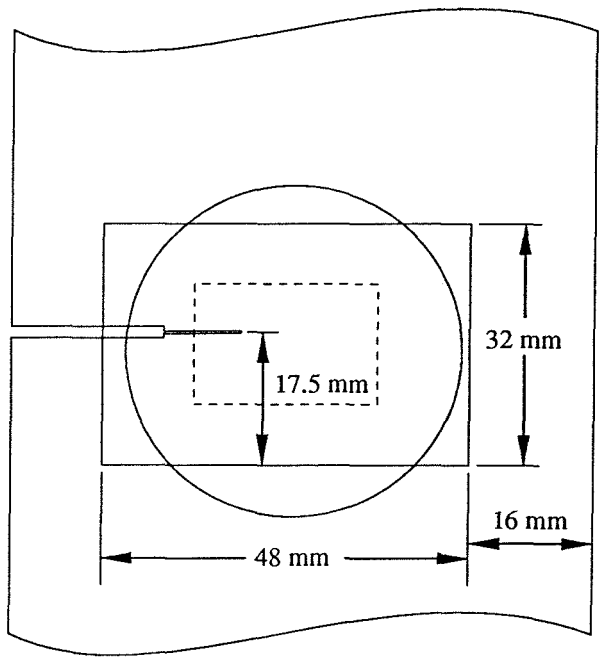


FIGURE 100. The same section of the specimen as in Figure 99 showing the location of the field of view seen through the camera. The smaller dashed–line rectangle is what is seen through the camera when the TC-200 teleconverter is used with the 200 mm f/4 lens.

The field of view seen through the camera is shown in Figure 100, superimposed on the crack and the circle representing the diffraction grating as in Figure 99. The field of view when the teleconverter is added to the 200 mm f/4 lens is shown in the smaller dashed–line rectangle. The center of the camera field is also offset with respect to the crack tip. The offset ahead of the crack is necessary in order to capture the saddle points in the displacement fringe patterns. The offset below the crack line as shown in Figure 100 is for two reasons. The first is to take advantage of the offset of the diffraction grating and the assumed symmetry of the displacement field with respect to the crack line to gather data from farther away from the crack than could be achieved if the field of view was centered on the line of the crack.

The second reason is that the diffraction grating in the top part of Figure 100 had more air bubbles than the lower part, which would make data reduction (especially automated data reduction) more difficult. It would be better if the crack tip itself could be centered in the field of view, as that would minimize the shadow spots caused by large out-of-plane displacement gradients. (See Chapter 2.) However, this arrangement was intended to yield a large magnification and still include the saddle points in the fringe patterns. In order to center the crack tip in the field of view and retain the magnification and the view of the saddle points, a larger format camera would be necessary.

4.4.3 Load–Point Deflection

The value of the J–integral (Section 4.2.1) may be determined for this specimen geometry from the load and the deflection of the central loading point. The deflection of the central loading point was determined from the deflection measured at each end of the specimen, as the central load point was held fixed while the outer ends of the specimen were moved via the I beam attached to the hydraulic ram. At one end, the displacement was measured using a dial gauge, and at the other end the displacement was measured using an LVDT. It was intended to record the output from the load cell and the LVDT throughout the experiment, but the data acquisition failed and all of the electronically recorded data was lost. The deflection measured by the dial gauge and the deflection measured by the LVDT were recorded by hand at each load step. In order to estimate the compliance of the loading frame, a large steel bar was put in place of the specimen, and the deflection of the bar was measured with the same dial gauge and LVDT arrangement up to about 30,000 N. This beam was 80 mm wide by 21 mm deep by 300 mm long, Young's modulus was assumed to be 200 GPa and Poisson's ratio was taken as 0.3. The deflection of the steel bar was calculated through beam theory, including the shear deflection since the beam was rather short, and then the deflection of the steel bar was subtracted from the overall deflection to determine the deflection of the loading frame. The total deflection of the beam under load was least–squares fit to a linear function of load, and the beam theory estimate is also linear with load, so that the correc-

tion for the machine compliance is linear with the load as well. The compliance of the loading frame was estimated to be 5.6×10^{-6} mm/N, and assumed to be constant over the loading range. The corrected measured displacements and the displacements as calculated from the finite–element code are shown in Figure 101.

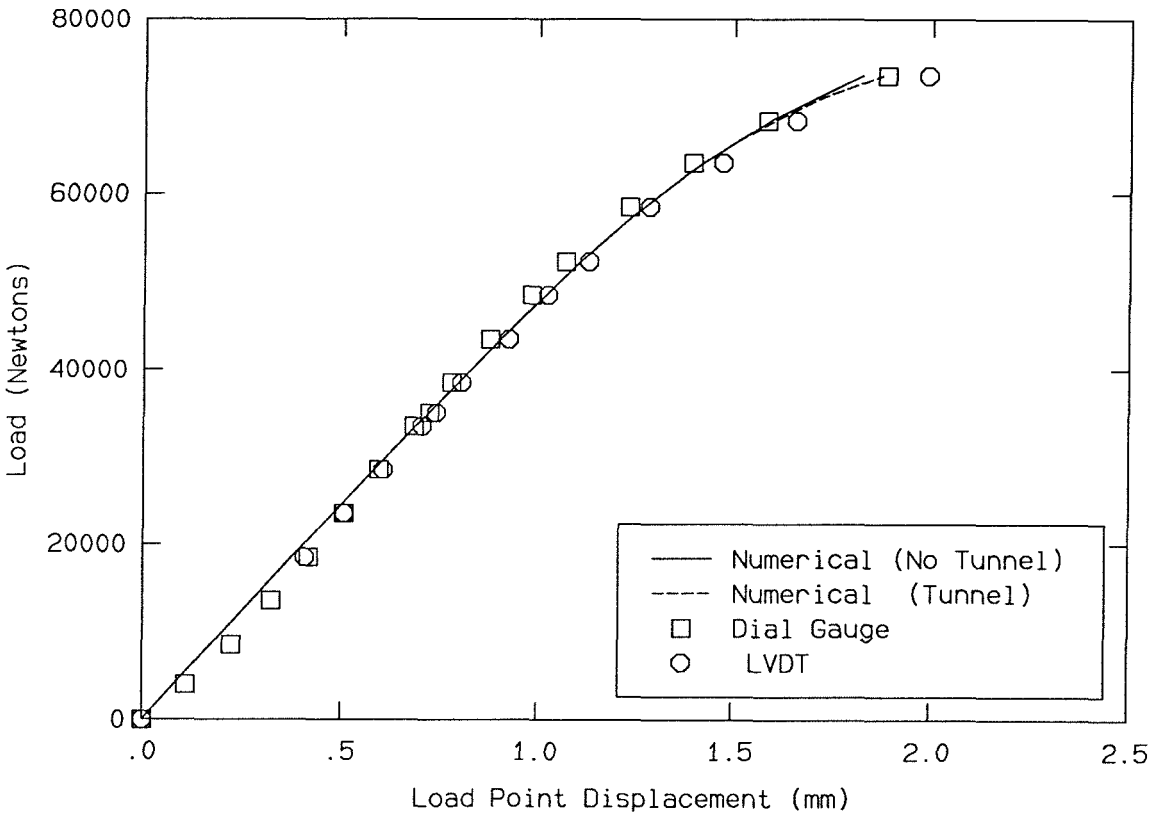


FIGURE 101. The load–point deflection as measured in the experiment shown with the deflection calculated from the finite–element code. The data from the numerical model are shown for both the case with no tunnel and the case where the tunnel has been included by releasing nodes along the crack plane. The measured displacements are corrected by an estimated machine compliance.

The difference between the LVDT and the dial gauge is approximately linear with load, and is most likely an indication of a slight rigid rotation of the specimen.

This comparison between the experiment and the finite–element model should

match closely, as it is a measurement taken far away from the local effects of the notch and its associated plastic zone. Any small discrepancies between the experiment and the numerical model near the notch should have died out or perhaps averaged out at such distances from the notch tip.

The smallest division on the dial gauge used was one ten–thousandth of an inch, or 0.00254 mm. The LVDT used had a linear range of about 10 mm, and a resolution of about 0.025 mm. Thus, the maximum error in the measurements themselves should be within 0.025 mm. The largest source of error in the measurement is the estimation of the compliance of the load frame. Since the compliance estimation was made with the loading machine in the same configuration as was used in the experiment, any error in the compliance estimate is due to the calculation of the deflection of the beam. If it is assumed that the beam theory calculation is within 10 % of the actual value of the beam deflection, then the compliance of the load frame is estimated to lie between 5.2×10^{-6} mm/N and 6.1×10^{-6} mm/N, which is an uncertainty of about ± 0.23 mm at 50000 Newtons and about ± 0.36 mm at 80000 Newtons.

The S shape of the experimental measurement at low loads is also apparent in the compliance measurement made with the beam. It is likely due to the tightening of the threaded rod connections used in the loading frame. This could also have been removed from the experimental data, but it was thought that simply accounting for a linear variation was a more conservative approach.

4.4.4 J–Integral Calculation

A simple method of calculating the J–integral has been proposed for this three–point–bend geometry [15,26,68]. For a three–point–bend specimen such as the one in this experiment, for which the width of the specimen is one quarter of the span between the supports, the J–integral is given by

$$J = \frac{2}{(w-a)\,t} \int_0^{\delta} P d\delta \tag{4.4.1}$$

where w is the width of the specimen, a is the crack length, t is the specimen thickness, P is the applied load and δ is the displacement of the loading point. This simple relation between the value of the J–integral and the work done by the applied loading may be compared with the J–integral value estimated from a path integral employed by the numerical code. For the specimen in the experiment, $w = 76.0$ mm, $a = 30.0$ mm and $t = 10.0$ mm.

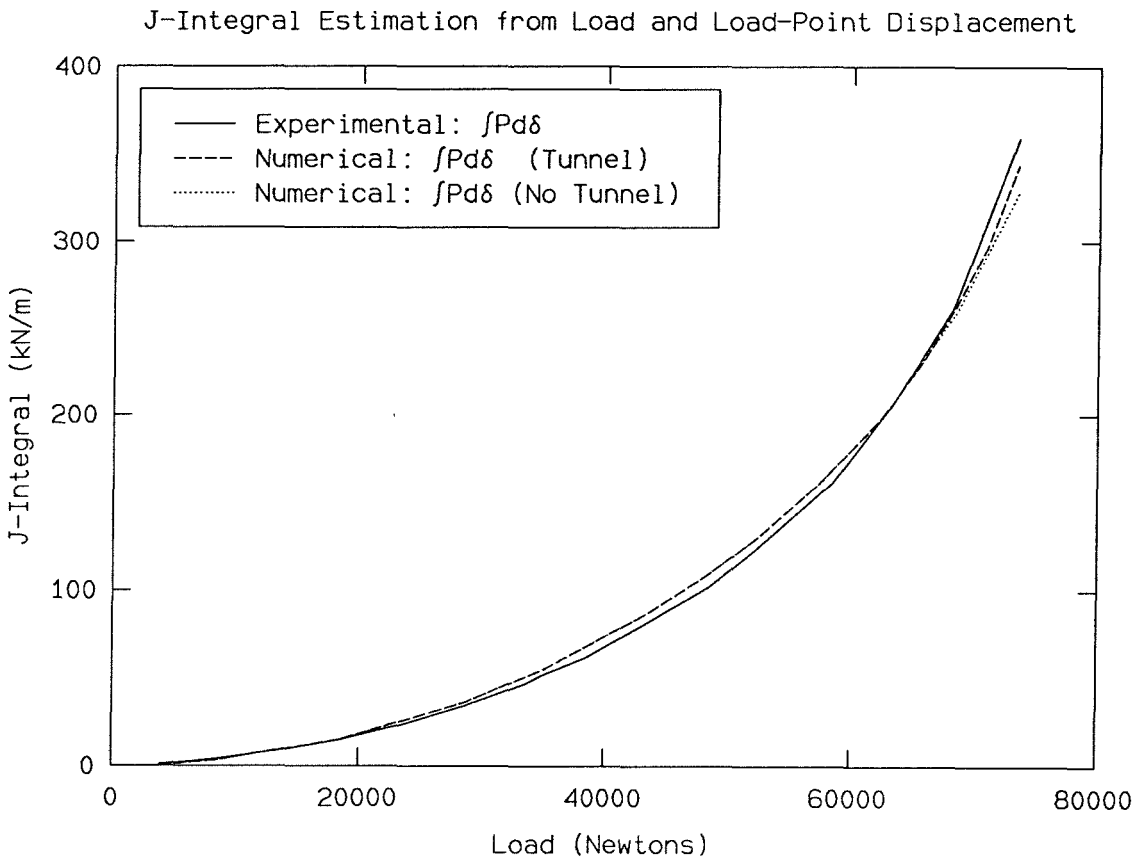


FIGURE 102. The value of the J–integral calculated from equation (4.4.1) using data from the experiment and the numerical model.

The data of Figure 101 are integrated following equation (4.4.1) using a trapezoidal rule to yield the variation of the J–integral with applied load shown in Figure 102. The experimental displacement was taken as the average of the LVDT and the dial gauge measurement to remove the rigid rotation. Also shown is the value of the J–integral calculated from the numerical model using equation (4.4.1). The good

agreement is expected from the close match of the load and load–point displacement data from the experiment and the numerical model from Figure 101.

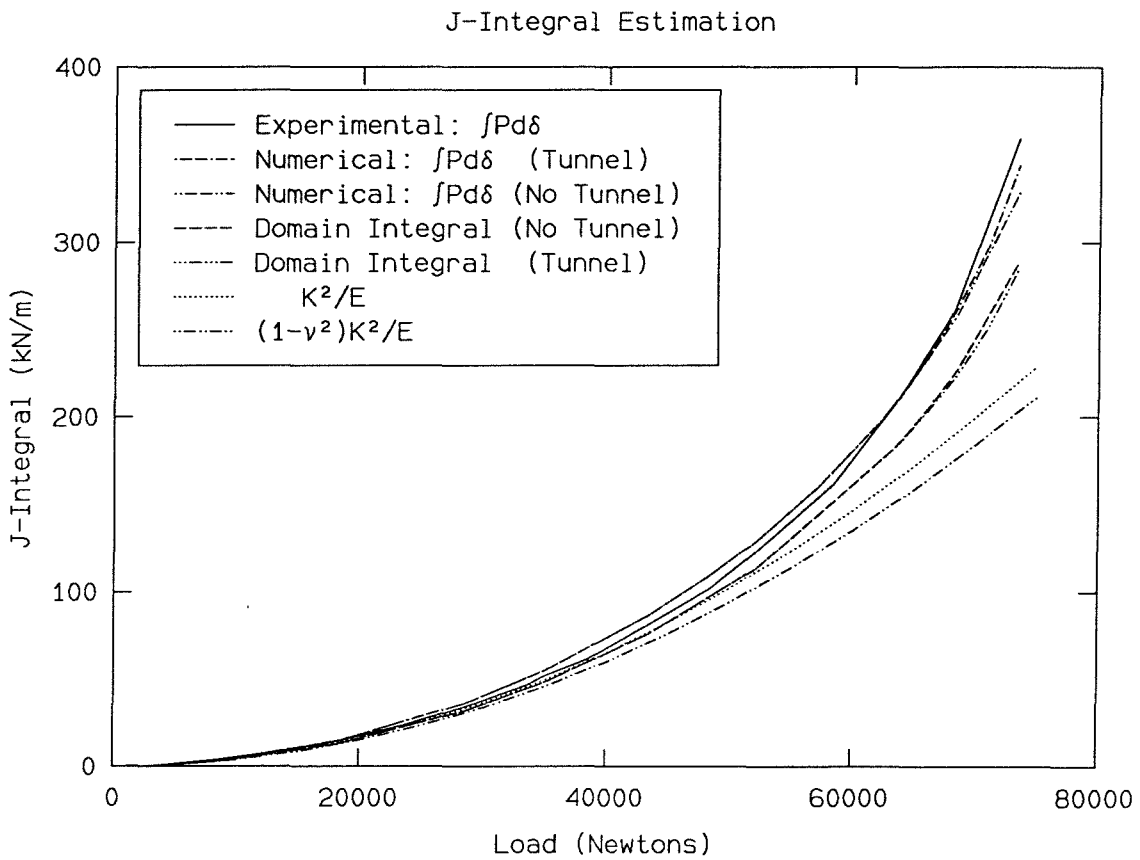


FIGURE 103. The value of the J–integral calculated from equation (4.4.1) compared with the domain integral calculation of the finite–element model and the relation between J and the elastic stress–intensity factor K. For elastic materials under plane stress, $J = K^2/\mathrm{E}$, while for plane strain, $J = (1 - \nu^2)K^2/\mathrm{E}$.

In Figure 103, the variation of the J–integral with load as calculated from equation (4.4.1) is compared with the variation of J with load from the domain–integral calculation performed in the finite–element calculation in Section 4.3, and with the variation of J for elastic materials determined from the stress–intensity factor K for this geometry from Section 4.2.5. The stress–intensity factor is found to be $K = (2795\ \mathrm{m}^{-3/2})(P)$ where P is the applied load.

4.5 Experimental Measurement of Crack Tunneling

One important possible deviation between the finite–element model and the experiment would be tunneling of the crack in the interior of the physical specimen. Such internal crack propagation may occur even when there has been no visible crack advance at the surface of the specimen, and the finite–element model used for this study did not incorporate the void growth failure model of Narasimhan, Rosakis and Moran [52], owing to the difficulty in converging to a solution after void growth has begun. Thus, the finite–element model itself had no way to simulate any tunneling of the crack in its solution. The tunneling of the crack was instead included artificially in the finite–element model by releasing nodes along the crack plane corresponding to measured crack–tunnel profiles determined from similar specimens.

In order to incorporate the crack tunneling into the finite–element code, the size and shape of the tunnel needed to be measured as a function of the applied load. Once the shape of the tunnel is known as a function of the load, the appropriate boundary conditions of the finite–element model would be released to approximate the new shape of the crack in the experiment, and the comparison between the experiment and the model may be extended.

Additionally, one could use the measured onset of tunneling to estimate critical parameters from the finite–element model, such as a critical stress or strain at which fracture initiates. Narasimhan, Rosakis and Moran [52] correlated the incipient failure caused by void growth to a critical local value of the J–integral as a failure criterion.

In order to estimate the crack tunneling as a function of load, a multispecimen approach was employed. Six specimens were loaded in the three–point–bend arrangement used for the interferometric measurements. Each specimen was loaded to a different maximum load, and then removed from the loading frame. Two specimens were unloaded to zero load to examine the path of unloading; the remaining four specimens were unloaded slightly from the maximum load and then razor blades were inserted to wedge open the crack in an effort to ensure that the tunnel surface was open to the atmosphere. The specimens were heat–tinted by placing them in

the kiln for approximately thirty-five minutes at 300 °C, as recommended in ASTM Standard E399–81 [5]. Heat tinting marks the extent of the crack advance in the interior of the specimen by coloring the specimen surface that is open to the atmosphere when the specimen is heated. Actually, the suggested time for the heat tinting was ten minutes per inch thickness of the specimen. Perhaps because the other dimensions of the specimens were much larger than the thickness, it was found that a longer time was needed to produce a dark–blue color on the surface of the specimens.

The surface color changes through yellow to brown to purple to blue as the specimen heats up. The color change is likely due to the formation of molybdenum or chromium oxides on the surface. Molybdenum and chromium have a more negative energy of oxidation than iron, so they will oxidize preferentially. This preferential oxidation is what keeps stainless steel from rusting, although stainless steel has a much higher chromium content, which forms a thick oxide layer on the surface and keeps oxygen from penetrating to the iron beneath. It is important to observe the specimen to ensure that it is removed when it reaches an acceptable blue color, as heating for a longer period of time (in this case, one hour or longer) can cause the color to disappear. One of the specimens was heated for too long, and its surface does not show as marked a delineation between the surface available before heat tinting and the surface formed when the specimen was broken open after heat tinting.

After heat tinting, each specimen was cooled in liquid nitrogen to ensure a brittle response to dynamic loading. The specimens were then broken in the drop–weight tower, so that the size and shape of the crack propagation in the interior of the body could then be measured.

The specimen geometry for the tunneling tests is shown in Figure 104. This geometry differs slightly from the geometry of the specimen used in the interferometric study because the steel was obtained from leftover cut–off pieces from an earlier set of experiments. The height of these specimens is therefore 7.11 cm instead of 7.60 cm. However, the uncracked ligament length of 4.60 cm was kept the

same so that the stress distribution ahead of the crack would be comparable. The loading fixture was the same as was used for the interferometric test, so the applied load and bending moment would be the same for both experiments. The nominal specimen thickness was 1.00 cm to match the interferometric study, but the actual specimen thicknesses varied between 0.96 cm and 1.01 cm. It is assumed that the response to the load scales linearly with the thickness of the specimen.

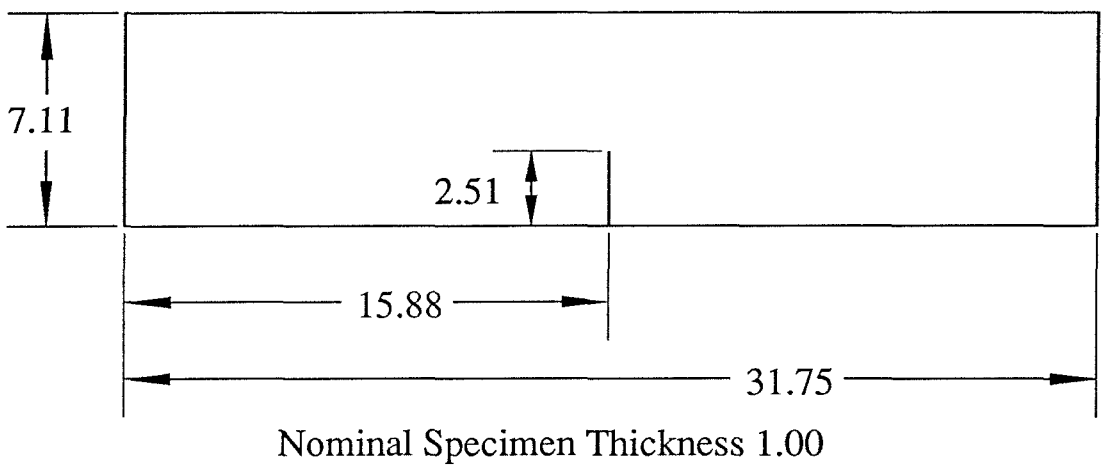


FIGURE 104. Three–point–bend specimen geometry for the tunneling tests. The material is a ductile heat treatment of 4340 steel.

4.5.1 Tunnel Profiles

It was found that no tunneling could be observed at all for the two loads at 52300 Newtons and at 63500 Newtons. What is observed at these load steps is blunting of the notch [14,26], permanent deformation at the tip of the crack corresponding to the opening and "squaring off" of the notch tip. Notch blunting can be seen in photographs taken through a microscope of Specimen 5, which was loaded to 77000 Newtons. Blunting begins at loads below 50000 Newtons, before any real crack propagation has occurred. Small amounts of tunneling were apparent

for the specimens at the load levels of 68500 and 73500 Newtons, which are the two highest load steps for which data were taken using the interferometers. The tunnel at 68500 Newtons is on the order of 0.3 mm in the center of the specimen, and at 73500 Newtons the tunnel is approximately 0.8 mm deep in the center of the specimen. The onset of tunneling at about 65000 Newtons was also inferred by Zehnder [91] from his strain gage data of a similar test.

The extent of the tunneling was measured via the microscope. The position was determined using the vernier on the microscope stage, and the tunnel depth was determined from the scale in the eyepiece of the microscope. Measurements were taken at one-half millimeter intervals. Both sides of the fractured specimen were measured, and the results averaged. The values from either side of the centerline of the specimen were then averaged to produce a symmetric profile for input to the finite–element code. These symmetric profiles are shown in Figure 105 for loads of 68500, 73500 and 77000 Newtons.

Although the interferometric experiment was carried out only to a maximum load of 73500 Newtons, the progress of the tunneling crack was investigated at loads higher than 73500 Newtons. The purpose of this investigation was to examine the relation between the branch of the crack that continues in the plane of the original crack where the tensile stresses are highest, in contrast to the cracks that propagate from the corners of the notch along the planes of maximum shear. It was found that after a small amount of tensile cracking in the original crack plane, the shear cracking and the tensile cracking would proceed simultaneously, until the two branches of the *shear cracks* from opposite sides of the specimen met and the tensile crack was eliminated. This finding ran contrary to the expectation that the tensile crack would tunnel deeply into the specimen to its final position until the shearing stress became large enough to propagate the shear cracks.

In Figure 105, only the profile of the tunneling crack is visible. What is not shown is the extent of the shear cracking at the corners of the notch. For the higher loads, the shear cracks are on the same order as the tunnel depth itself, and the eventual final failure mode of the specimen is through the shear cracking.

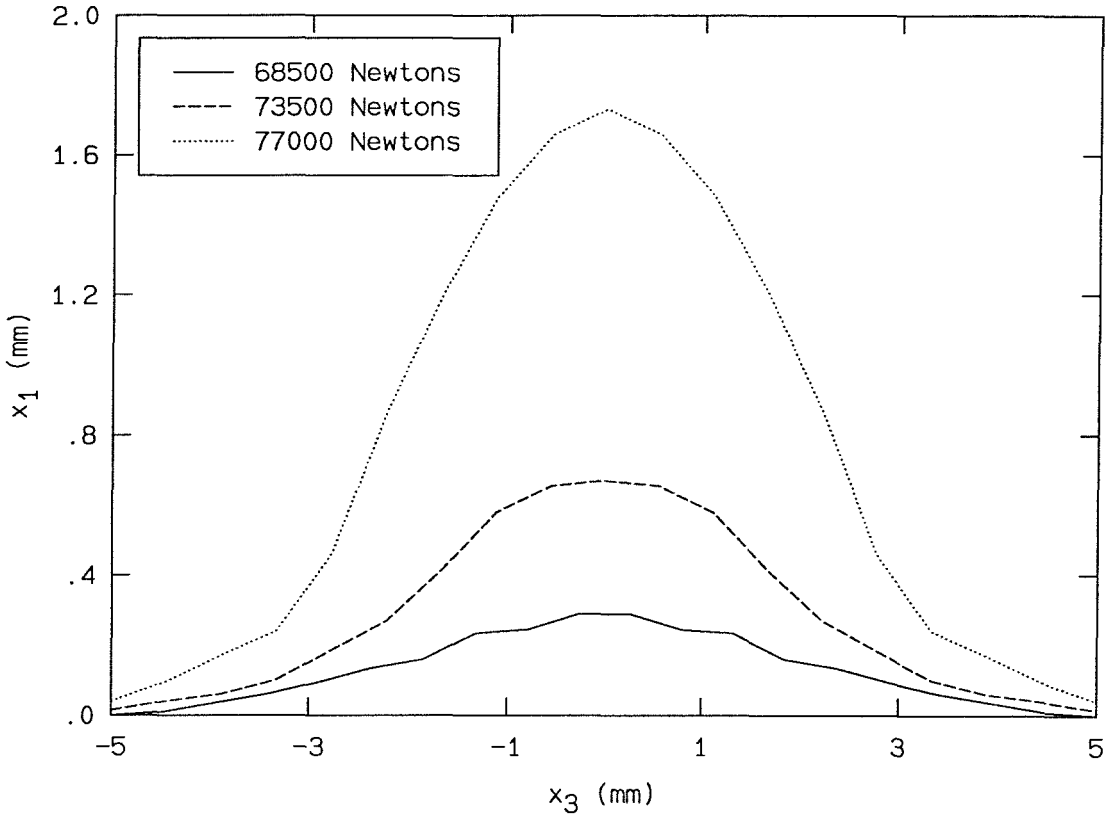


FIGURE 105. Extent of tunneling in the interior of the specimen for three different load levels.

4.5.2 Load–Point Deflection

The deflection of the central loading point as a function of load was measured for each specimen for comparison with the numerical model and the specimen used for the interferometric test. The deflection of the central loading point was determined from the deflection at each end of the specimen, as the central loading point was held fixed while the outer ends of the specimen were moved via the I beam attached to the hydraulic ram. At one end, the displacement was measured using a dial gauge, and at the other end the displacement was measured using an LVDT. The compliance of the loading frame was estimated to be 5.6×10^{-6} mm/N, and was assumed to be constant over the loading range, as in Section 4.4.3. The calculated deflection of the loading point is shown in Figure 106 for all six specimens, along

with the deflection calculated using the finite–element code. The curves plotted represent the average between the dial gauge reading and the LVDT measurement, in order to remove the effect of any rigid rotation of the specimen.

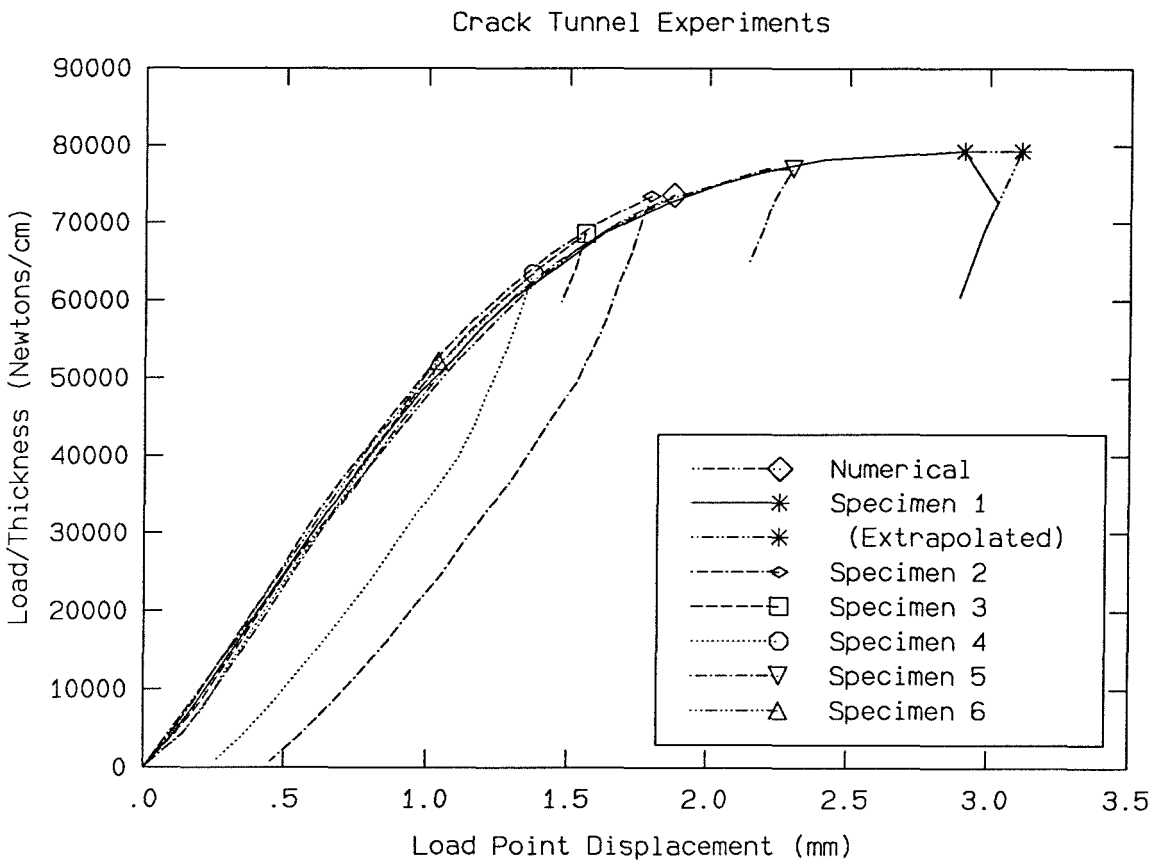


FIGURE 106. The load–point deflection versus load/thickness for each of the six tunnel specimens and the finite–element model. The symbols represent the highest load reached for each specimen. The last point on the curve for Specimen 1 is an extrapolation of the assumed linear unloading to estimate the actual maximum displacement, since that specimen was undergoing slow continuous fracture at constant load between the recorded data points. Those curves that do not unload to zero load correspond to the specimens that were unloaded slightly and the cracks then wedged open with razor blades to ensure that the tunnel surfaces were exposed to atmosphere during the heat–tinting procedure.

The agreement among the various specimens and with the numerical model

seems good. This indicates that the slightly different geometry is probably not significant, and that the choice to keep the uncracked ligament length the same as in the interferometric experiment and the numerical model was probably correct for making this comparison. Tunneling was found to initiate between 63500 Newtons and 68500 Newtons, or between the maximum load of Specimen 3 and the maximum load of Specimen 4.

The data from the numerical model shown in Figure 106 are from the simulation that included the tunnels calculated from these experiments. The tunnel was included artificially in the finite–element model by releasing nodes along the crack plane corresponding to measured lengths of crack tunnel determined from Specimens 2 (68500 N) and 3 (73500 N).

4.5.3 J–Integral Estimation

Although the specimen geometry is slightly different for the three–point–bend specimens used in these tunneling experiments from the three–point–bend geometry of the interferometric test and the numerical calculation, the same method of calculating J will be used as described in Sections 4.2.2 and 4.4.4. This method is correct for a span-to-width ratio of 4, whereas the tunneling specimens have a span-to-width ratio of 4.25. However, the good agreement between the load and load–point displacement records as compared with the numerical model indicates that this should be as good a J calculation as that performed using the interferometric specimen in Section 4.4. For a span-to-width ratio of 4, the J–integral is given by

$$J = \frac{2}{(w-a)t} \int_0^{\delta} P d\delta \tag{4.4.1}$$

where w is the width of the specimen, a is the crack length, t is the specimen thickness, P is the applied load and δ is the displacement of the loading point.

The result of applying equation (4.4.1) to the load and load–point displacement data of Figure 106 is shown in Figure 107. The data from the numerical model are also calculated using equation (4.4.1) on the load and load–point displacement data

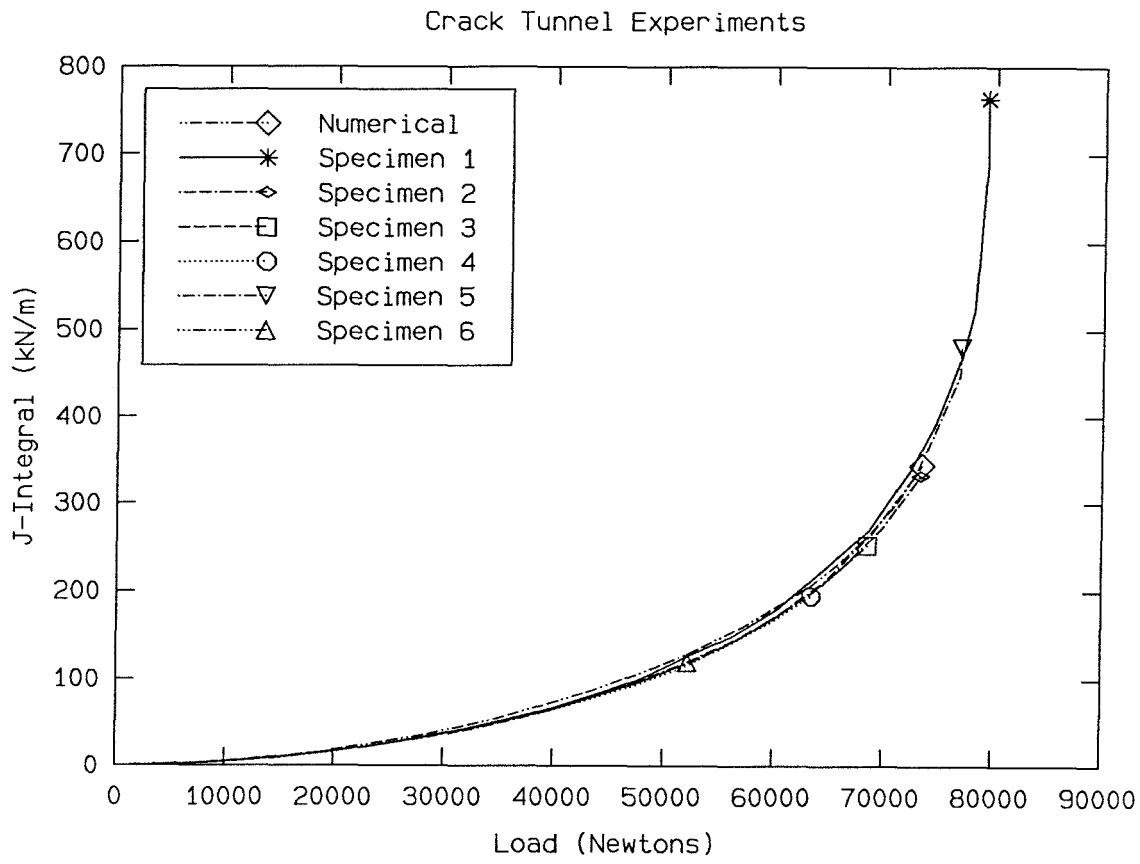


FIGURE 107. The J–integral calculated from the load and load–point displacement data from Figure 106 using equation (4.4.1).

shown in Figure 106.

Tunneling initiated at a load between 63500 Newtons and 68500 Newtons, between the maximum loads achieved using Specimens 3 and 4. The J value reached with Specimen 4 was 195 kN/m, and the J value reached with Specimen 3 was 252 kN/m, both calculated using equation (4.4.1).

Estimation of the critical value of J is ordinarily accomplished through a comparison of J and the average, measured crack advance, Δa, across the thickness of the specimen, as shown in Figure 108 [7,14,26]. The experimental data shown come from the measured crack profiles of Figure 105 for Specimens 2, 3 and 5, and the calculated J values of Figure 107 for the same three specimens. The solid

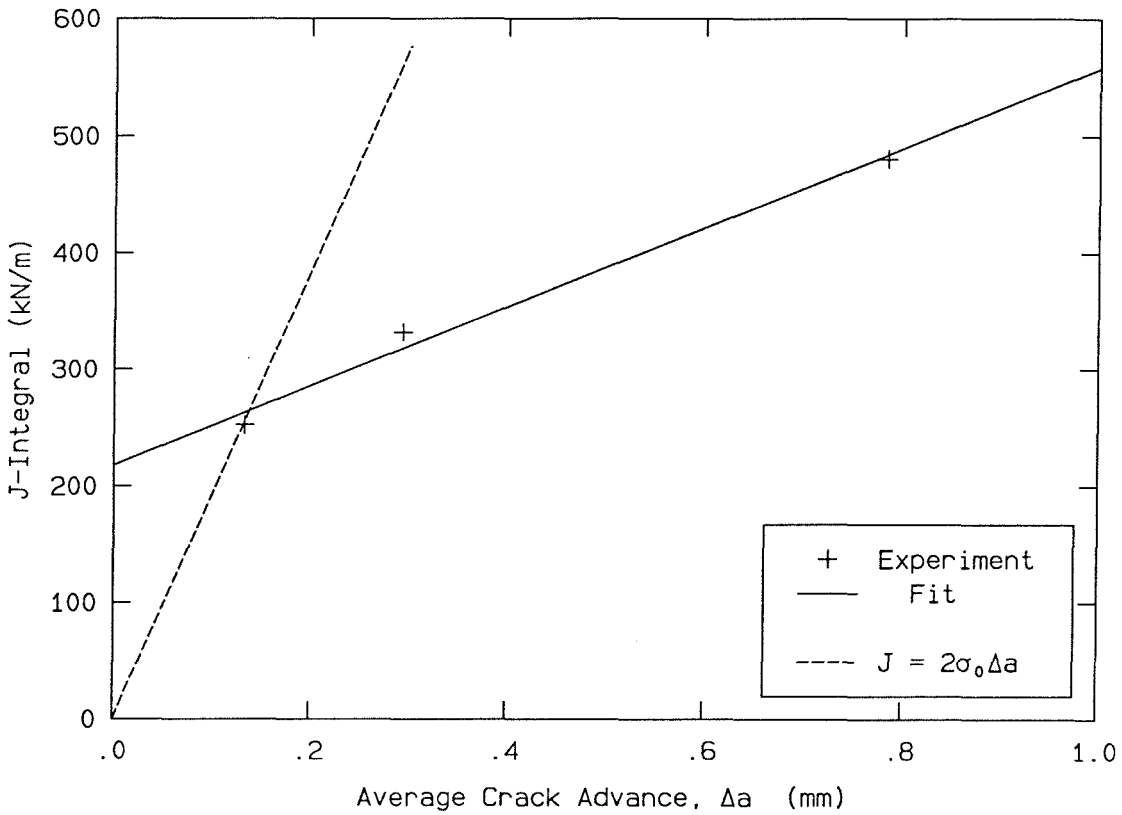


FIGURE 108. The variation of the crack advance, Δa, averaged through the thickness, of Specimens 2, 3 and 5, with the corresponding measured value of J. The dashed line, $J = 2\sigma_0 \Delta a$, is called the blunting line, it represents the estimated virtual crack advance of a sharp crack as it blunts into a semicircular shape.

line represents a least–squares fit through the three experimental data points. The dashed line is known as the crack–blunting line; it represents the estimated effective crack advance that occurs as a sharp crack blunts, but before actual tunneling begins. Use of the blunting line in calculating a critical value for J is intended for specimens that have been fatigued to produce a sharp initial crack, not the finite notch geometry of the specimens used for the test. The blunting line is calculated by assuming that the sharp crack blunts to a semicircular notch with a radius of half the crack opening displacement, estimated as J/σ_0. The crack advance that is due to blunting is thus $\Delta a = J/2\sigma_0$. The intersection of the least–squares fit of the

experimental data with the blunting line is at $J = 265\,\text{kN/m}$ and $\Delta a = 0.14\,\text{mm}$. (The value of the yield stress, σ_0, is 960 MPa, calculated from the uniaxial tests of Section 3.3.)

Since the specimens used in the test start with a rounded notch tip, it may be better to use the intersection of the line fit through the data with the vertical axis, $\Delta a = 0$. At that intersection, the value of J is found to be $J = 218\,\text{kN/m}$. This value compares well with the upper and lower bounds estimated from the J values calculated for Specimens 3 and 4 of between 195 kN/m and 252 kN/m.

Additionally, Atkins and Mai [8] suggest that for materials that work–harden appreciably, a stress value between the yield stress and the ultimate stress be substituted for the yield stress σ_0 in the equation for the blunting line. Also, they include an example of 316 stainless steel for which the blunting line is best described by $J = 4\sigma_{\text{flow}}\Delta a$, where σ_{flow} is the average of the yield stress and the ultimate tensile stress of the material. Since the steel used in these experiments does work–harden appreciably, the critical value of J is likely below the 265 kN/m calculated using the blunting line given by $J = 2\sigma_0\Delta a$.

As mentioned in Section 4.3.8, the void growth model of Narasimhan, Rosakis and Moran [52] estimated the critical value of the local J value to be 250 kN/m.

CHAPTER 5

Results

5.1 Fringe Photographs

Some sample fringe photographs are shown in the following two figures. Figure 109 shows fringes corresponding to the u_1 displacement, taken at 18.5 kN. The photograph is five times life size. Note the pattern at the edge of the field of view farthest from the notch tip, where the fringes are tending toward crossing one another. This pattern indicates a saddle point in the displacement field. Figure 110 shows the equivalent u_2 field at 18.5 kN. The photograph is five times life size. This pattern has an even clearer view of a saddle point in the displacement field. This saddle point remains nearly fixed in location with respect to the notch tip at all loads; it indicates the bending axis of the beam.

There are two interesting qualitative features that develop as the load increases; these features are illustrated in the following photographs. The first feature is the change in the shape of the u_1 fringe patterns as plastic deformation begins. When the deformation is purely elastic, the u_1 fringes at $\theta = \pm 90°$ are approximately parallel to the crack line. As plasticity occurs, the u_1 fringes curve around, in a seeming effort to touch the crack tip at $\theta = \pm 90°$. This effect is apparently due to the high shear strains emanating from the "corners" of the notch. Behind the crack, the specimen is nearly stress–free, so that the deformation is effectively a rigid rotation. This is evident from the fringe patterns, as the u_1 fringes are nearly parallel to the x_1 axis, and the u_2 fringes are nearly parallel to the x_2 axis, with the spacing of the two fringe patterns being approximately equal. Thus,

$$\frac{\partial u_1}{\partial x_1} \approx \frac{\partial u_2}{\partial x_2} \approx 0.$$

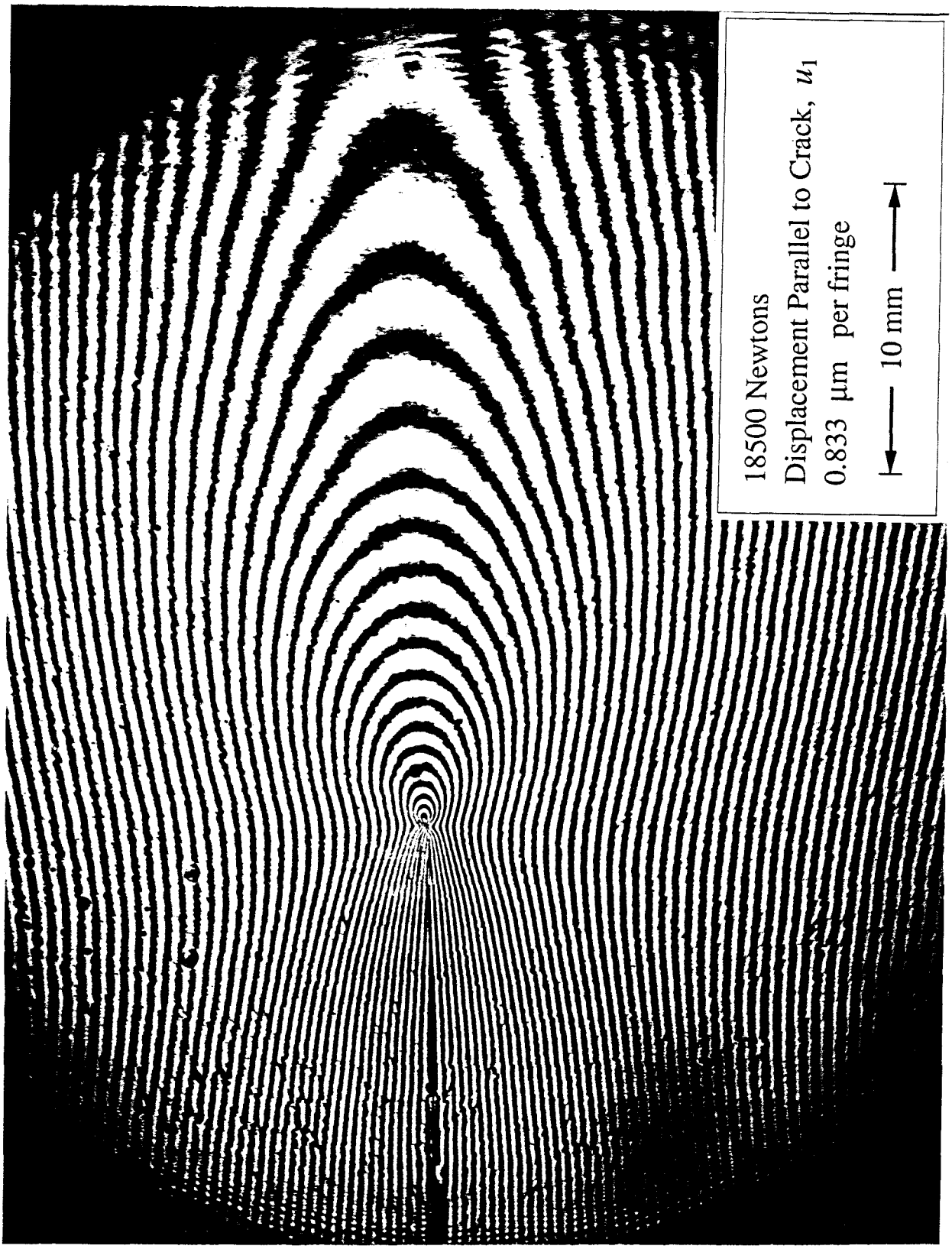


FIGURE 109. Sample u_1 fringe pattern. 18.5 kN, magnification 5 times life size.

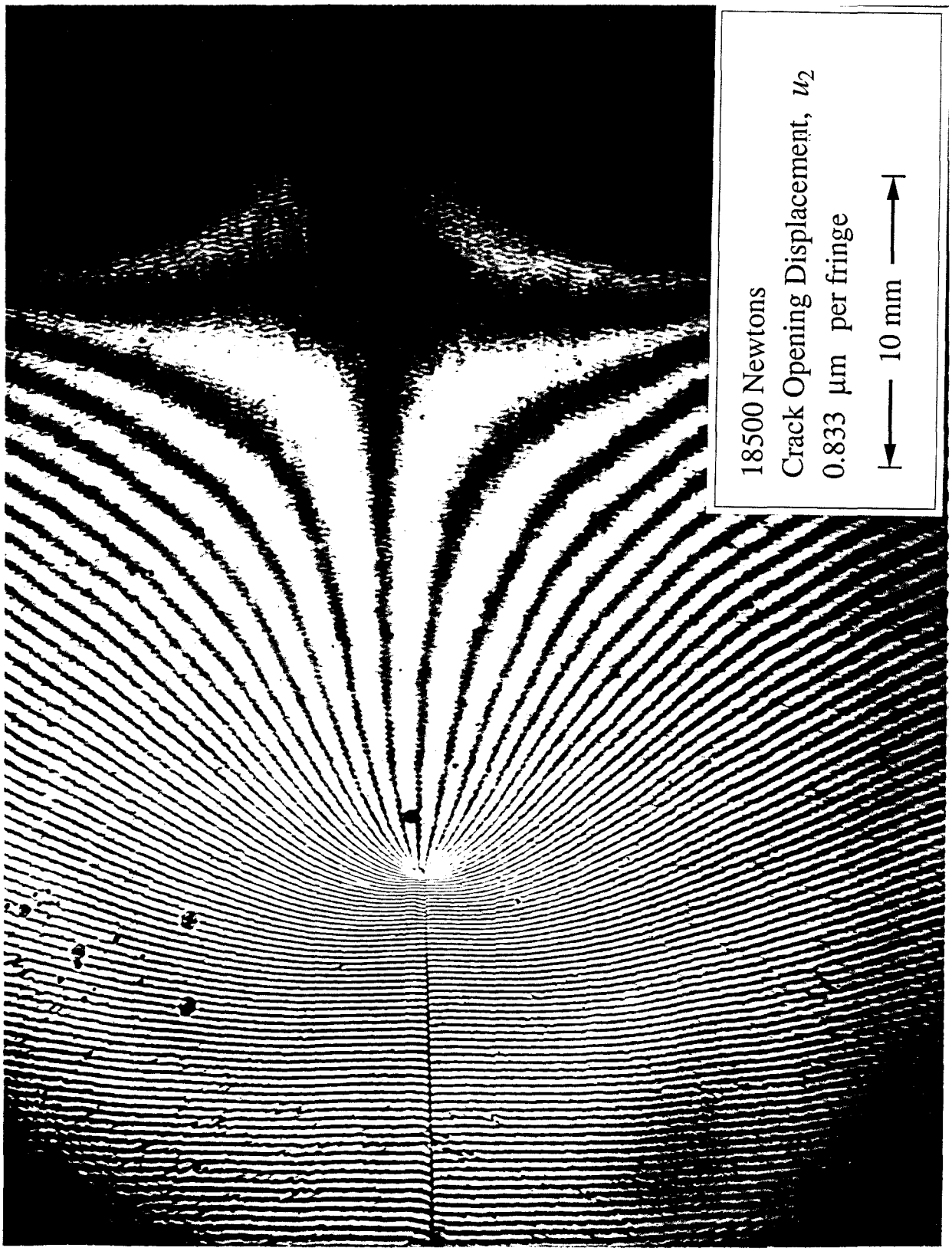


FIGURE 110. Sample u_2 fringe pattern. 18.5 kN, magnification 5 times life size.

Above the crack, u_1 is increasing with increasing x_2, and u_2 is decreasing by the same rate with increasing x_1. Thus,

$$\epsilon_{12} = \frac{1}{2}\left[\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right] \approx 0.$$

As plastic deformation begins, the u_1 fringes in the regions perpendicular to the crack line begin to loop towards the notch tip, indicating that the ϵ_{12} shear strain is no longer zero.

This feature may be illustrated by comparing Figure 111 with Figure 112. Figure 111 shows the u_1 fringe pattern at 4.0 kN; at this load the body should be completely elastic. Figure 112 shows the equivalent fringe pattern at 35.0 kN. The plastic zone in this case is on the order of 3 mm in radius. Both photos are ten times life size. Also note the beginnings of a shadow spot near the notch tip in Figure 112, a result of the aperture limitation of the camera.

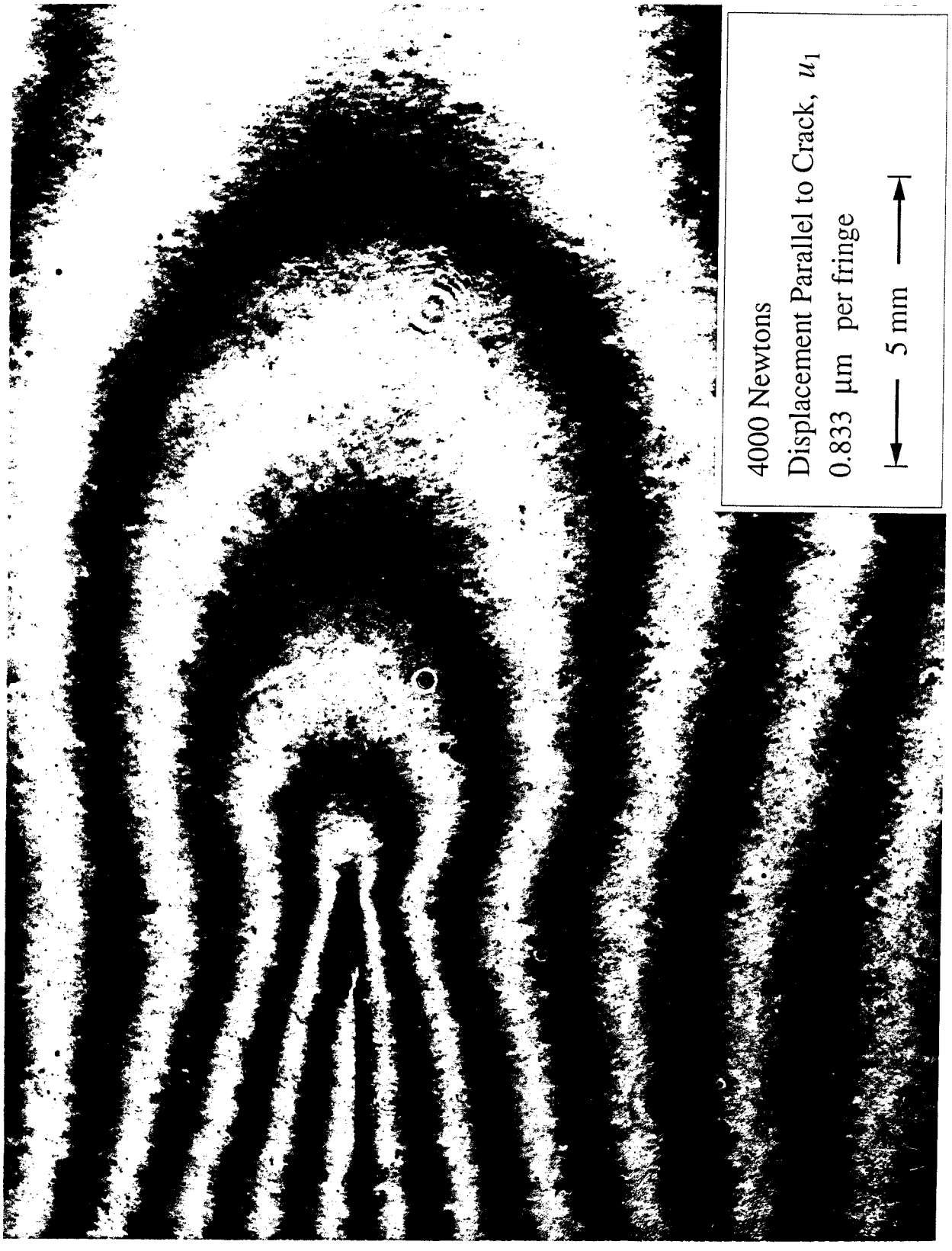


FIGURE 111. Elastic u_1 fringe pattern. 4.0 kN, magnification 10 times life size.

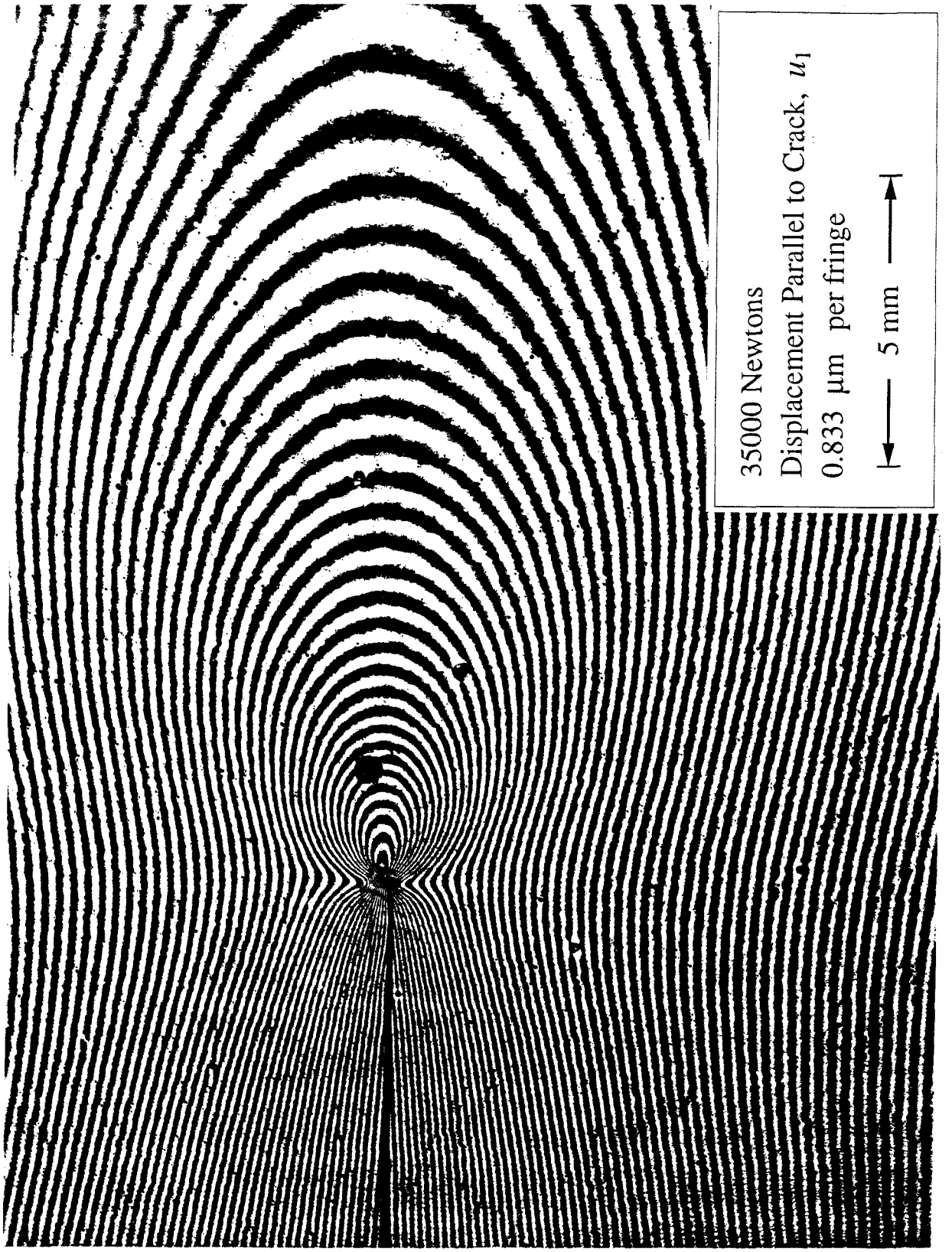


FIGURE 112. Plastic u_1 fringe pattern. 35.0 kN, magnification 10 times life size.

The second interesting qualitative feature to be seen in the photographs is the shape of the u_2 fringes ahead of the crack at very high loads. At lower loads, the fringes emanating from the notch tip follow paths along which the x_2 position is a monotonic function of the x_1 position. At the higher loads, however, in a small region ahead of the notch tip, the x_2 position of the fringes has a local maximum, so that the fringes follow paths that arc back toward the crack line before traveling away from crack line. Alternatively, this effect might be described as a region of lower fringe density that is bracketed by two regions of higher fringe density, one of which is the region of the notch tip itself. One proposed explanation for the second region of higher fringe density farther from the notch is tunneling of the crack [T.V. Hareesh, private communication]. The suggestion is that the tunneled crack tip would produce a stress concentration ahead of the original notch, causing a region of higher fringe density. This is not the case, however, as the extent of the tunnel is on the order of only one tenth of the distance to the region of higher fringe density. The tunnel may play some part in causing this fringe pattern, as illustrated in Section 5.4, where the numerically calculated plastic zones are displayed. The inclusion of the tunnel produces a region of unloading at the free surface that matches faily well with the location of the region of lower fringe density ahead of the notch tip.

The development of this barrel shape in the fringe pattern may be seen by comparing Figures 113 and 114. Figure 113 shows the u_2 fringe pattern at 52.3 kN, and Figure 114 shows the u_2 fringe pattern at 73.5 kN. Note the increase in size of the shadow spot, and the high fringe density behind the notch in Figure 114 that is due to the rotation of the crack flanks. Both photos are ten times life size.

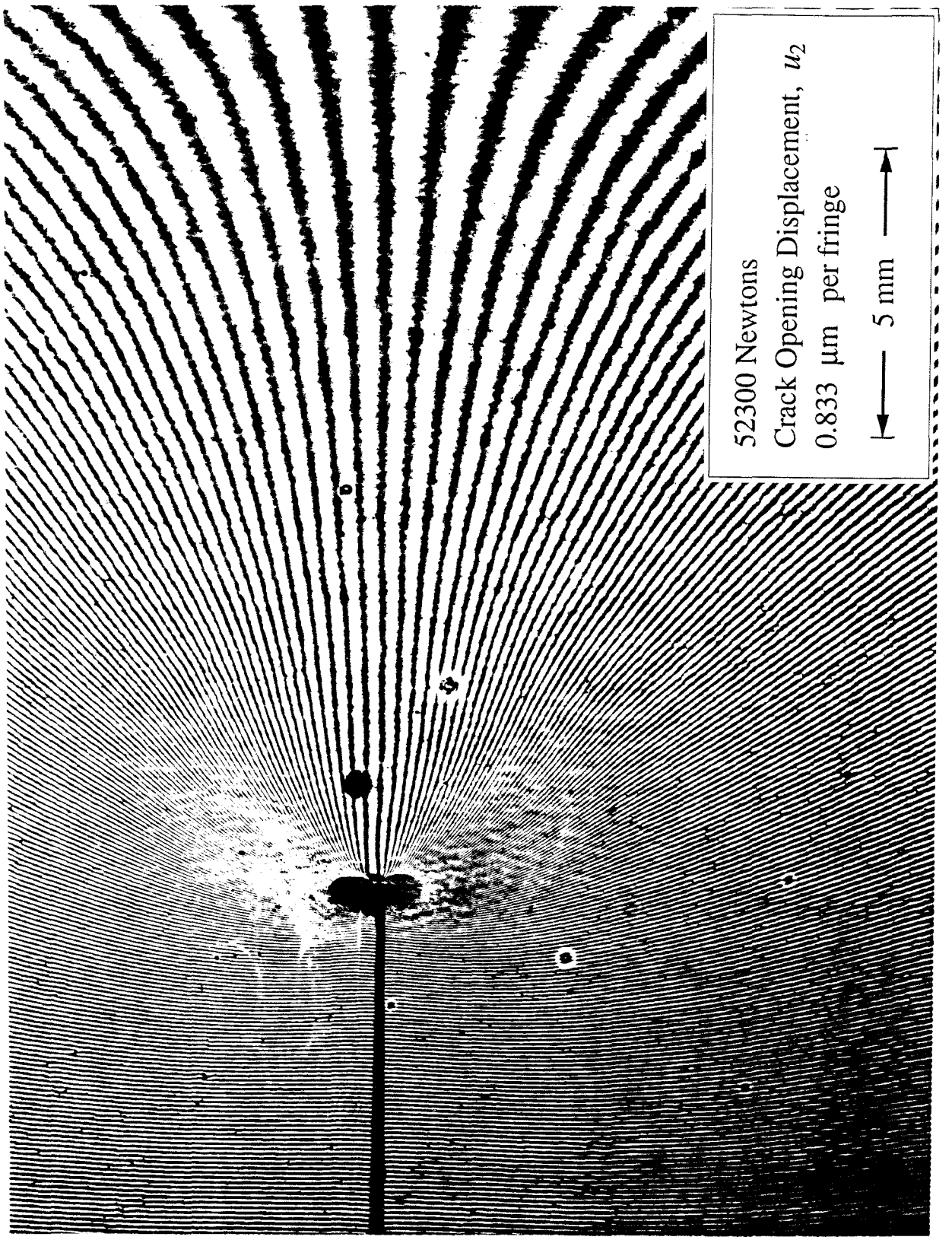


FIGURE 113. u_2 fringe pattern. 52.3 kN, magnification 10 times life size.

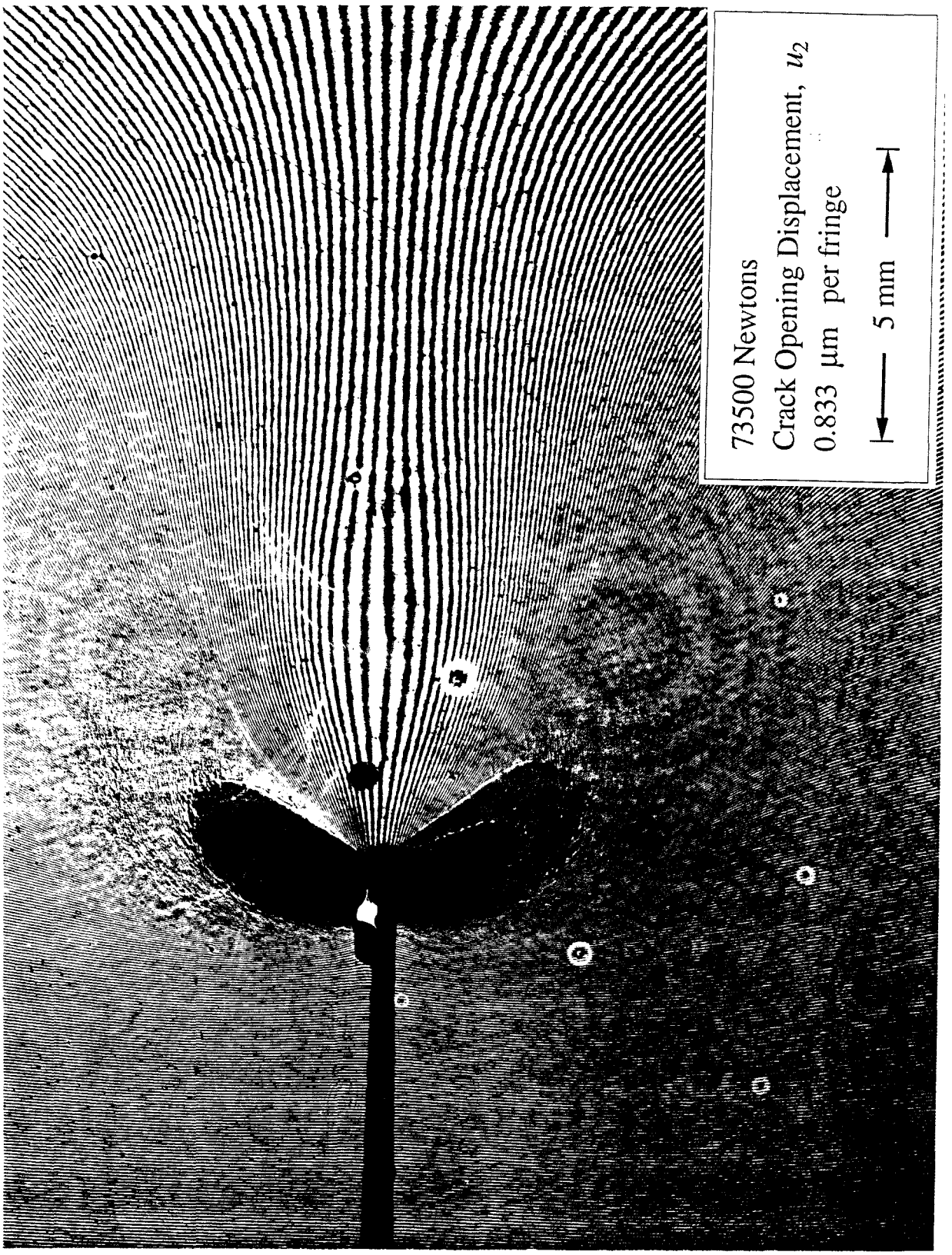


FIGURE 114. u_2 fringe pattern. 73.5 kN, magnification 10 times life size.

5.2 Comparison of Displacements

The u_1 and u_2 displacement fields measured in the experiment may be considered to be composed of three parts: the Mode I deformation that is modeled in the numerical code, a Mode II deformation brought about by any asymmetry in the loading, and a rigid rotation.

$$\begin{aligned} u_1 &= u_1^{\mathrm{I}} + u_1^{\mathrm{II}} + u_1^{\mathrm{R}} \\ u_2 &= u_2^{\mathrm{I}} + u_2^{\mathrm{II}} + u_2^{\mathrm{R}} \end{aligned} \tag{5.2.1}$$

(The superscripts I, II and R denote the Mode I contribution, the Mode II contribution, and the rigid rotation, respectively.) In Mode I, u_1^{I} is symmetric with respect to the plane of the crack (the x_1 axis or, equivalently, the line $\theta = 0$), while u_2^{I} is antisymmetric with respect to the plane of the crack.

$$\begin{bmatrix} u_1^{\mathrm{I}}(x_1,-x_2) = u_1^{\mathrm{I}}(x_1,x_2) \\ u_2^{\mathrm{I}}(x_1,-x_2) = -u_2^{\mathrm{I}}(x_1,x_2) \end{bmatrix} \quad \text{or,} \quad \begin{bmatrix} u_1^{\mathrm{I}}(r,-\theta) = u_1^{\mathrm{I}}(r,\theta) \\ u_2^{\mathrm{I}}(r,-\theta) = -u_2^{\mathrm{I}}(r,\theta) \end{bmatrix} \tag{5.2.2}$$

In Mode II, the opposite is true, so that u_1^{II} is antisymmetric with respect to the plane of the crack (the x_1 axis or, equivalently, the line $\theta = 0$), while u_2^{II} is symmetric with respect to the plane of the crack.

$$\begin{bmatrix} u_1^{\mathrm{II}}(x_1,-x_2) = -u_1^{\mathrm{II}}(x_1,x_2) \\ u_2^{\mathrm{II}}(x_1,-x_2) = u_2^{\mathrm{II}}(x_1,x_2) \end{bmatrix} \quad \text{or,} \quad \begin{bmatrix} u_1^{\mathrm{II}}(r,-\theta) = -u_1^{\mathrm{II}}(r,\theta) \\ u_2^{\mathrm{II}}(r,-\theta) = u_2^{\mathrm{II}}(r,\theta) \end{bmatrix} \tag{5.2.3}$$

The rigid rotation may be found as

$$\begin{aligned} u_1^{\mathrm{R}}(x_1,x_2) &= x_1(\cos\gamma - 1) - x_2 \sin\gamma \\ u_2^{\mathrm{R}}(x_1,x_2) &= x_1 \sin\gamma + x_2(\cos\gamma - 1) \end{aligned} \tag{5.2.4}$$

where γ is the angle of rotation. It will be assumed that γ is very small, so that the rigid rotation may be approximated by

$$\begin{aligned} u_1^{\mathrm{R}}(x_1,x_2) &= -x_2\gamma \\ u_2^{\mathrm{R}}(x_1,x_2) &= x_1\gamma \end{aligned} \tag{5.2.5}$$

With this approximation, the displacements that are due to rigid rotation have the same symmetries as the Mode II deformation. As a result, the Mode I deformation may be extracted from the measured displacements as

$$\begin{aligned} 2u_1^{\mathrm{I}}(x_1, x_2) &= u_1(x_1, x_2) + u_1(x_1, -x_2) \\ 2u_2^{\mathrm{I}}(x_1, x_2) &= u_2(x_1, x_2) - u_2(x_1, -x_2) \end{aligned} \tag{5.2.6}$$

or alternatively,

$$\begin{aligned} 2u_1^{\mathrm{I}}(r, \theta) &= u_1(r, \theta) + u_1(r, -\theta) \\ 2u_2^{\mathrm{I}}(r, \theta) &= u_2(r, \theta) - u_2(r, -\theta) \end{aligned} \tag{5.2.7}$$

A representative sample of the data from the experiment is shown in the figures that follow, compared with the displacement from the numerical model. The comparison is made at six load levels, 4.0 kN, 8.5 kN, 18.5 kN, 35.0 kN, 52.3 kN and 73.5 kN.

The radial variation of the u_1 displacement is shown on lines at $\theta = 0°$ and $\theta = 90°$, and the angular variation of the u_1 displacement is shown at $r = 0.504t$ and at $r = 1.066t$. The data along $\theta = 90°$ at the largest radius indicate the maximum u_1 displacement measured in the experiment at that load level. Away from the notch, the u_1 variation along $\theta = 90°$ is quite linear, because of the far–field bending.

The radial variation of the u_2 displacement is shown on lines at $\theta = 60°$ and $\theta = 180°$, and the angular variation of the u_2 displacement is also shown at $r = 0.504t$ and at $r = 1.066t$. The data along $\theta = 180°$ at the largest radius indicate the maximum u_2 displacement measured in the experiment at that load level. Away from the notch, the u_2 variation along $\theta = 180°$ is also quite linear; behind the notch, the deformation corresponds to a rigid rotation. Along $\theta = 60°$, the effects of the far–field bending loads can be seen as the displacement first increases and then decreases as the location of the measurement moves across the bending axis of the specimen.

The u_1 displacement data from the experiment are shown as $u_1(r, \theta)$ for $\theta > 0$ compared with $u_1(r, -\theta)$ for $\theta < 0$, while the u_2 displacement data are shown as

$u2(r,\theta)$ for $\theta > 0$ compared with $-u2(r,-\theta)$ for $\theta < 0$, in order to emphasize the Mode I deformation.

In addition, since the interferometric measurement is a relative measurement, there is a rigid translation to fix for comparison between the experiment and the finite–element model. In reducing the data from each fringe photograph, the fringe nearest the notch tip along the x_1 axis was arbitrarily chosen as zero. The rigid u_2 translation is fixed through the assumption of Mode I symmetry about $\theta = 0$, so that $u_2^{\mathrm{I}}(r,0) = 0$. The numerical u_1 displacement has been arbitrarily translated to match the experimentally measured u_1 displacement along $\theta = 0$ at a radius of one half of the plate thickness. This radius was chosen because it is clearly visible in all of the photographs.

Also note that since each interferometer is adjusted independently at each load to minimize the fringes that are due to rigid rotation, the size of the rotation in each photograph at each load is a random quantity.

The displacement data from the numerical model are shown for both the free surface, $x_3/t = 0.5$, which is where the experimental measurement is made, and for the center of the specimen, $x_3/t = 0$, to illustrate the three–dimensionality of the deformation.

The most significant consistent discrepancy occurs in the u_1 displacement at large r, with the numerical code consistently underpredicting the magnitude of the displacement. This may be because the mesh at this location is too coarse (and therefore too stiff) to characterize the deformation accurately.

Another significant feature of the *numerical* displacement variations is the sharp upturn in both the u_1 and u_2 displacement just ahead of the notch tip. The change in the u_1 displacement was expected, because the notch surface is traction free, so that there is a local maximum in the σ_{11} stress ahead of the notch that leads to a local minimum in the u_1 displacement. This upturn is difficult to detect in the fringe photographs: At lower loads the upturn is less than the displacement increment between fringes, while at higher loads the near tip region is obscured

because of the aperture limitation of the camera.

The agreement between the measured displacements and the numerically calculated displacements is quite good for loads up to 52.3 kN. The discrepancy between the numerical model and the experimental measurement is largest at the highest load, 73.5 kN. There are several possible sources for this discrepancy. The first possible source is tunneling of the crack, and an attempt was made to remove this problem through a nodal–release procedure. A second possible source of dicrepancy is that the strains are so large at the higher loads that the assumption of linearized strains employed in the numerical model is no longer valid. A third possible source is continuous slow crack growth within the specimen. This effect is clearly visible in Specimen 1, in one of the tests to determine the tunneling behavior of the material; at a fixed ultimate load, the specimen deformed continuously. This slow crack growth may also have occurred in the interferometric investigation, although on a much smaller scale.

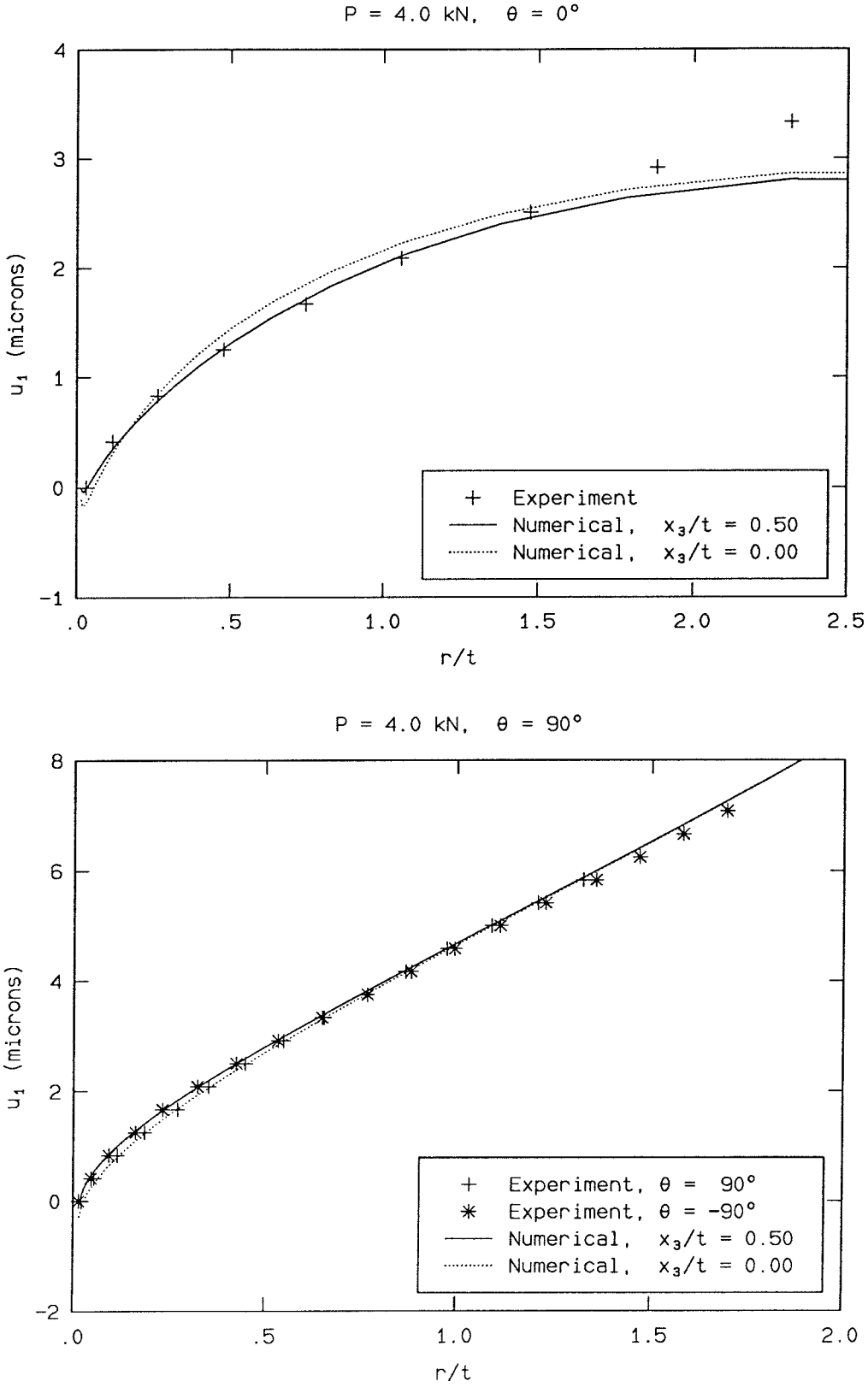


FIGURE 115. Radial variations of u_1, 4.0 kN.

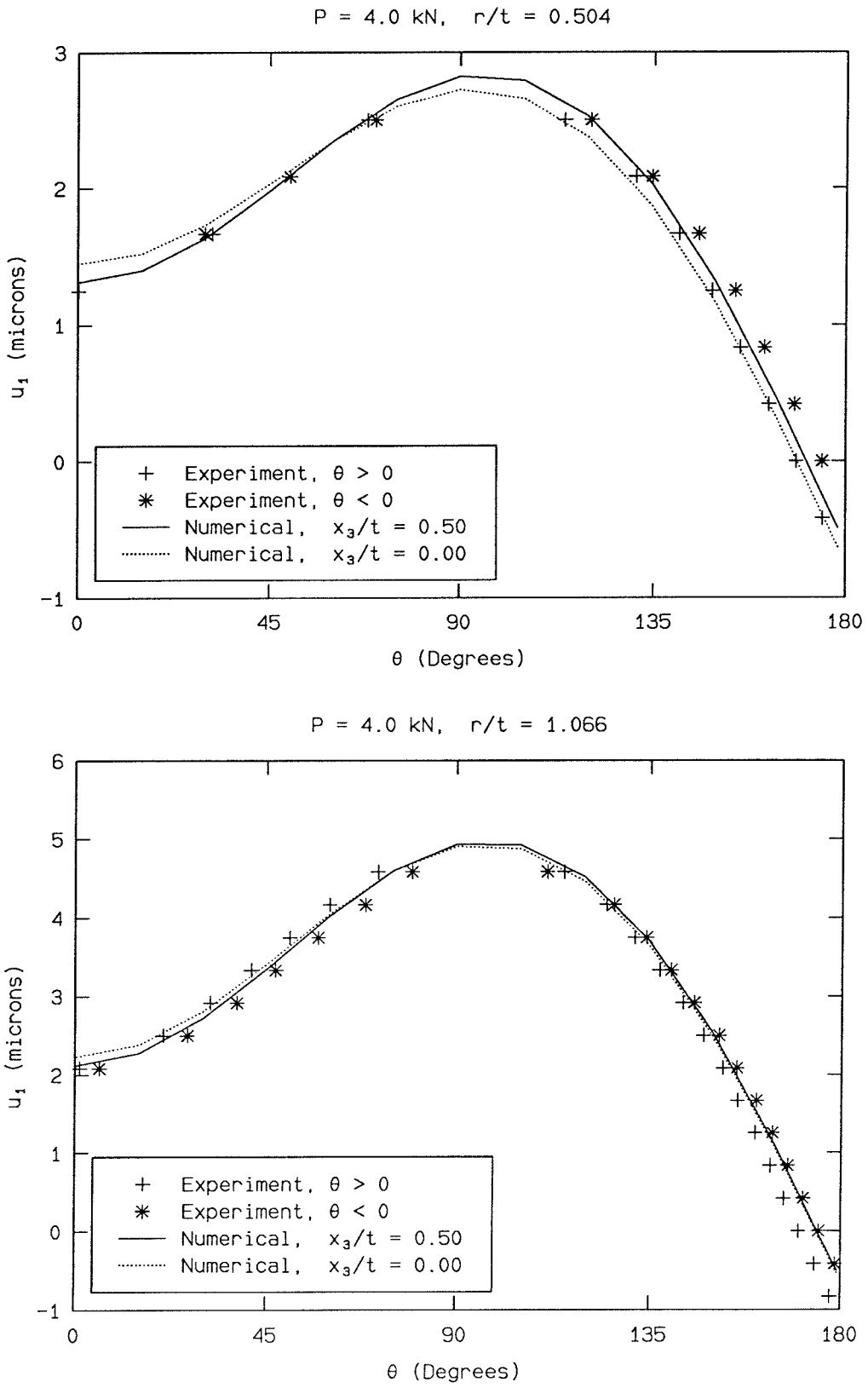


FIGURE 116. Angular variations of u_1, 4.0 kN.

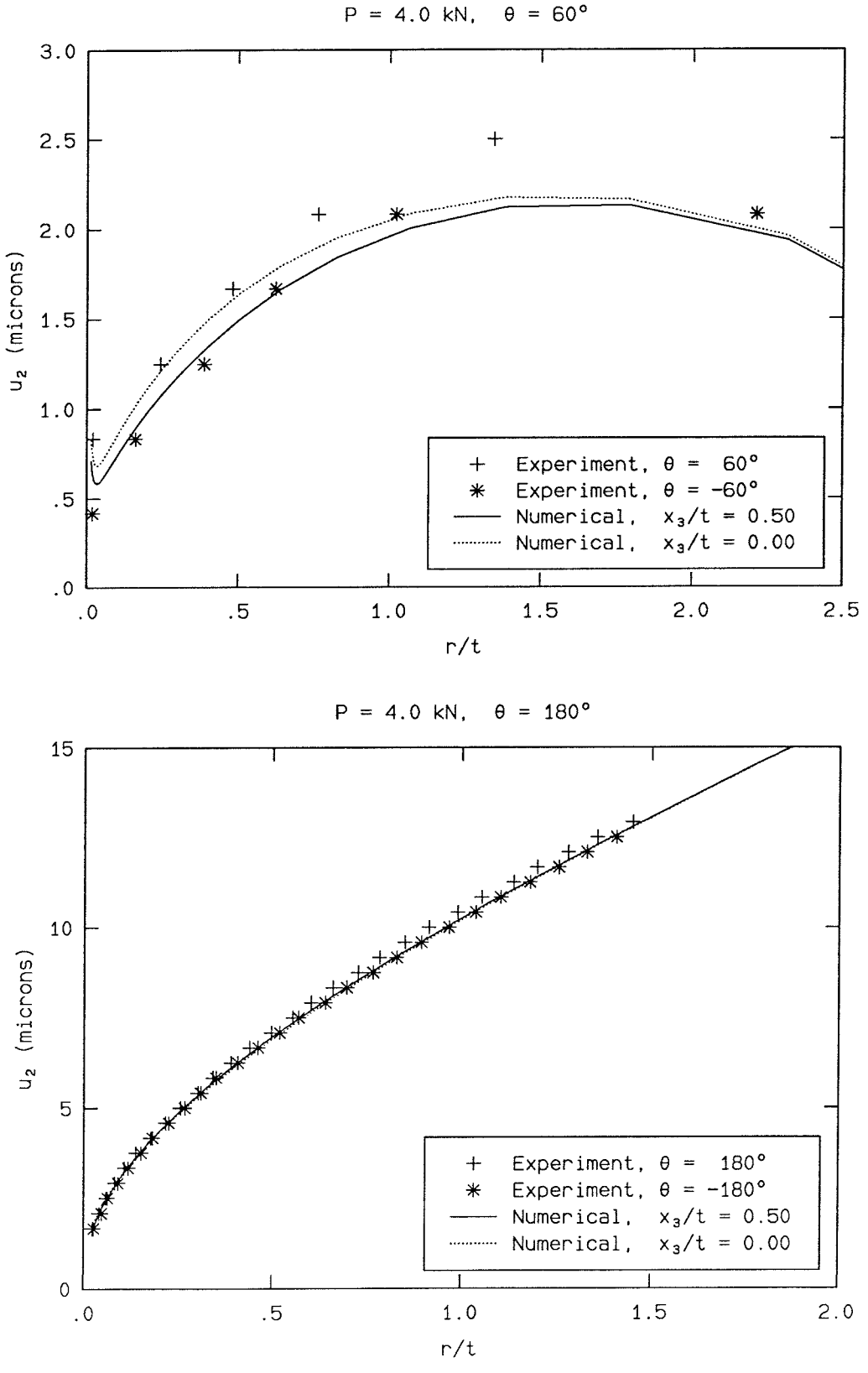


FIGURE 117. Radial variations of u_2, 4.0 kN.

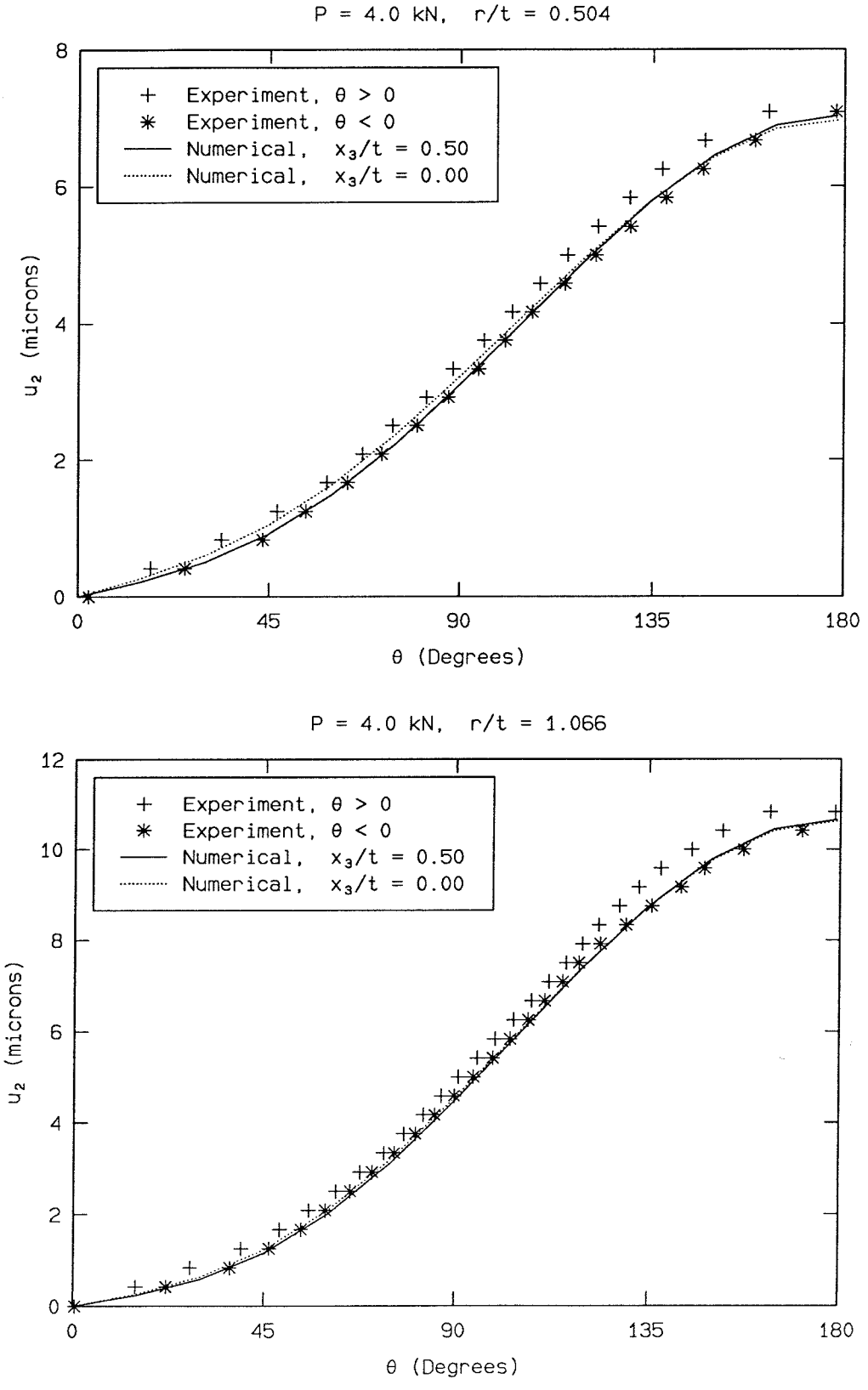


FIGURE 118. Angular variations of u_2, 4.0 kN.

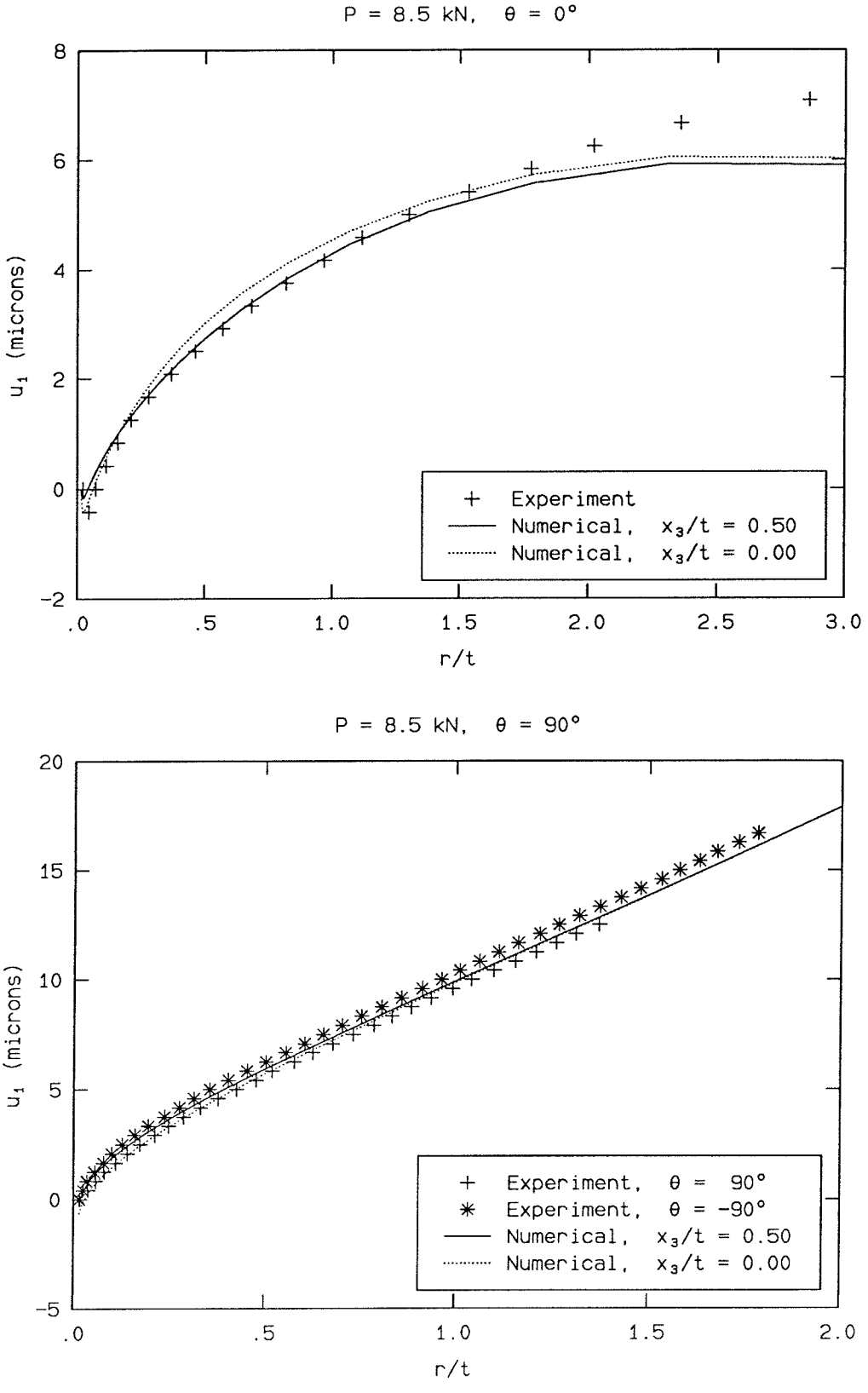


FIGURE 119. Radial variations of u_1, 8.5 kN.

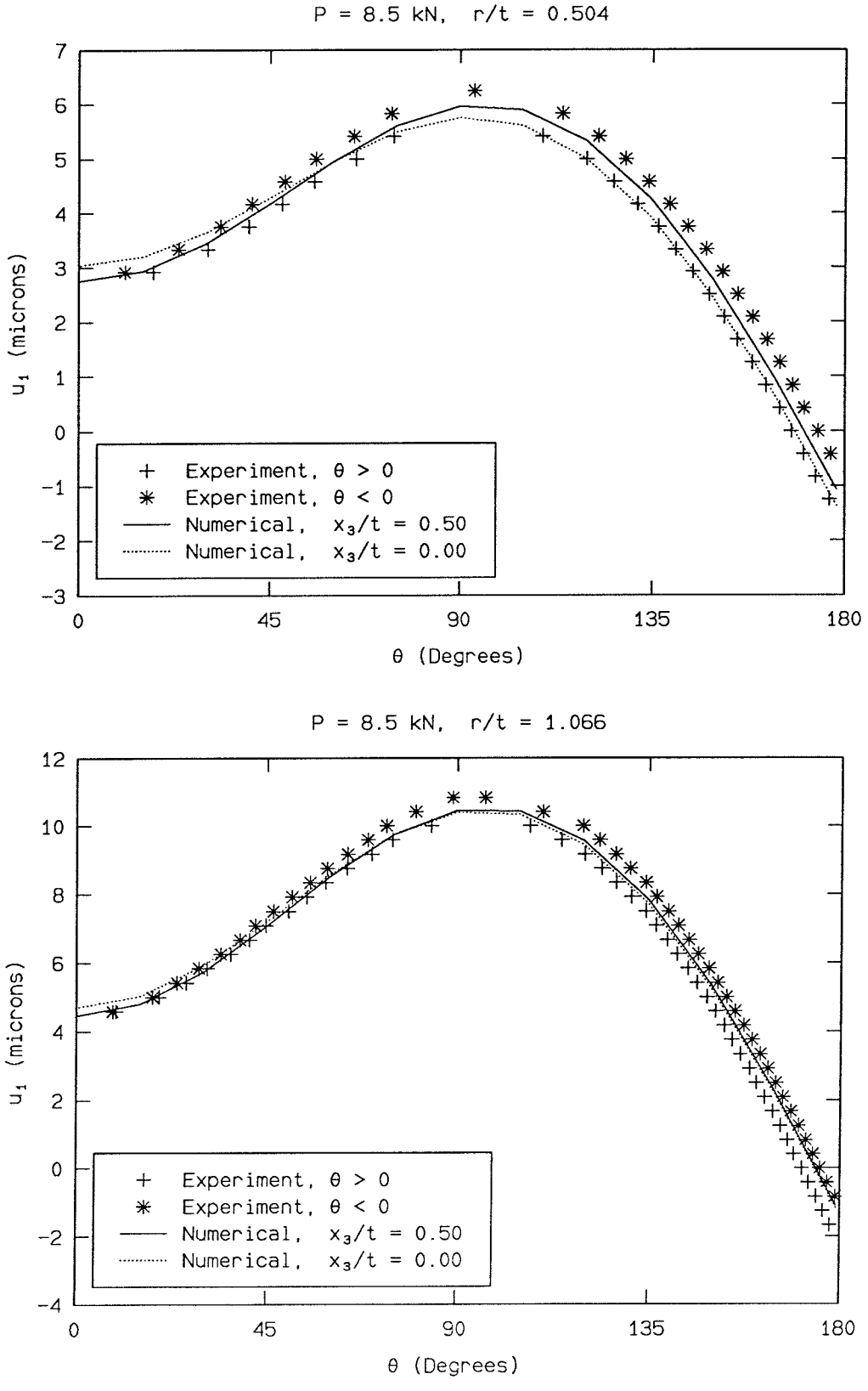


FIGURE 120. Angular variations of u_1, 8.5 kN.

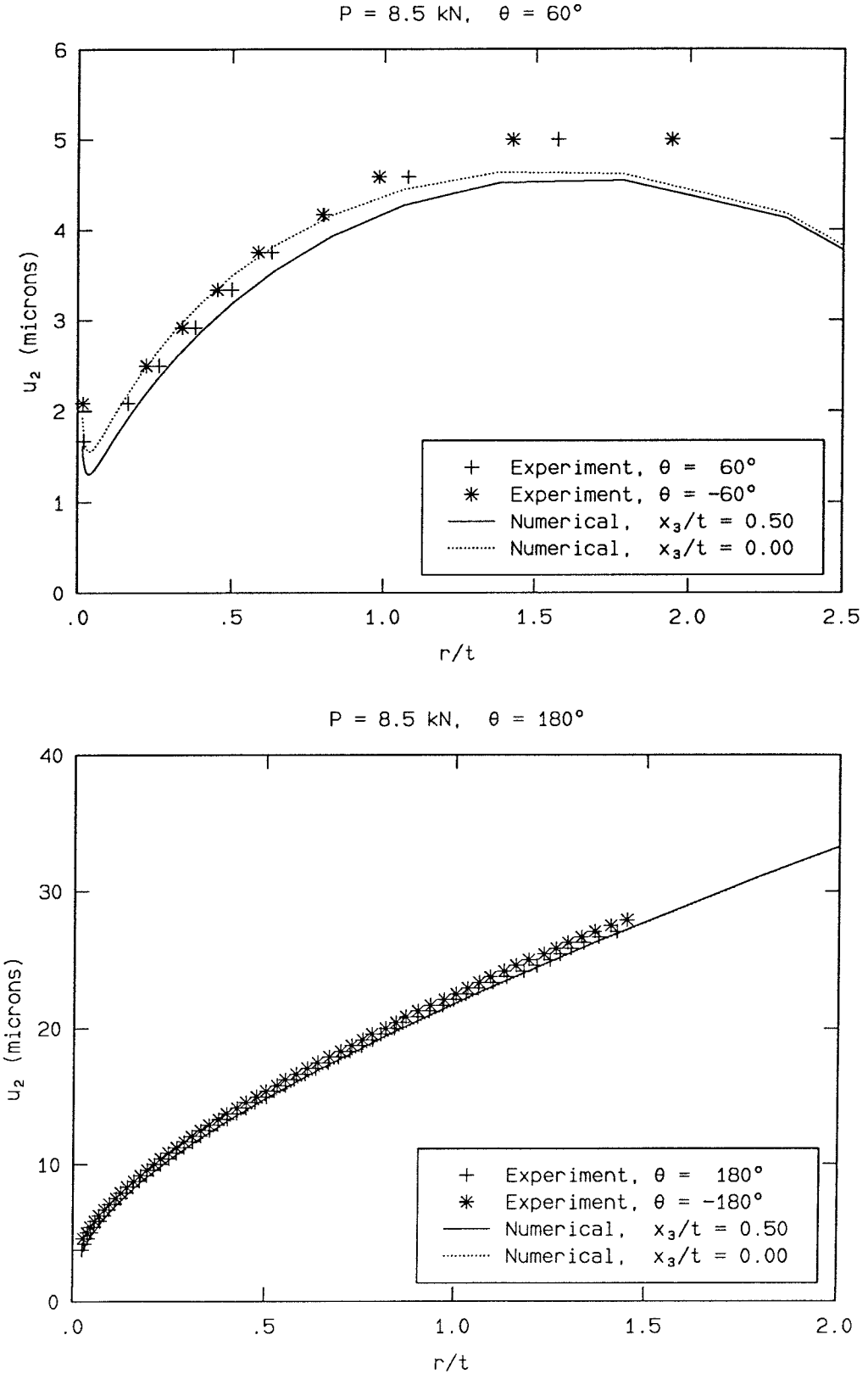


FIGURE 121. Radial variations of u_2, 8.5 kN.

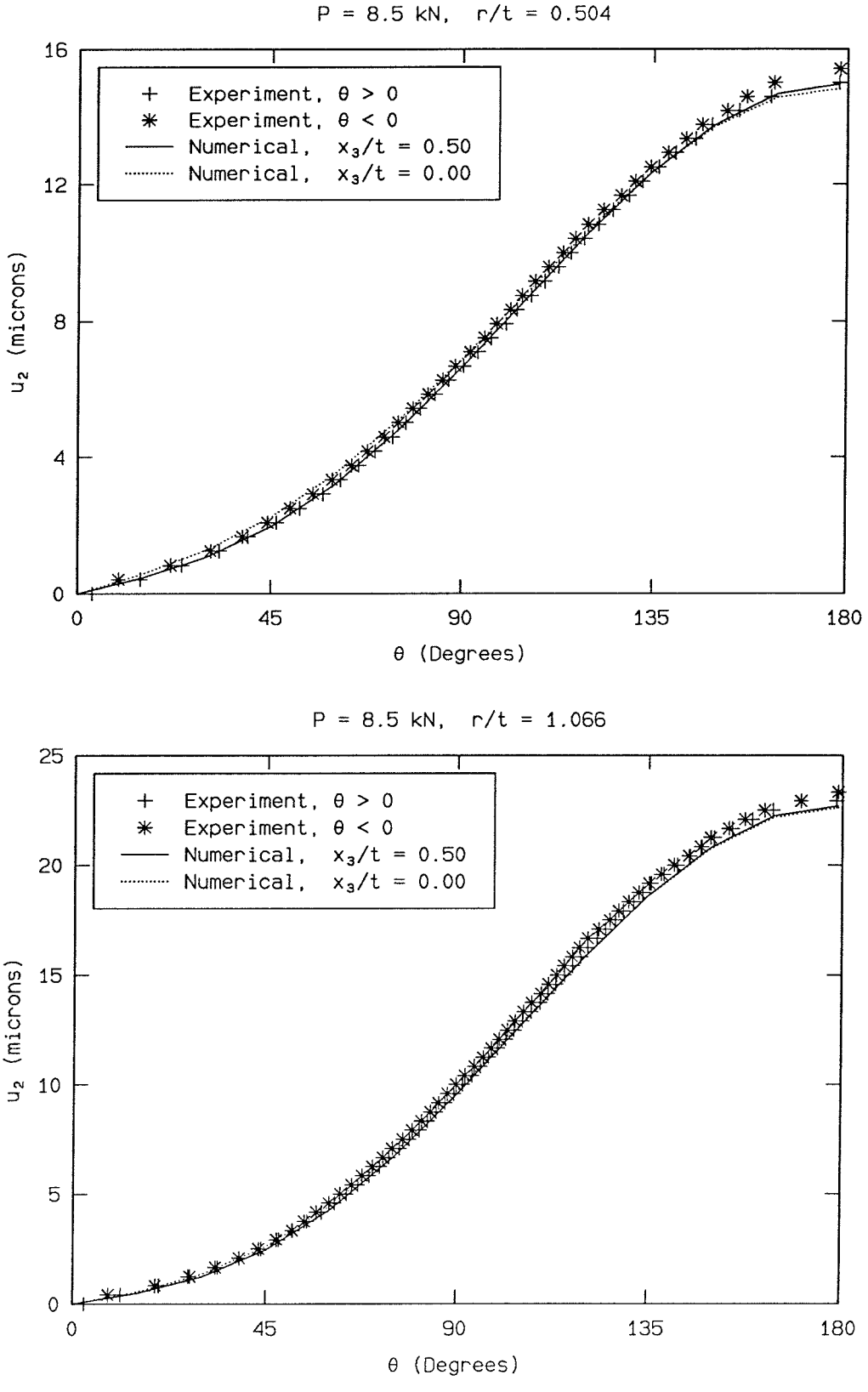


FIGURE 122. Angular variations of u_2, 8.5 kN.

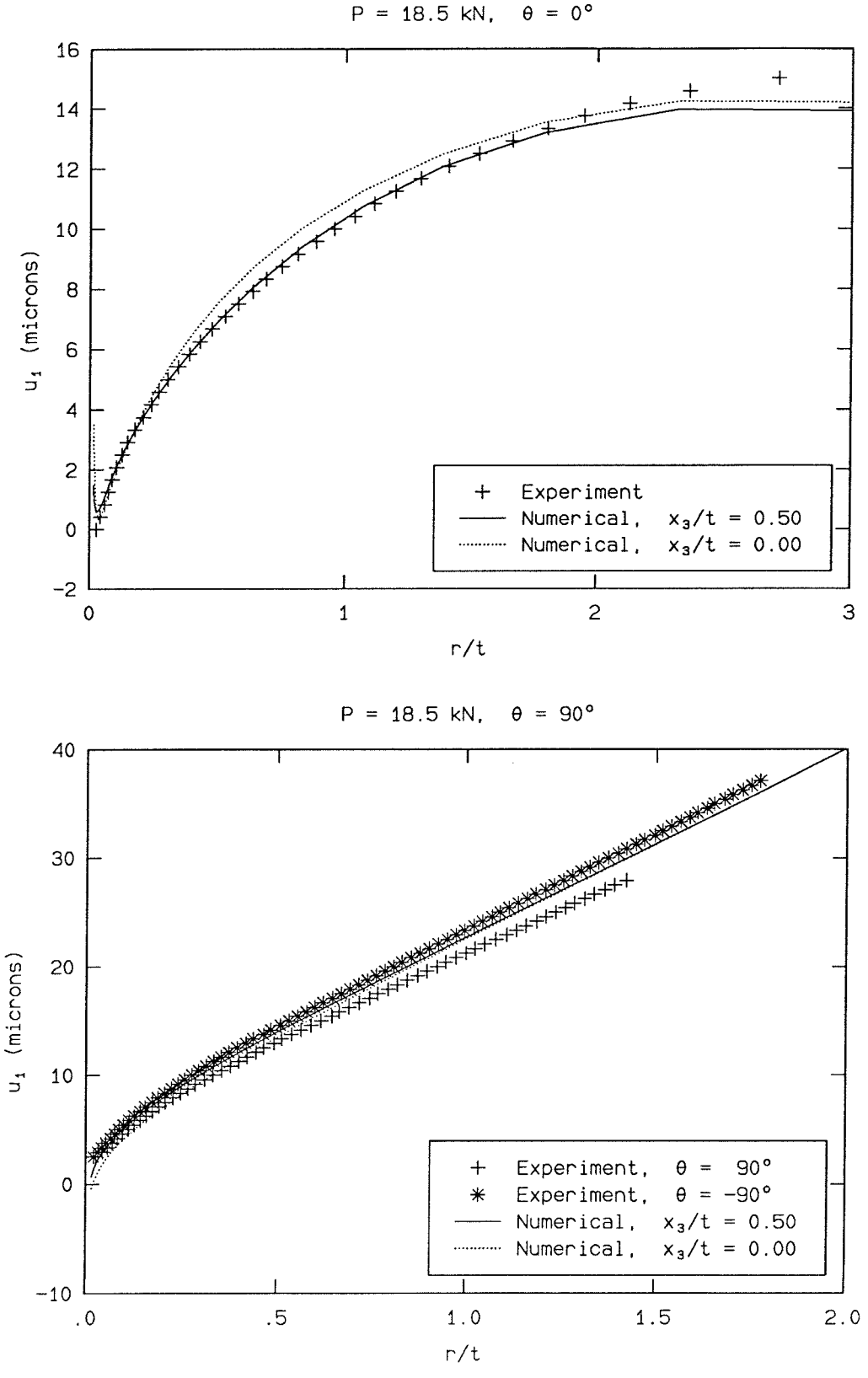


FIGURE 123. Radial variations of u_1, 18.5 kN.

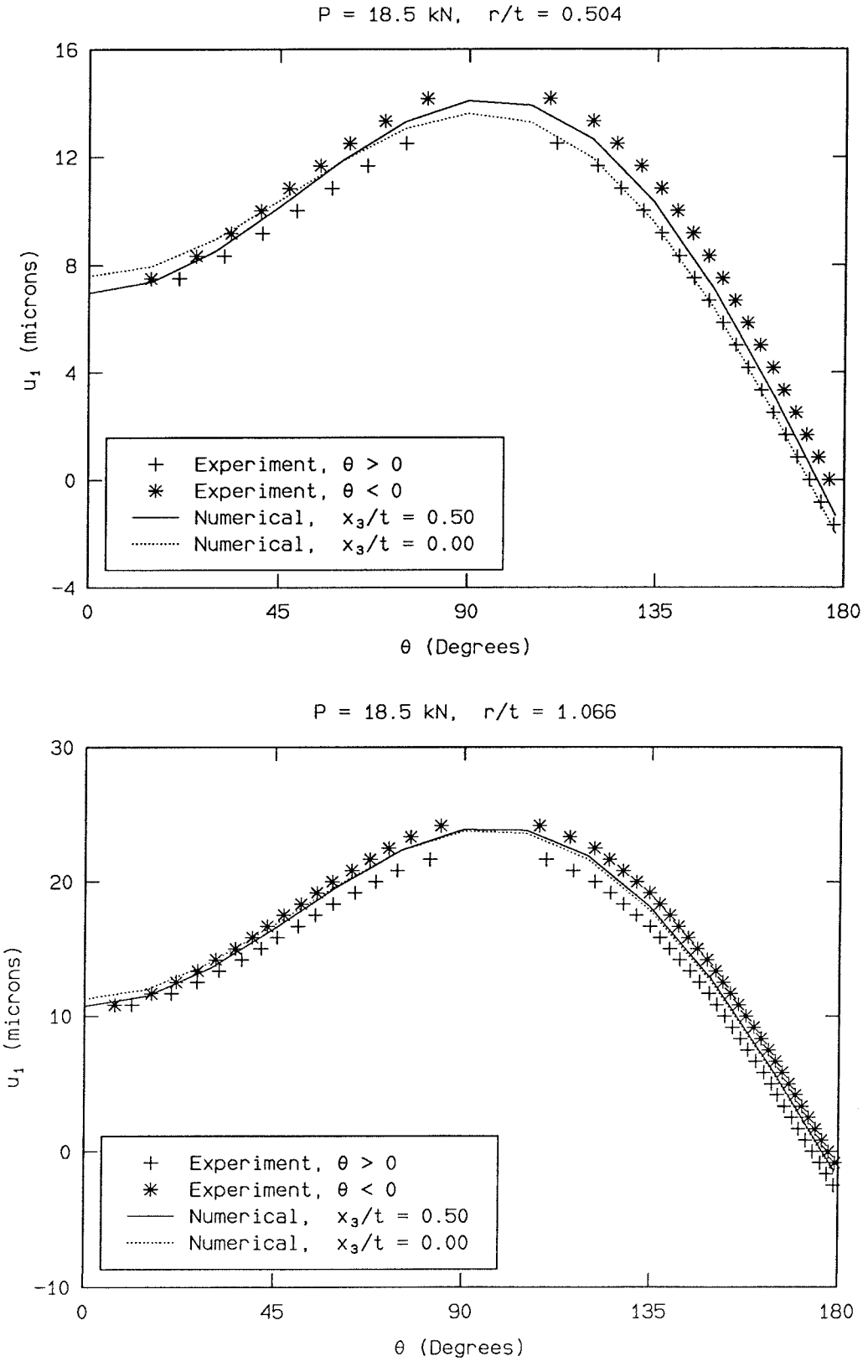


FIGURE 124. Angular variations of u_1, 18.5 kN.

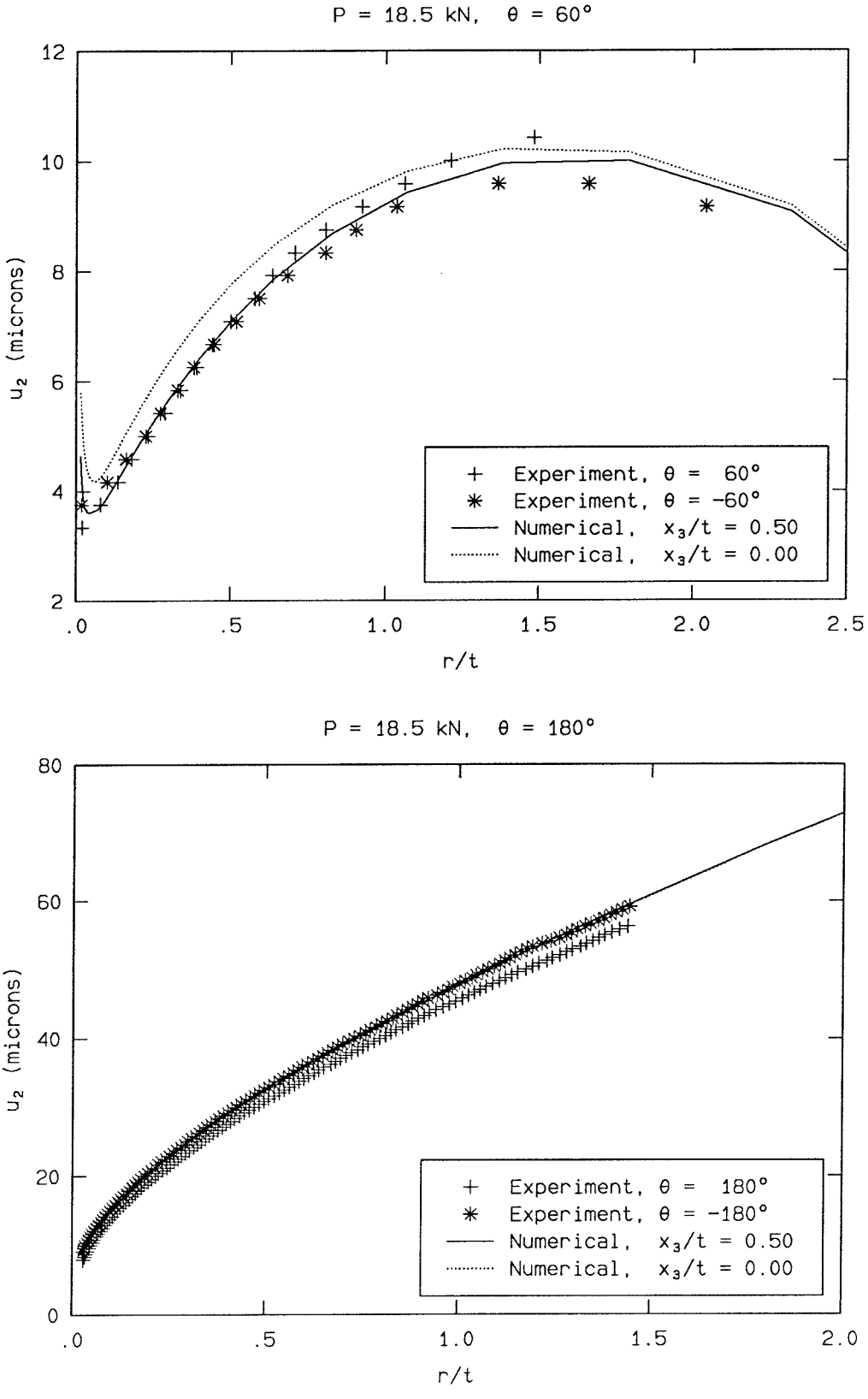


FIGURE 125. Radial variations of u_2, 18.5 kN.

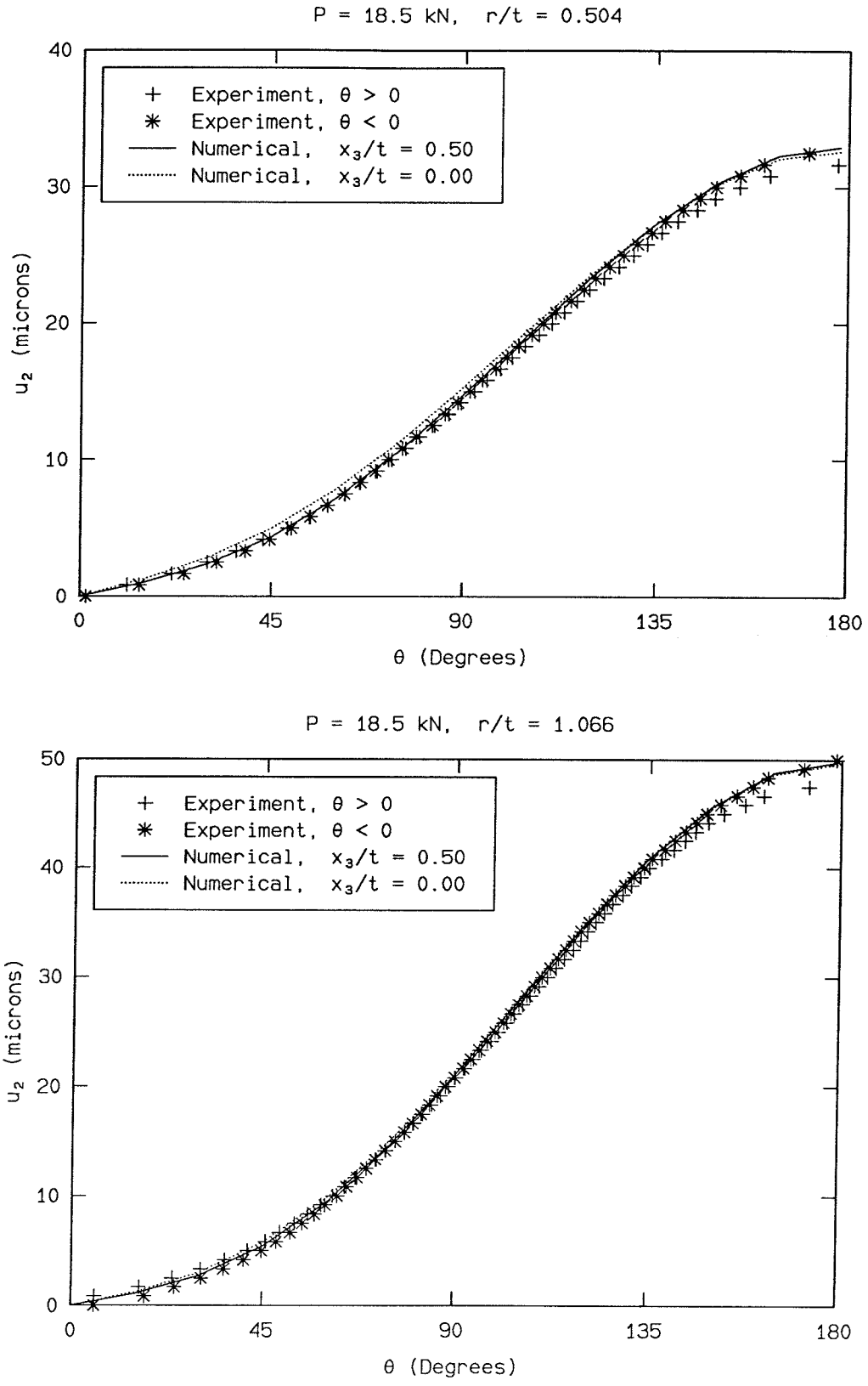


FIGURE 126. Angular variations of u_2, 18.5 kN.

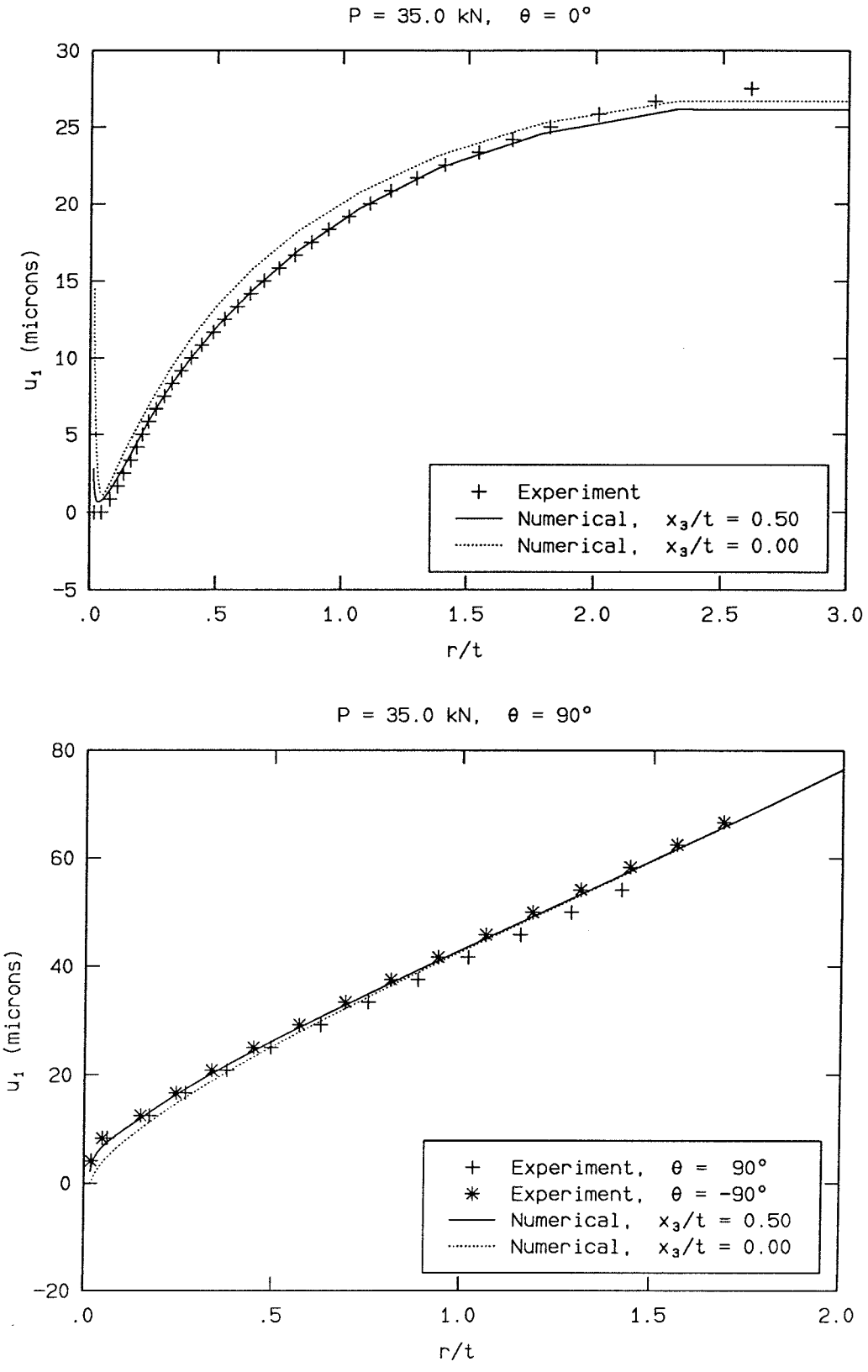


FIGURE 127. Radial variations of u_1, 35.0 kN.

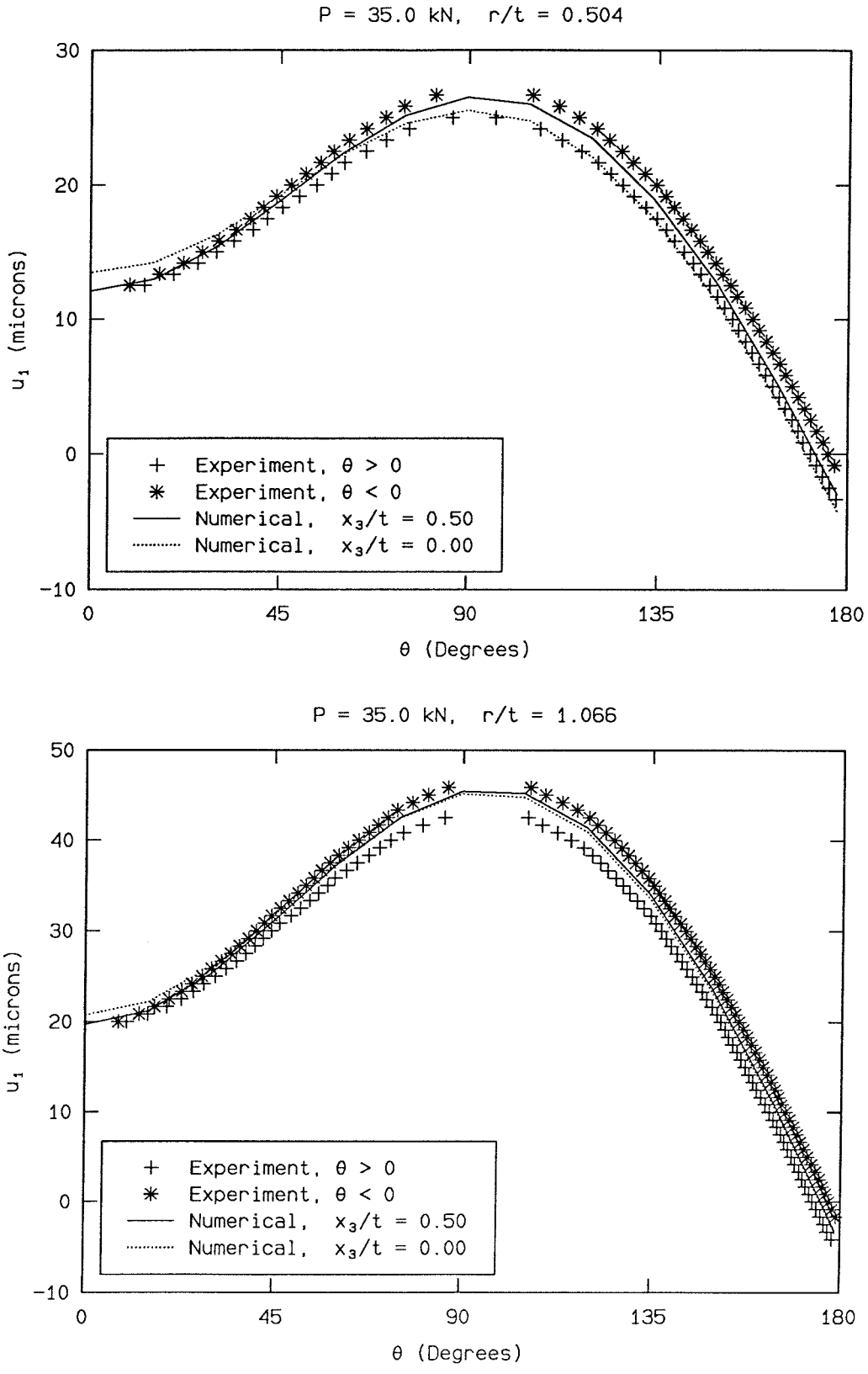


FIGURE 128. Angular variations of u_1, 35.0 kN.

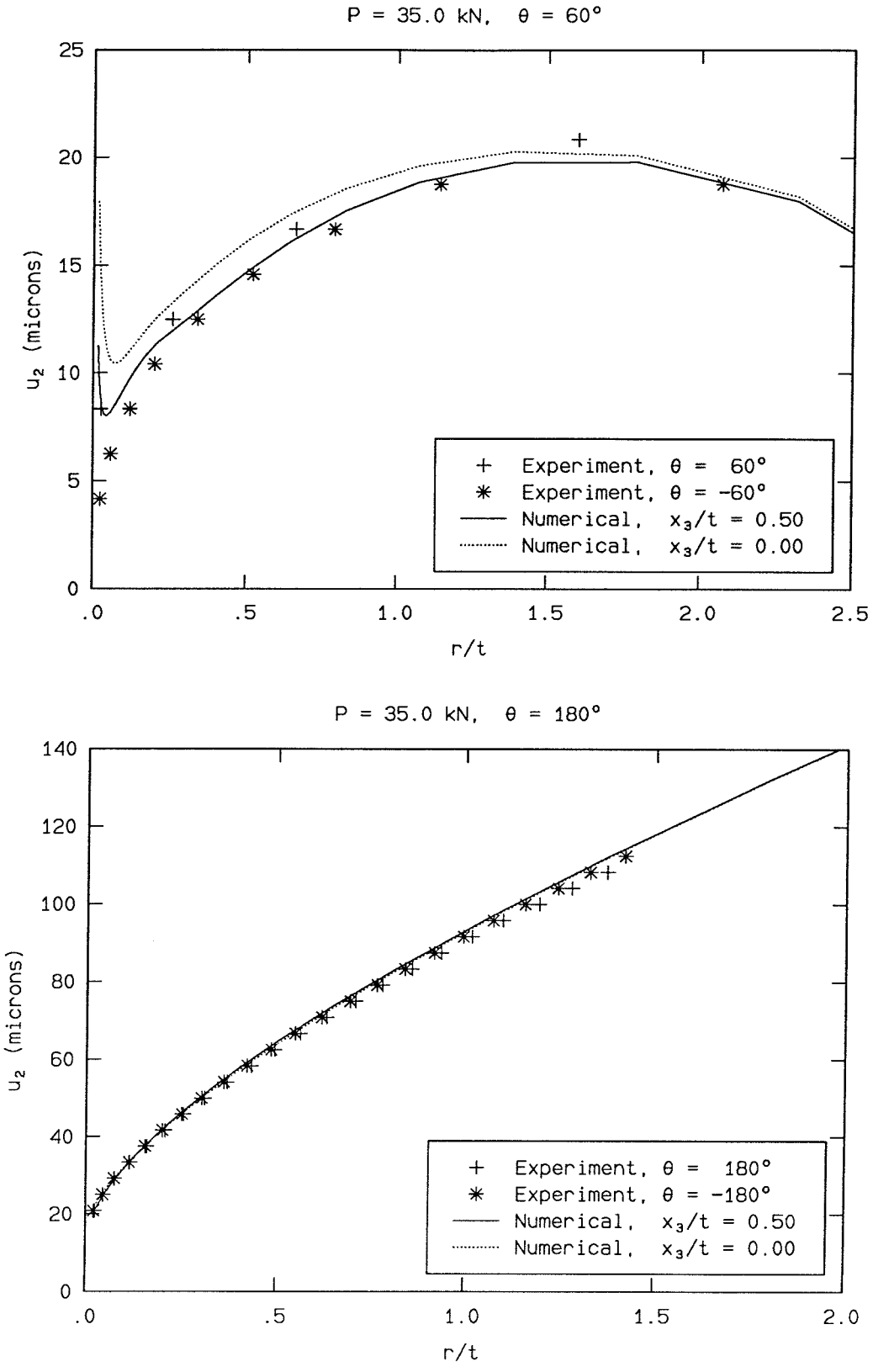


FIGURE 129. Radial variations of u_2, 35.0 kN.

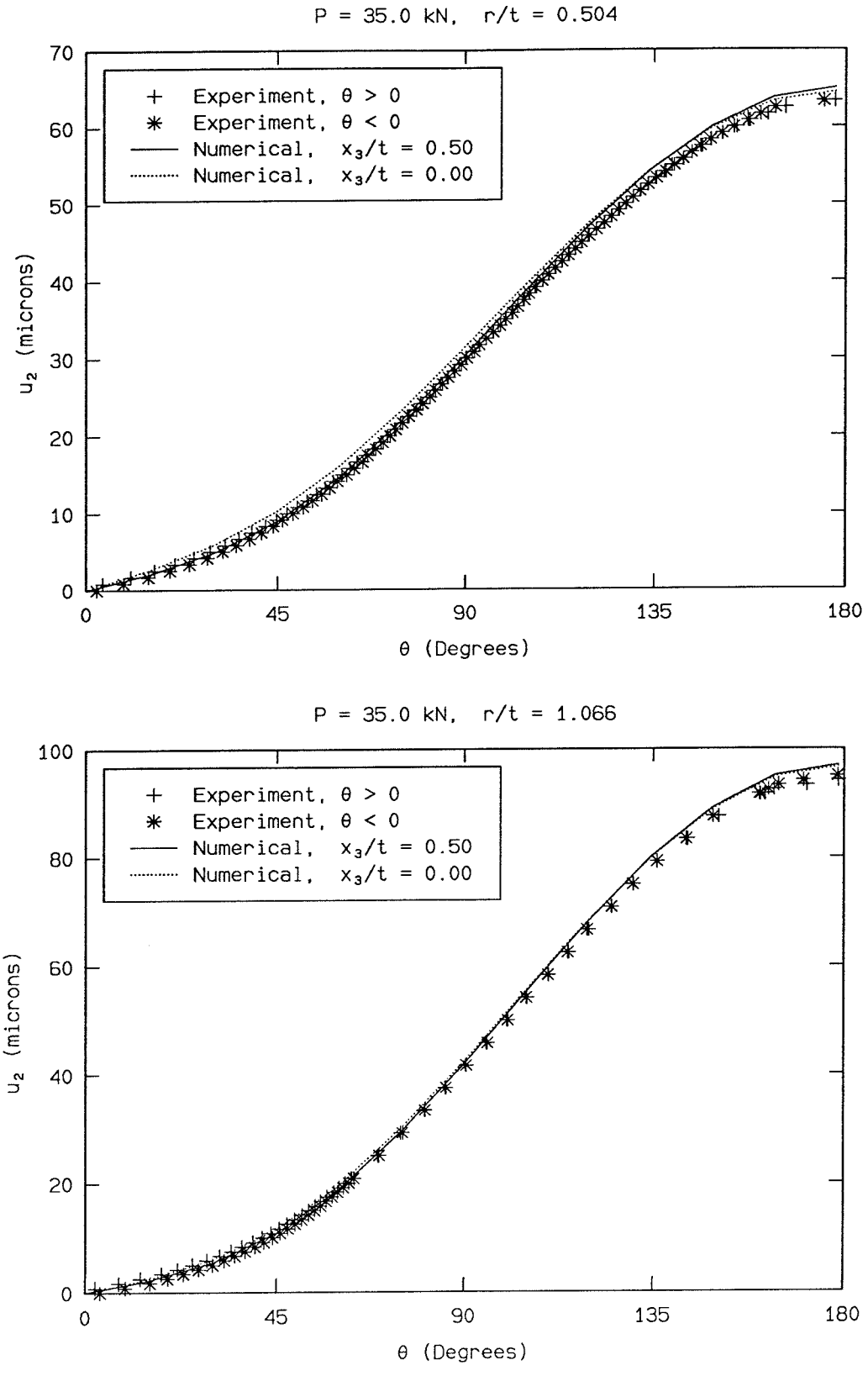


FIGURE 130. Angular variations of u_2, 35.0 kN.

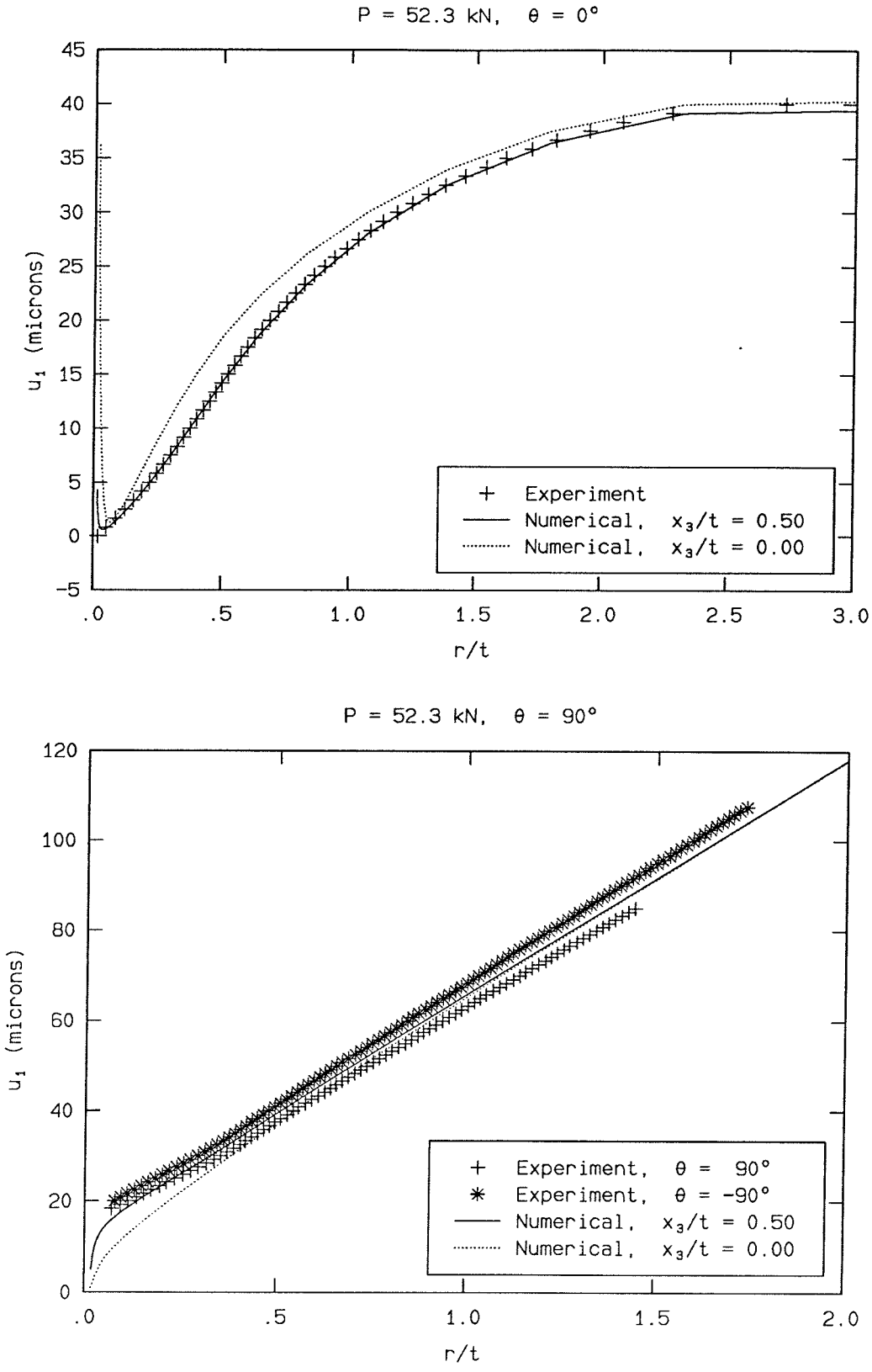


FIGURE 131. Radial variations of u_1, 52.3 kN.

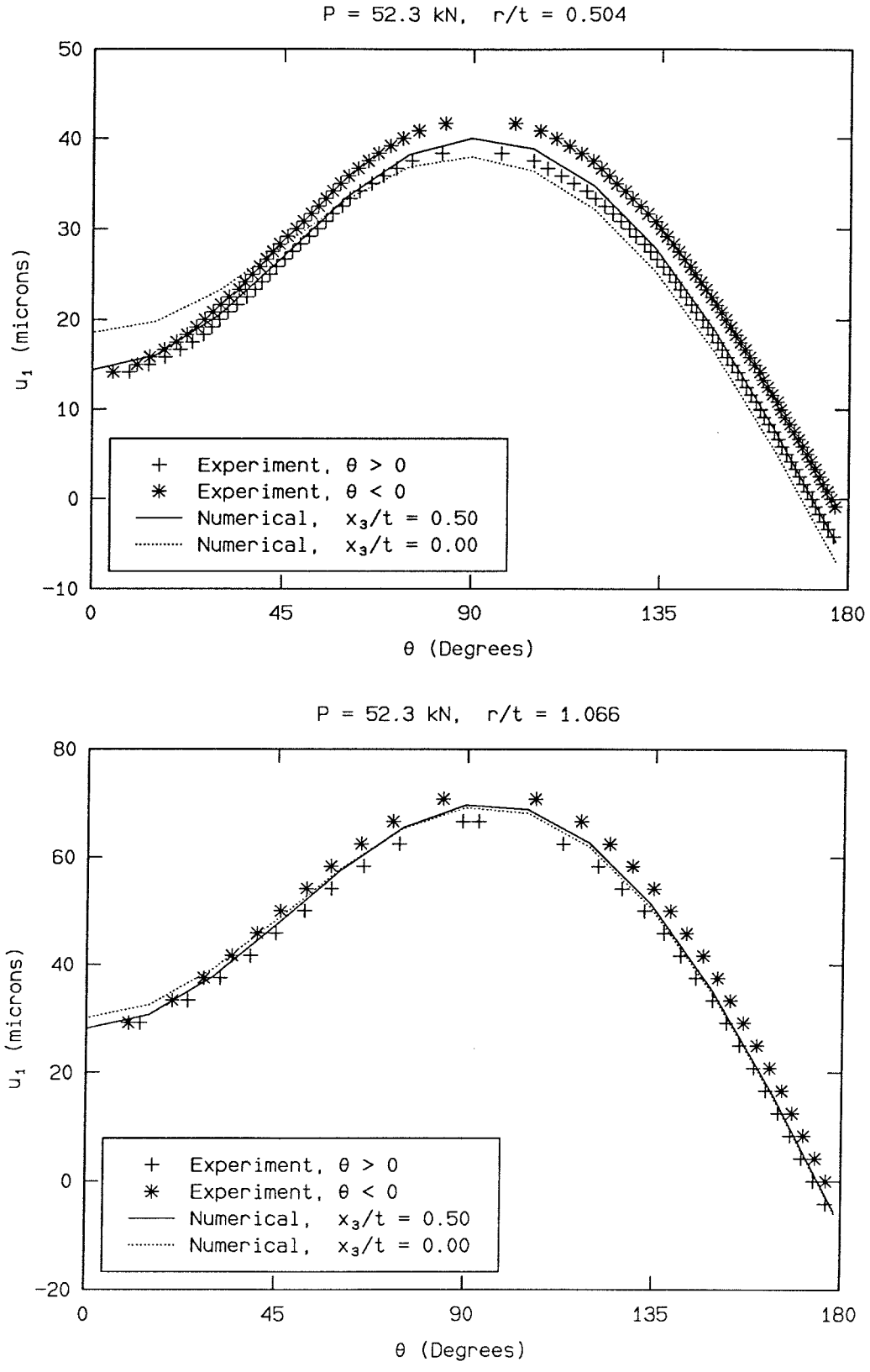


FIGURE 132. Angular variations of u_1, 52.3 kN.

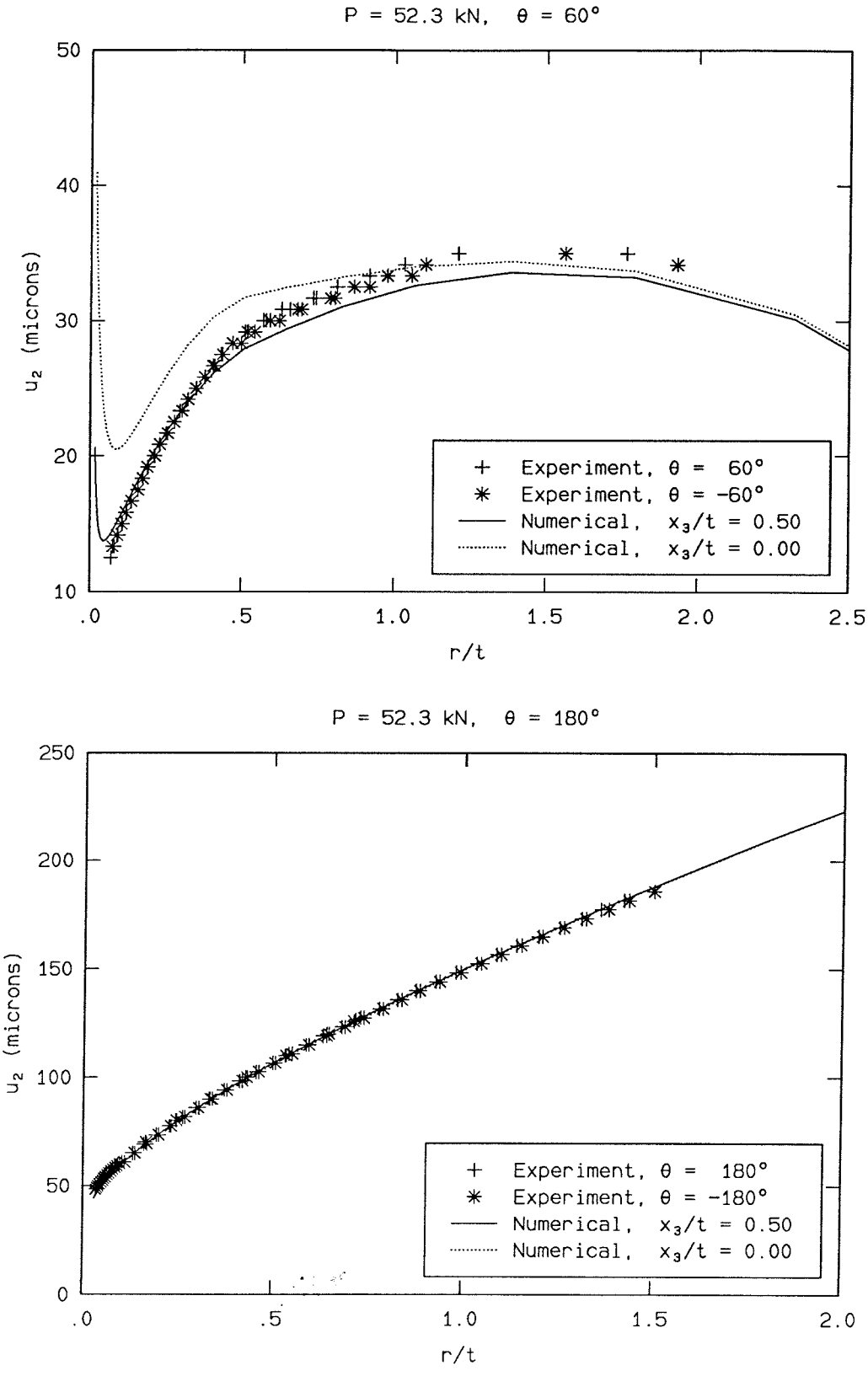


FIGURE 133. Radial variations of u_2, 52.3 kN.

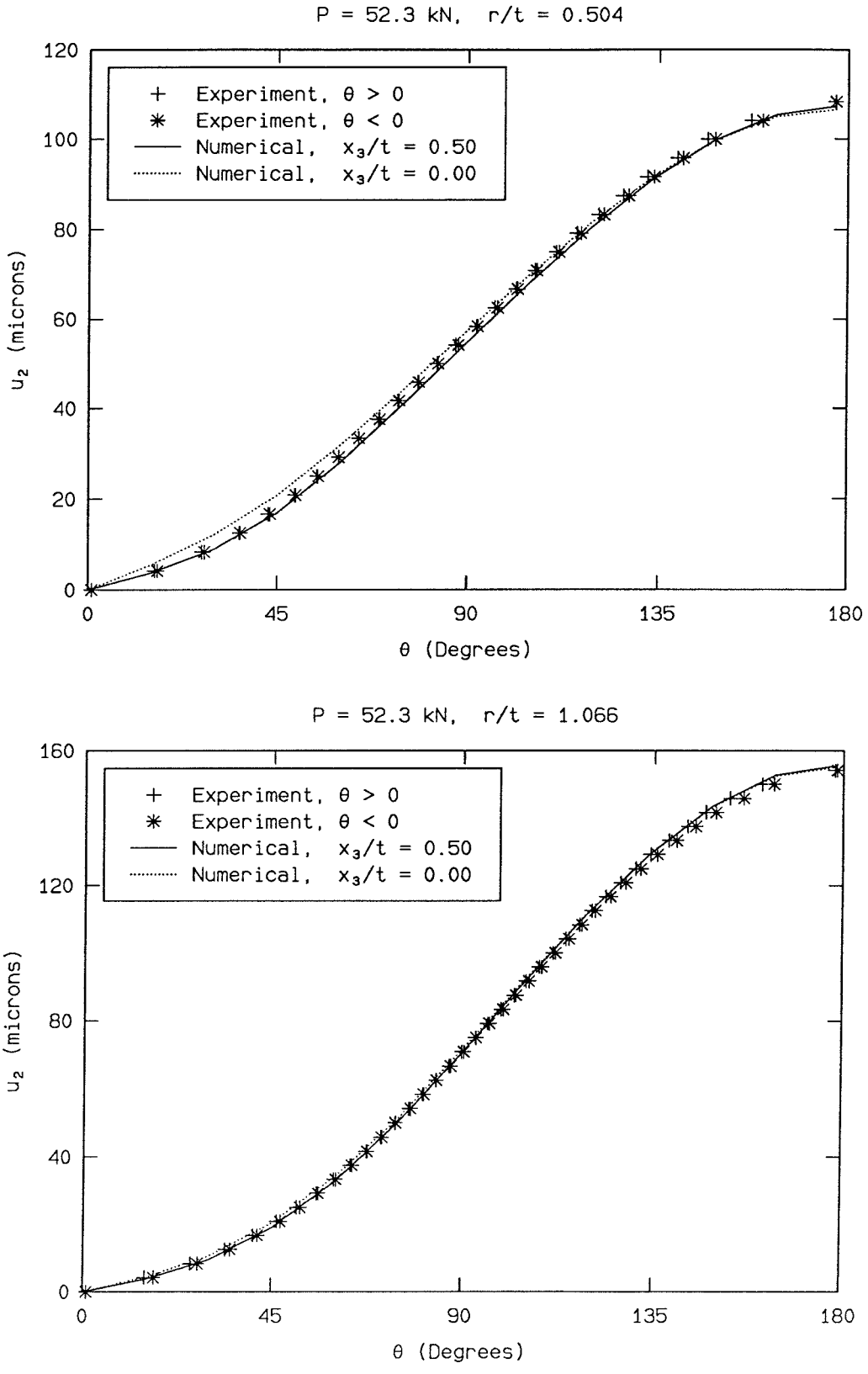


FIGURE 134. Angular variations of u_2, 52.3 kN.

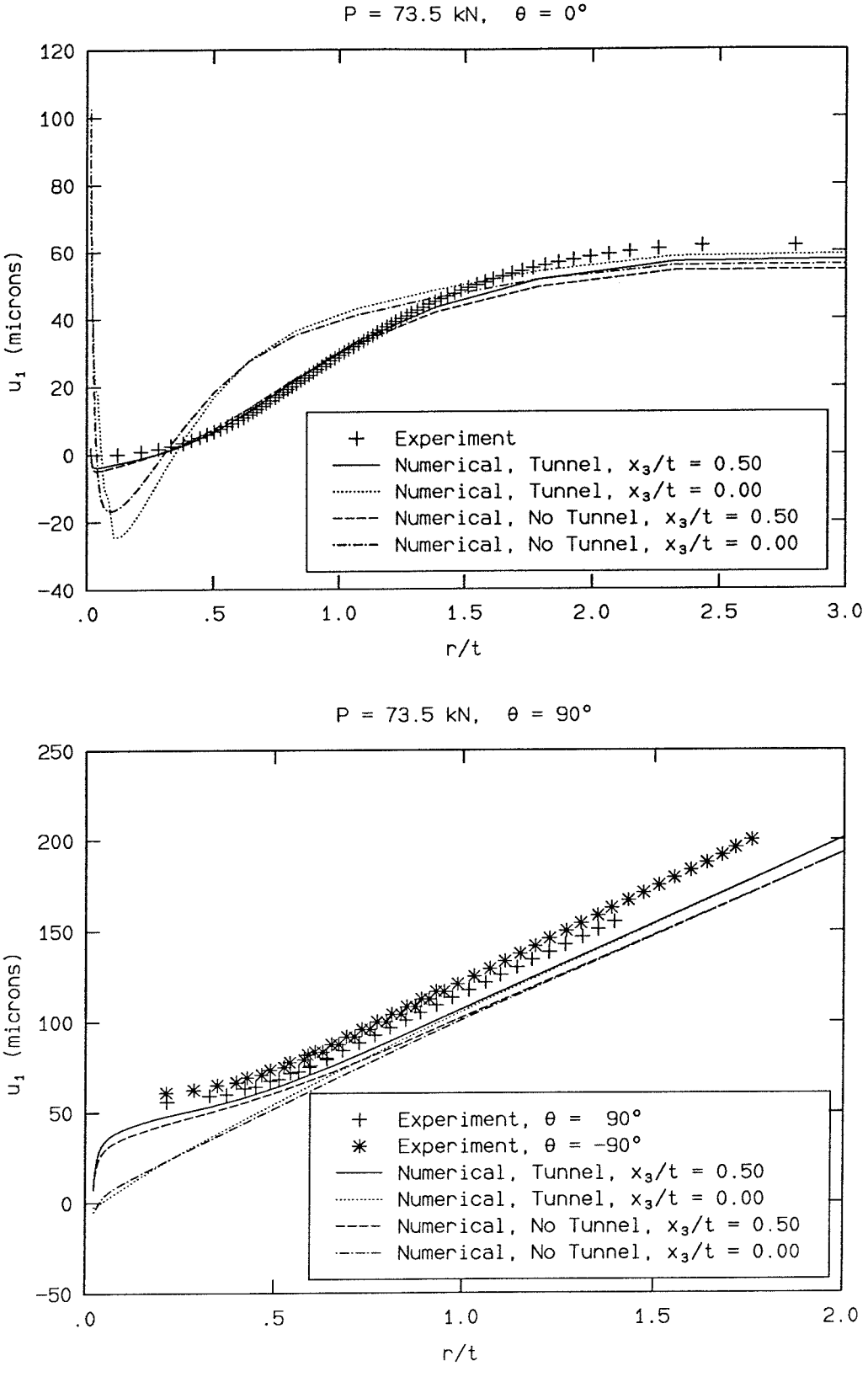


FIGURE 135. Radial variations of u_1, 73.5 kN.

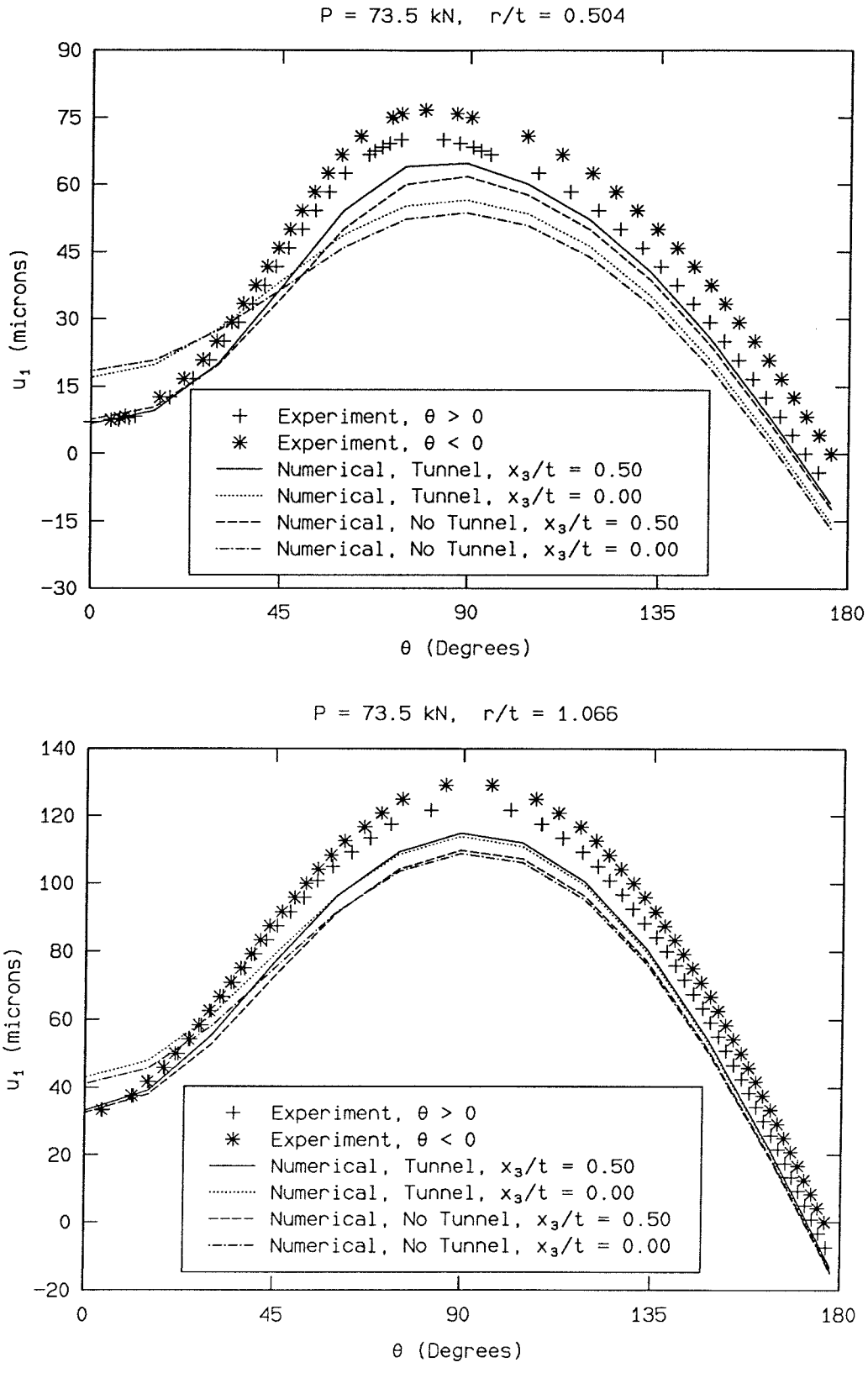


FIGURE 136. Angular variations of u_1, 73.5 kN.

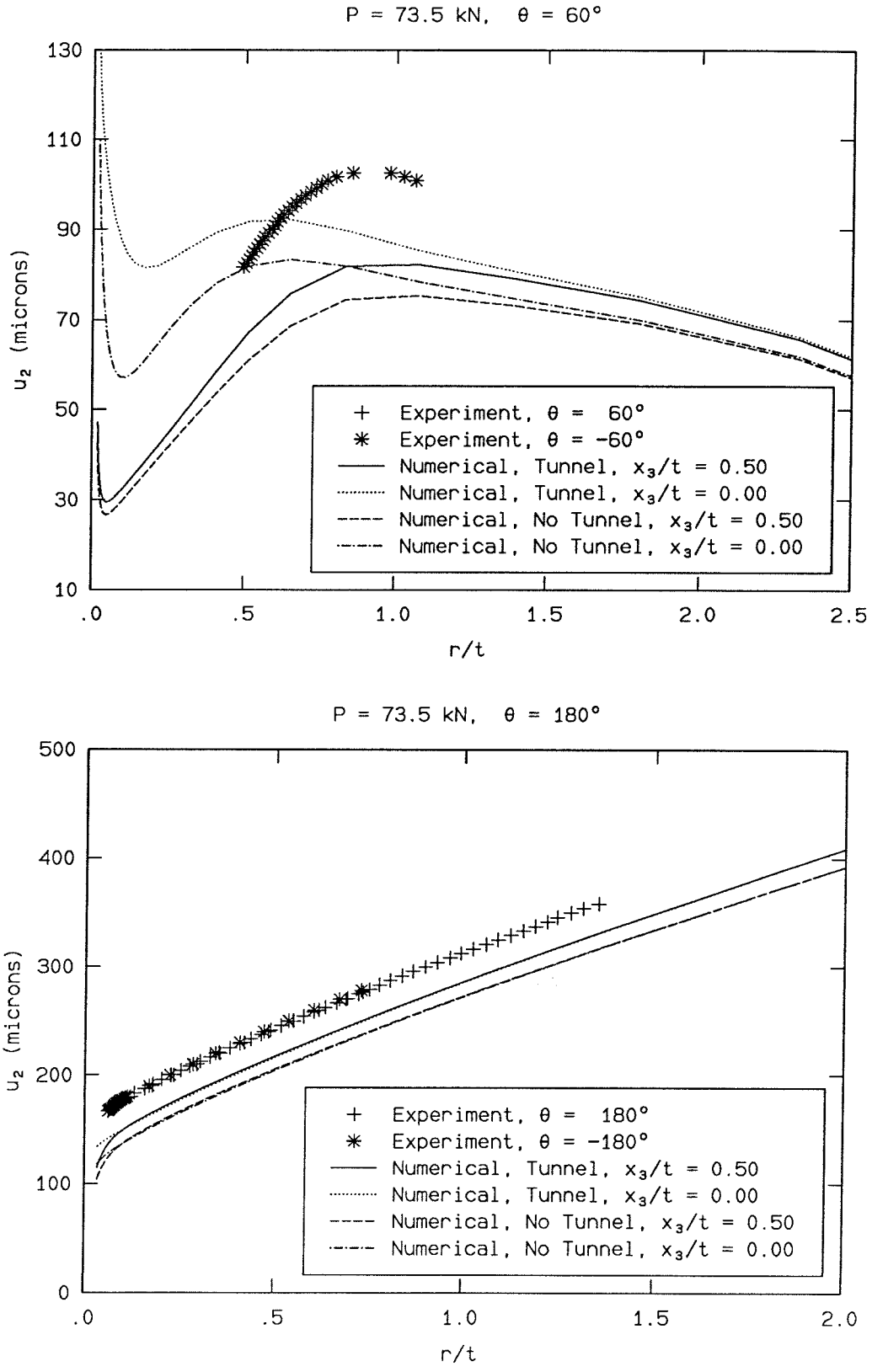


FIGURE 137. Radial variations of u_2, 73.5 kN.

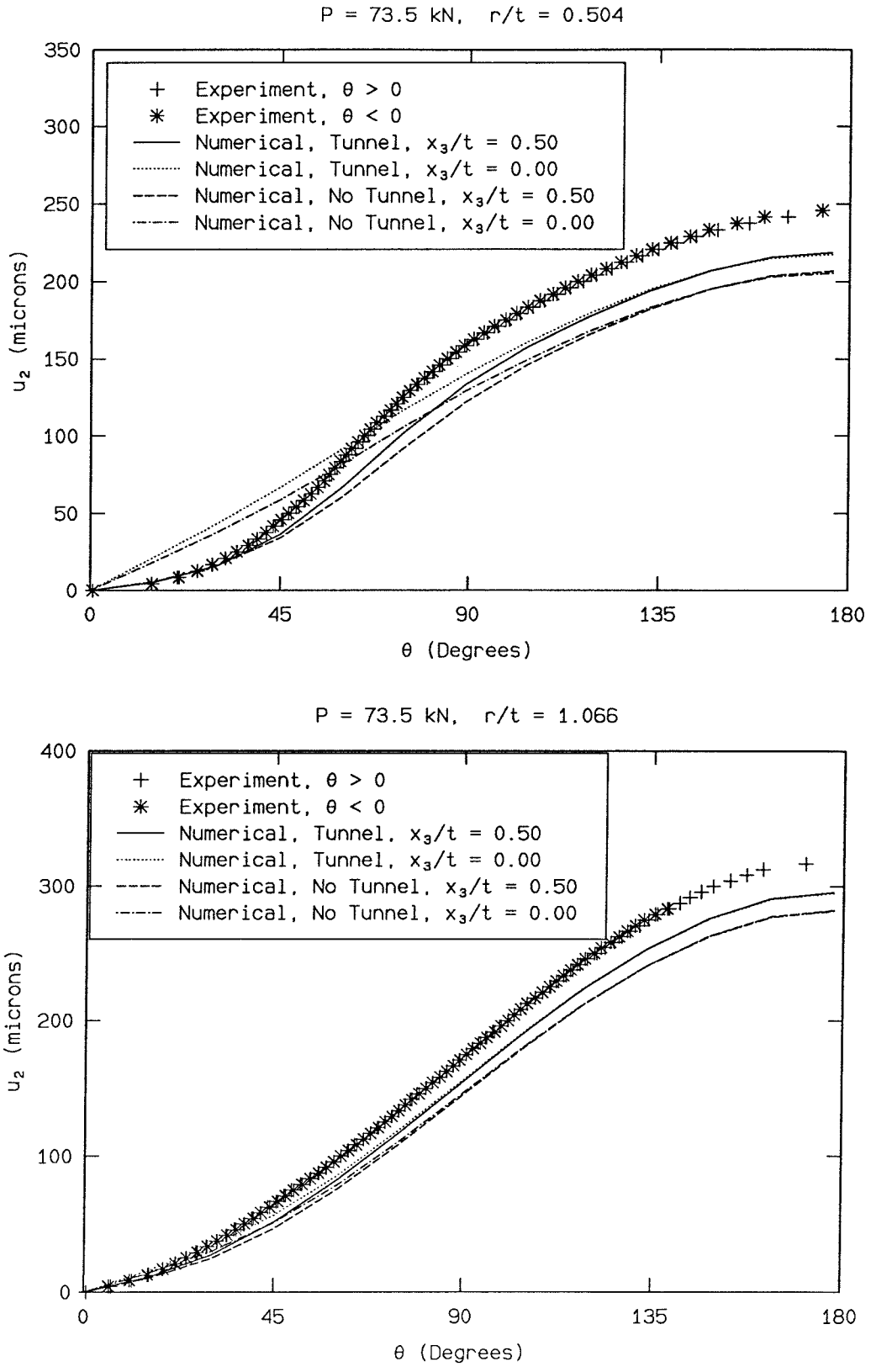


FIGURE 138. Angular variations of u_2, 73.5 kN.

5.2.1 Effects of Added Tunnel

Although the inclusion of the tunnel does not make the numerical data compare exactly with the experiment at 73.5 kN, the addition of the the tunnel does push the numerical displacements in the proper direction. In particular, one would like to determine whether the added tunnel introduces the effect seen in the u_2 fringe pattern ahead of the notch tip at 73.5 kN, where the fringes show a region of lower fringe density bracketed by two regions of higher fringe density, as seen in Figure 114. The added tunnel does cause some unloading at the free surface in the finite-element model, which is located in approximately the same position as the region of lower fringe density in the photograph (see Figure 114 in Section 5.1).

This effect will be shown graphically by plotting the angular variation of the u_2 displacement fringe for two radii, one that passes through the region of lower fringe density and one that passes through the region of higher fringe density farther from the notch tip. Experimental and numerical data for these two radii are shown in Figure 139 for the load of 52.3 kN, for comparison. It can be seen that the data for the two radii are nearly identical for $\theta < 45°$, after which the two curves in each plot deviate continuously with increasing θ. Figure 140 shows similar data for 73.5 kN. In this figure, the experimental data show a deviation between the two radii at very small theta, and the curves then grow closer at about $\theta = 75°$. The numerical data also show a deviation between the two radii at very small theta, which indicates that some part of this effect may be due to the size of the plastic zone. The added tunnel does seem to cause the proper sort of change in the data from the two radii at $\theta = 75°$, causing a larger increase in the data at $r/t = 0.504$ compared to $r/t = 1.066$.

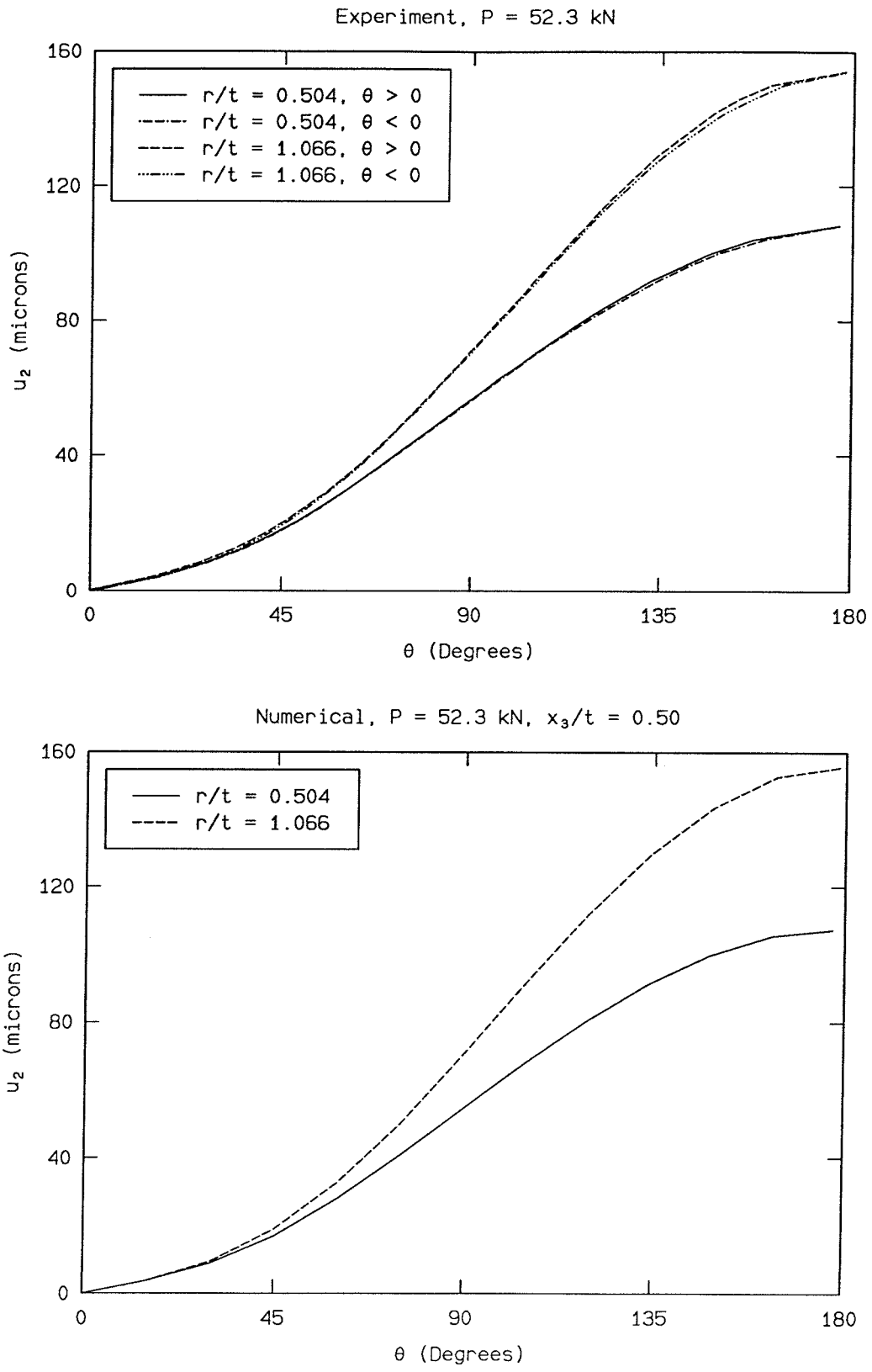
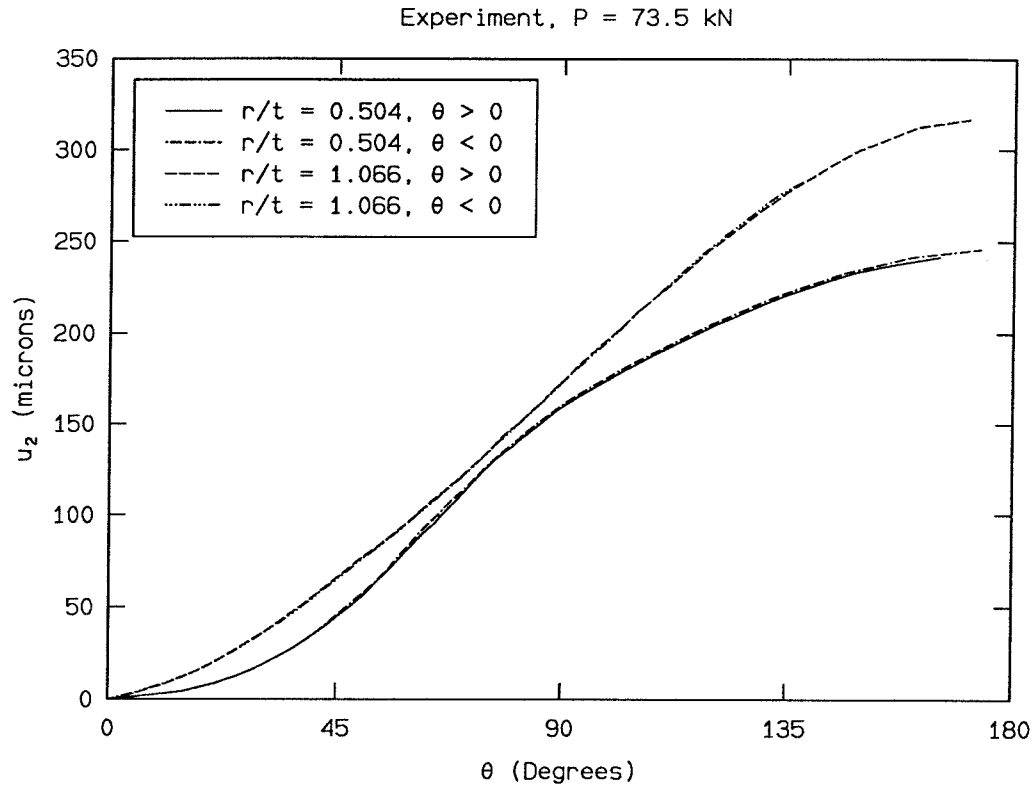


FIGURE 139. Angular variation of u_2. Experiment and Numerical.



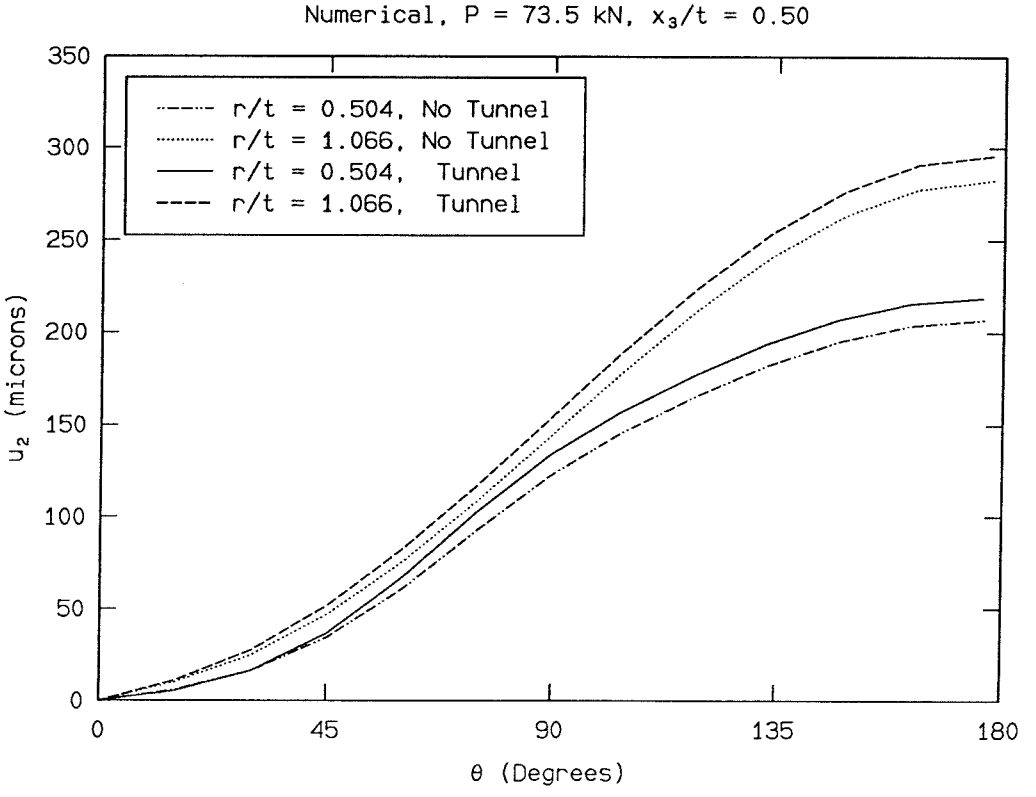


FIGURE 140. Angular variation of u_2. Experiment and Numerical.

5.3 Plastic Zones

An estimate of the size of the plastic zone surrounding the notch tip may be made from the finite–element model. The stresses and strains are calculated at the center Gauss point of each element. In the accompanying figures, the elements that are actively yielding are shown shaded in gray; these are elements in which the stress state is currently expanding the size of the local yield surface. When the tunnel is introduced through the release of the boundary constraints on the crack line, some elements that have previously been actively yielding begin to unload. These elements that are no longer actively yielding are outlined but not shaded. Those elements that are still deforming elastically are not drawn.

Plastic zones are shown for the layer of elements nearest the center of the specimen ($x_3/t = 0.079$) and the layer of elements nearest the free surface ($x_3/t = 0.487$). It was anticipated that the plastic zones of these two different layers would be somewhat different, as the center layer of elements is expected to approach plane–strain conditions, while the free–surface layer is expected to approach plane–stress conditions. Broek [14] indicates that the plane–strain plastic zone should be smaller than the plane–stress plastic zone, and that the shapes of the two plastic zones should be different.

For the Von Mises yield criterion used in the code, Broek [14] gives the radial extent of the plastic zone, r_p, as a function of θ as

$$\begin{aligned} \text{Plane Strain}: \quad r_p &= \frac{K^2}{4\pi\sigma_0^2}\left[\frac{3}{2}\sin^2\theta + (1-2\nu^2)(1+\cos\theta)\right] \\ \text{Plane Stress}: \quad r_p &= \frac{K^2}{4\pi\sigma_0^2}\left[1 + \frac{3}{2}\sin^2\theta + \cos\theta\right]. \end{aligned} \tag{5.3.1}$$

K is the Mode I stress–intensity factor, and σ_0 is the initial yield stress in uniaxial tension. For this specimen geometry, K is found from the load P and the thickness t as $27.95\text{m}^{-1/2}P/t$ (see Section 4.2). These analytical estimates for the plastic zone size are derived using Williams's [90] K–field series solution for a crack in a linearly elastic body; a sharp crack is assumed and only the first singular stress term is retained, so that the stresses are assumed to vary as $r^{-1/2}$. This simple analytical approximation must underestimate the size of the plastic zone, as the

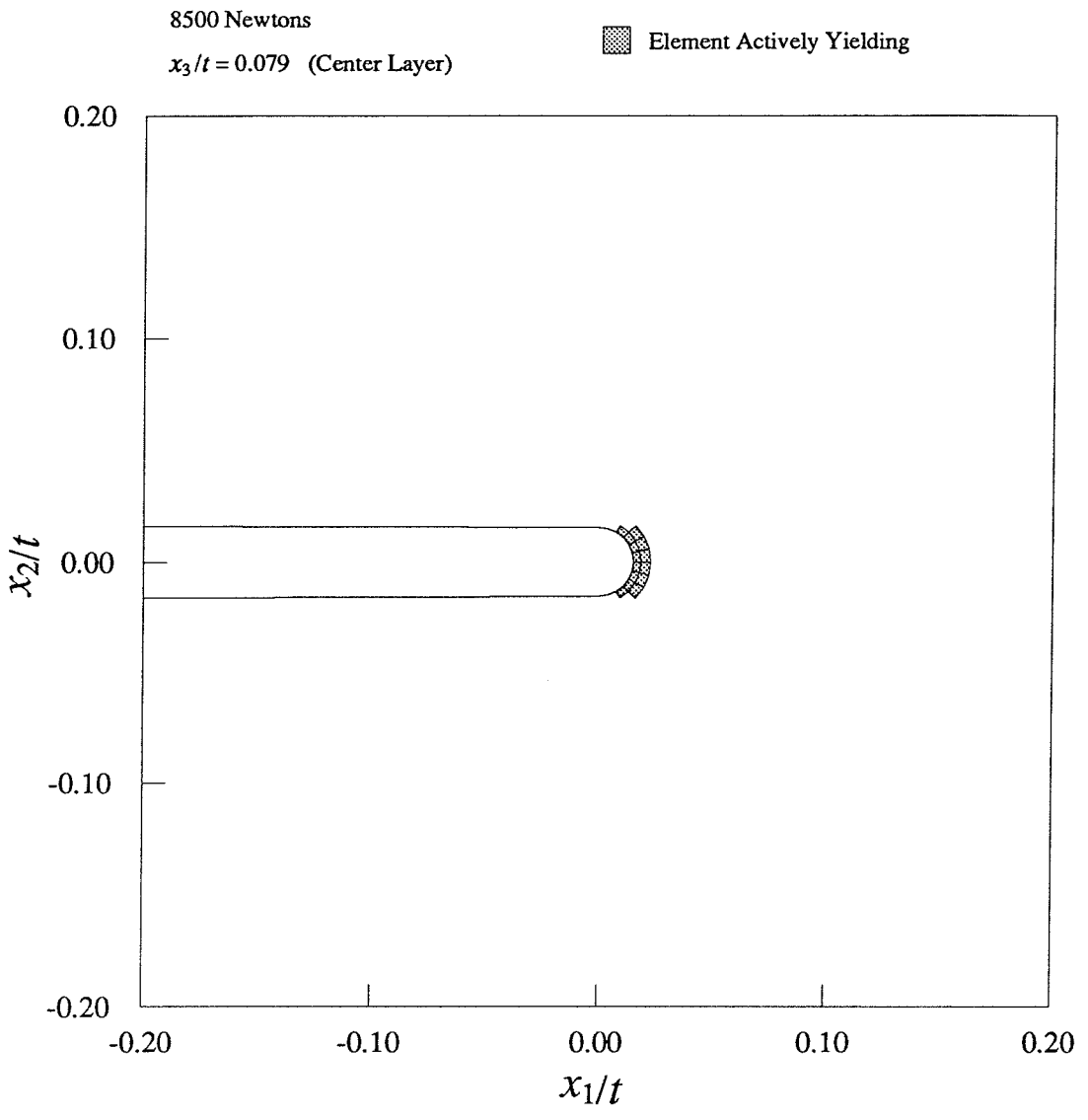


FIGURE 141. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is thirty times actual size. The plastic zone is shown on a deformed mesh.

stresses within the plastic zone must be smaller than the assumed elastic stresses, so that some of the load must be carried farther away from the crack, causing yielding at larger radii. In particular, the plastic zone should be considerably larger than these predictions along the line $\theta = 0$, where the highest stresses are expected to occur. The analytical estimates for the plastic zones are shown in the figures that

illustrate the actively yielding elements from the finite–element model. The origin of the coordinate frame is taken halfway between the center of curvature of the notch and the edge of the notch, as discussed in Section 4.2, and shown in Figure 86.

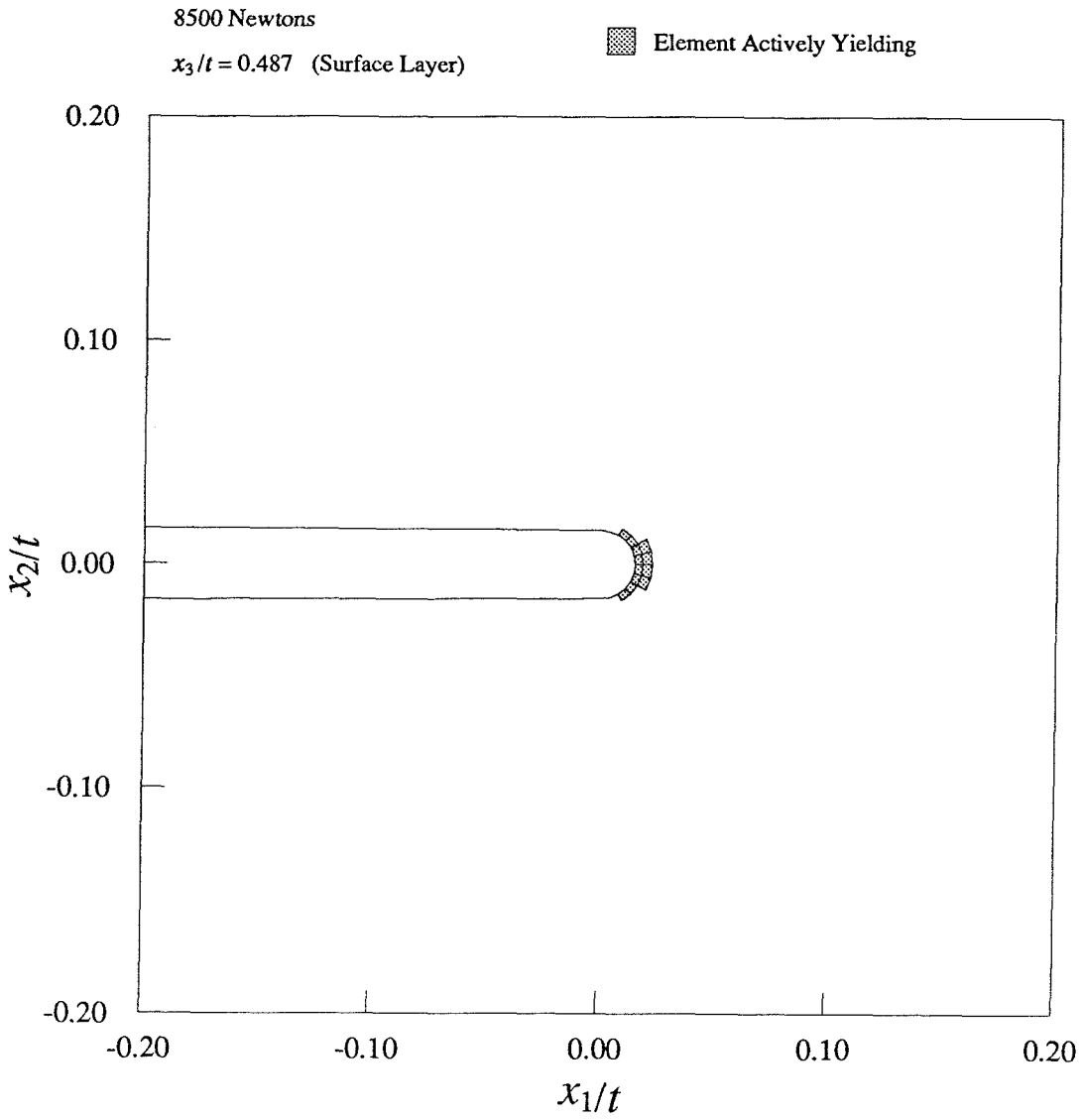


FIGURE 142. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is thirty times actual size. The plastic zone is shown on a deformed mesh.

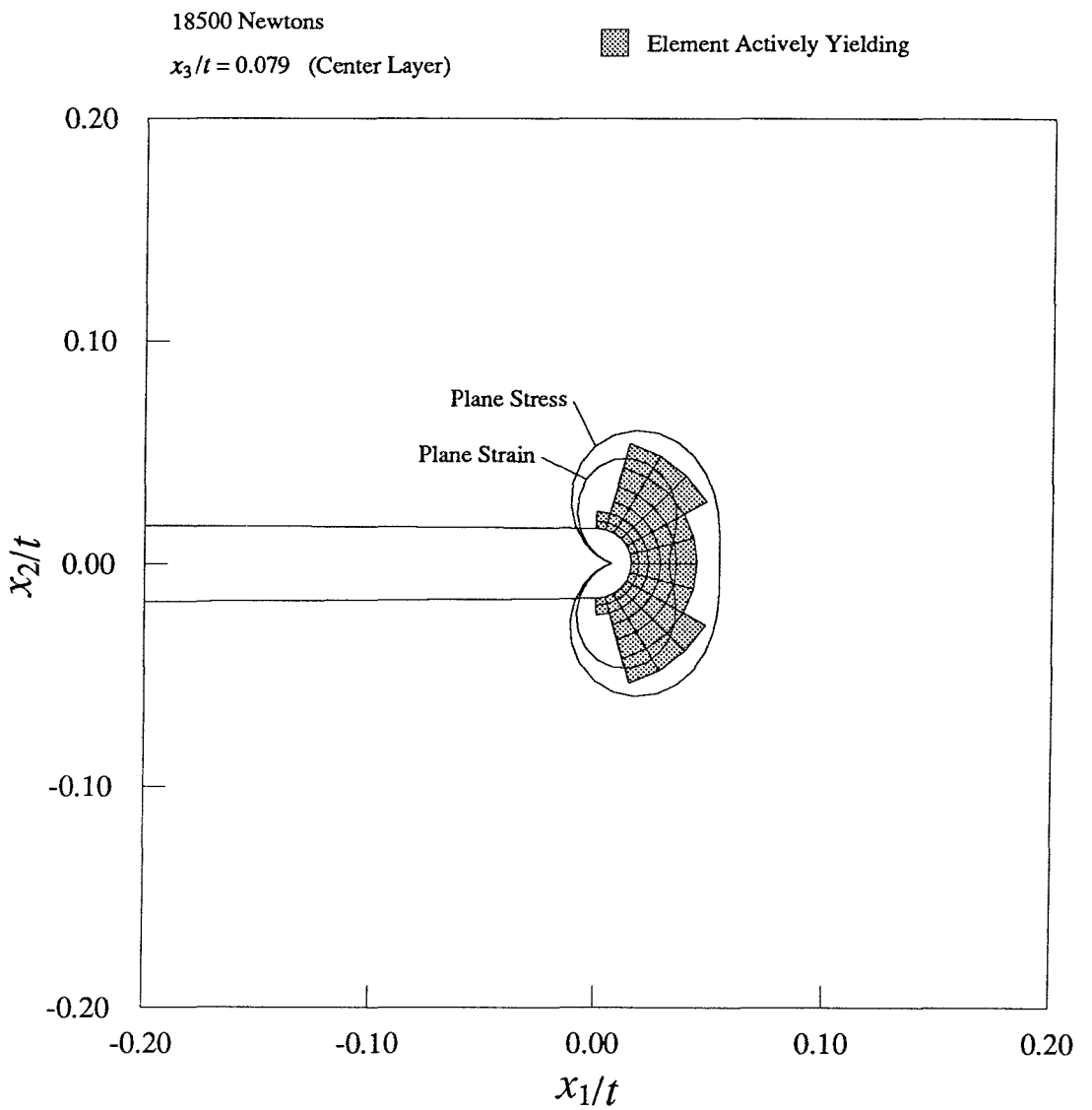


FIGURE 143. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is thirty times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

Figures 141 through 144 indicate the initiation of plastic deformation. These figures are shown on a very small size scale, in which the smallest elements are visible. Figures 141 and 142 show the plastic zones at the layer of elements nearest the center of the specimen and the layer of elements nearest the free surface taken

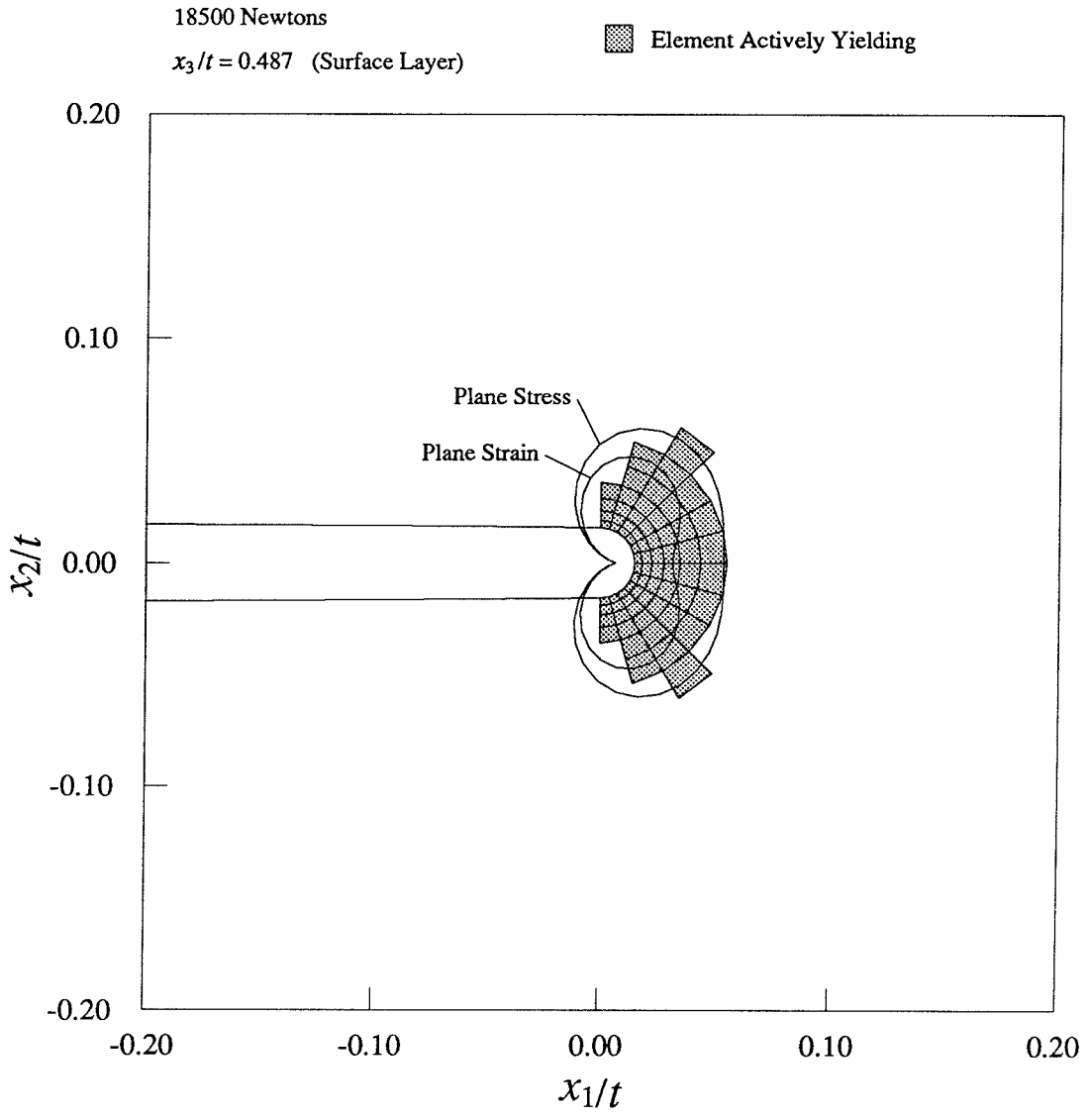


FIGURE 144. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is thirty times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

at 8500 Newtons, about 10% of the ultimate failure load. At this load, plastic deformation occurs only in the two rings of elements nearest the notch tip.

The analytical prediction is quite good at the load of 18.5 kN, and there seems to be good agreement with the Plane Strain estimate at the center of the specimen

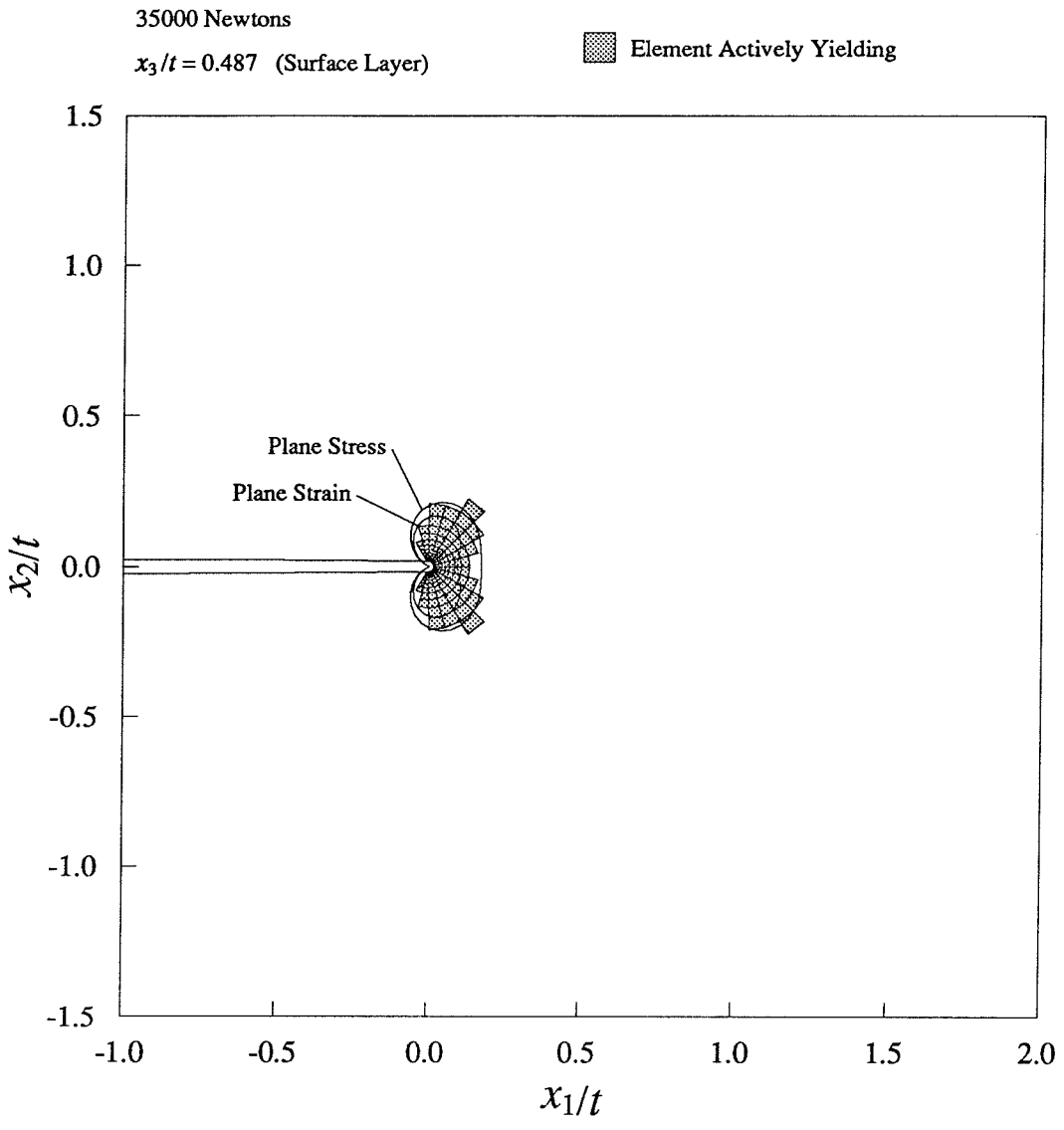


FIGURE 145. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

and the Plane Stress estimate at the free surface, as shown in Figures 143 and 144. It appears that this agreement would improve if the tip of the notch were taken as the origin of the K field estimate. However, at higher loads, the present study indicates that there is not a clear difference in either the size or the shape

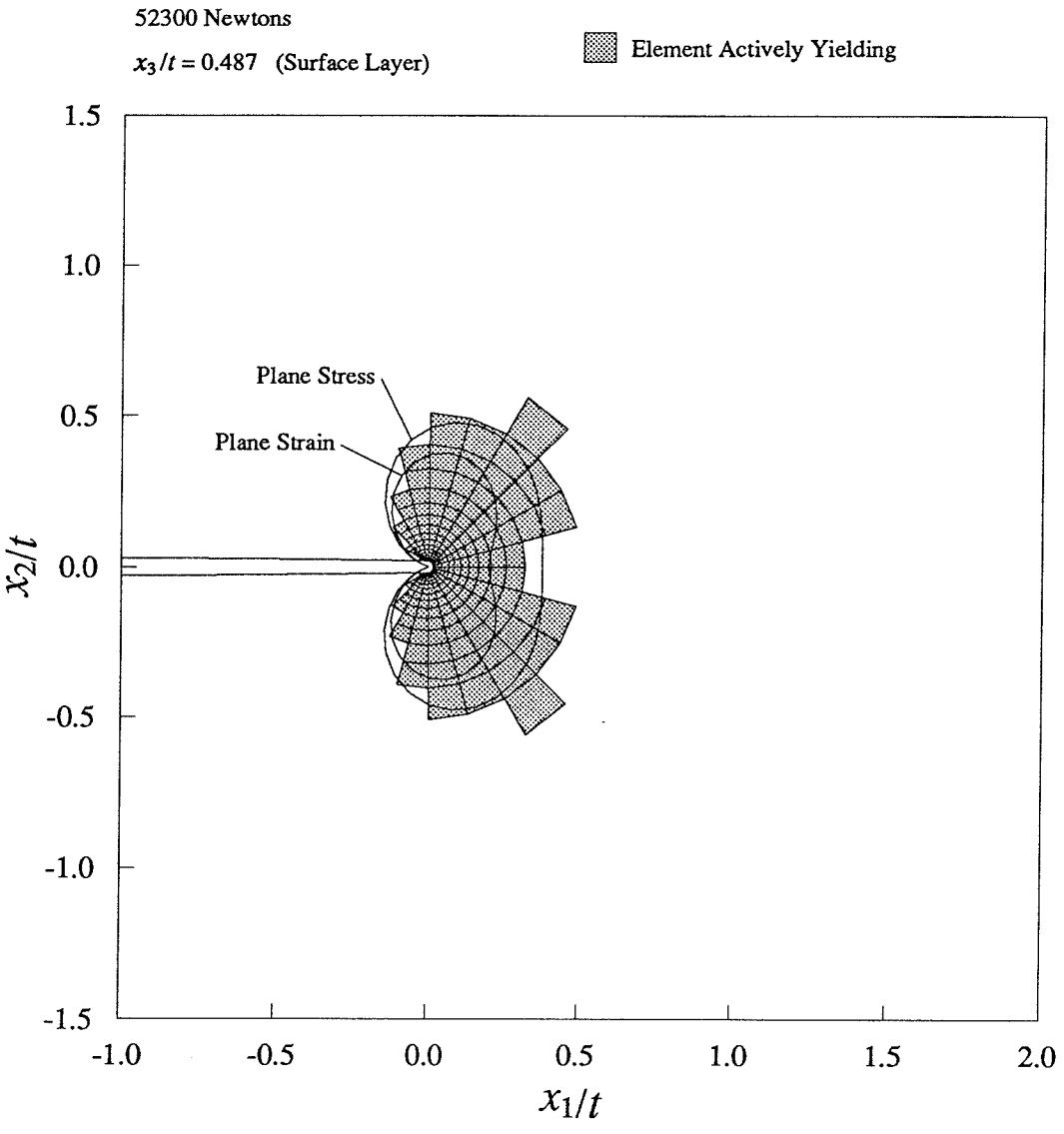


FIGURE 146. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

of the plastic zones from the two layers of elements, nor is there any clear trend in the development of the size and shape of the plastic zone at the center of the specimen as compared to the free surface. The difference in the size of the plastic zone between the two layers seems to be on the order of the size of one element in

any direction. At most, it seems that the plastic zone at the free surface extends slightly farther behind the notch tip and perpendicular to the notch at the tip than the plastic zone at the center of the specimen. The plastic zone at the center seems to extend slightly farther directly ahead of the notch.

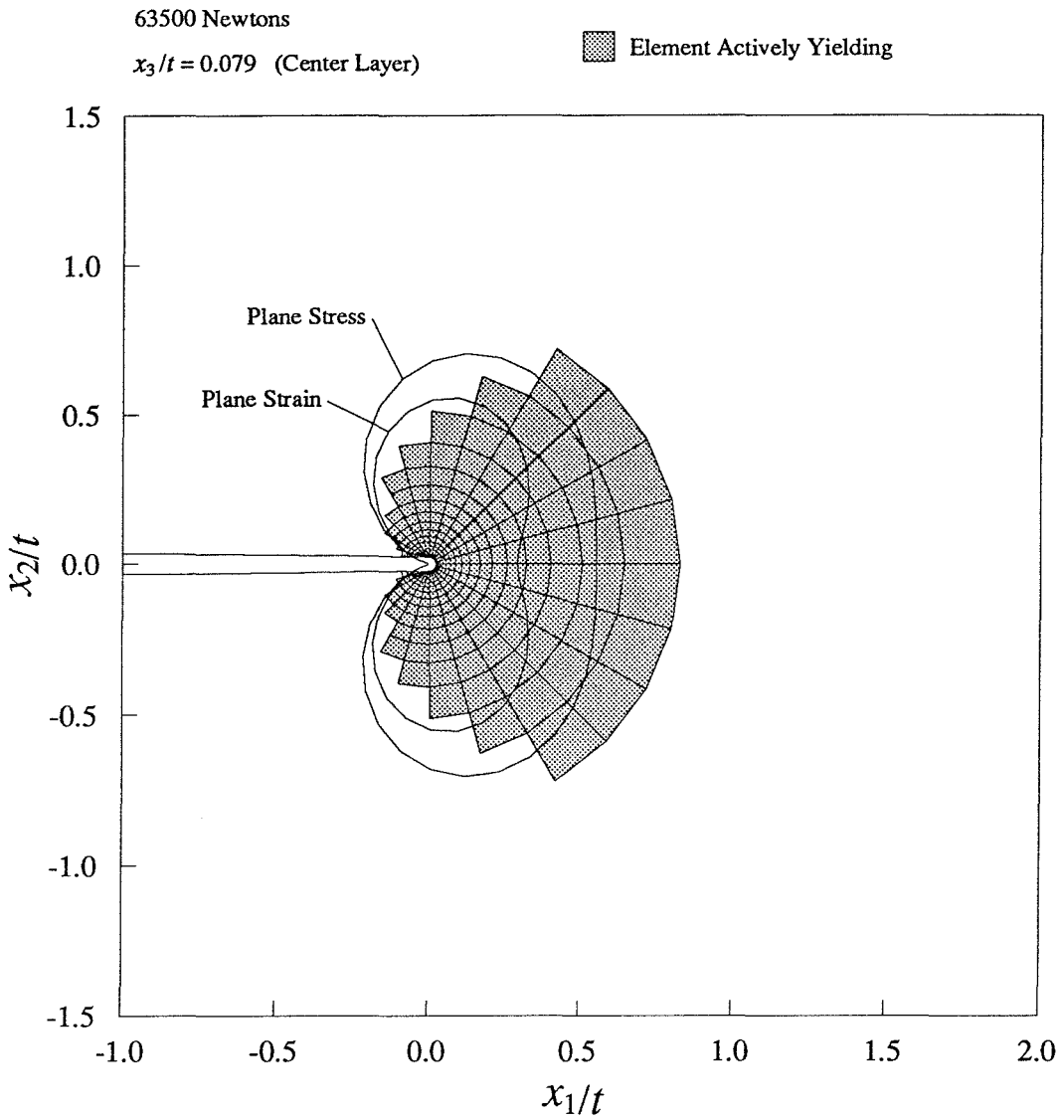


FIGURE 147. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

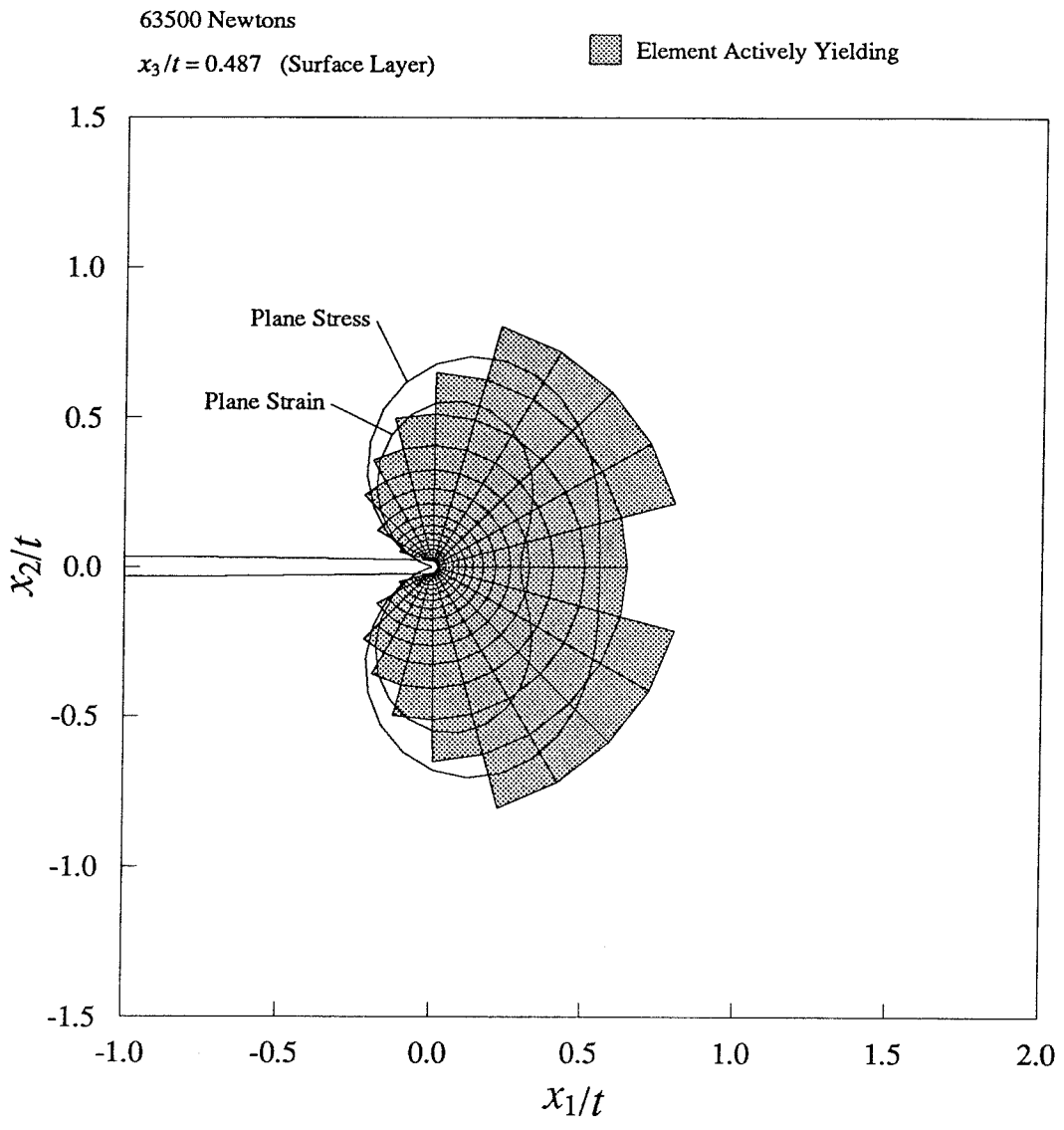


FIGURE 148. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

Figures 145 through 156 are intended to indicate the growth of the plastic zones at higher loads, as well as to indicate the size of the plastic zones with respect to the measurements shown in Section 5.2.

As expected, at the higher loads, the numerically estimated plastic zones are

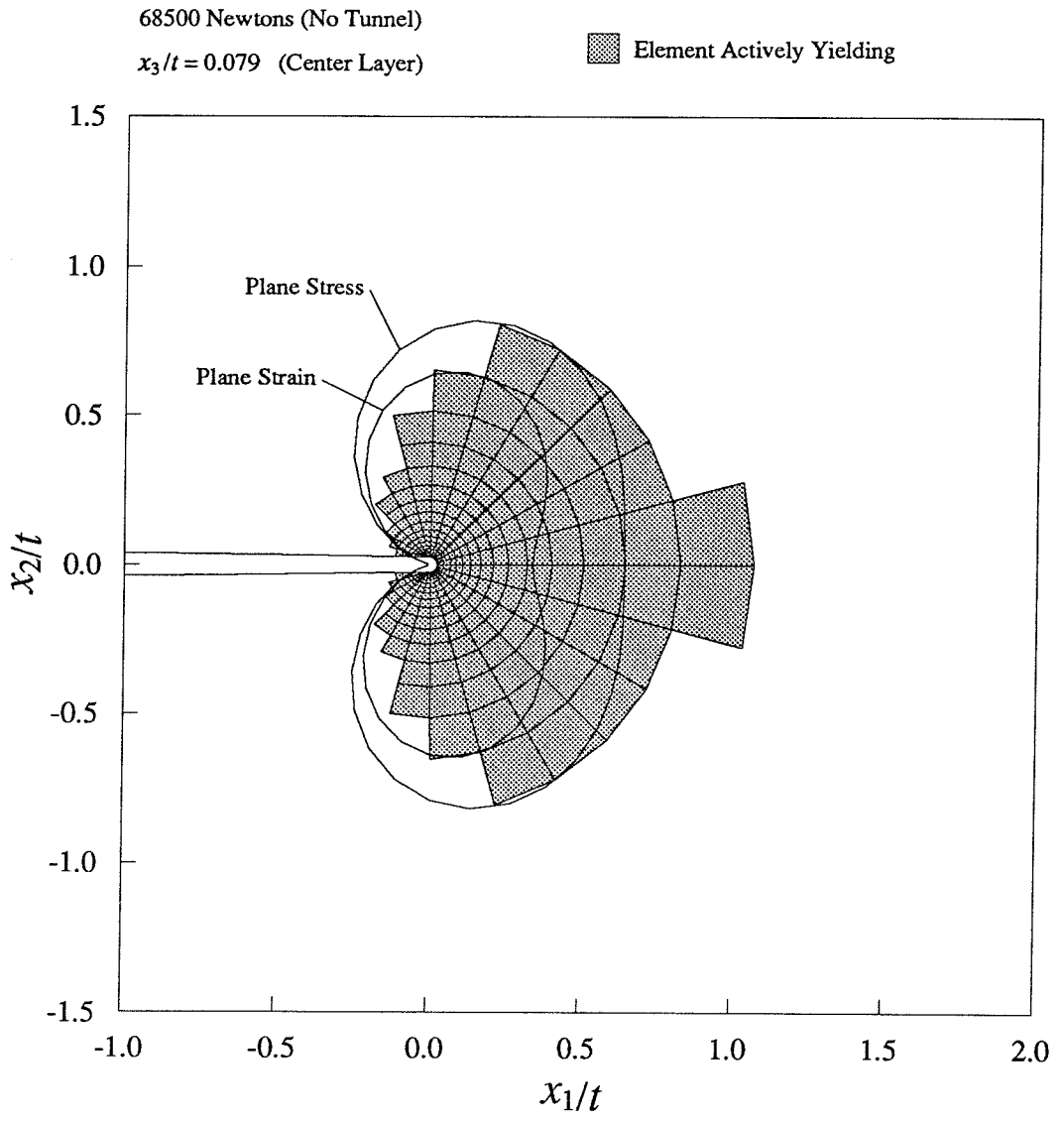


FIGURE 149. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

quite a bit larger than the analytical estimate along $\theta = 0$. However, the analytical estimate seems quite good at predicting the lateral extent of the plastic zone.

Aside from unloading in the region of the tunnel, and along the crack flank, the inclusion of the tunnel also results in some interesting regions of unloading in

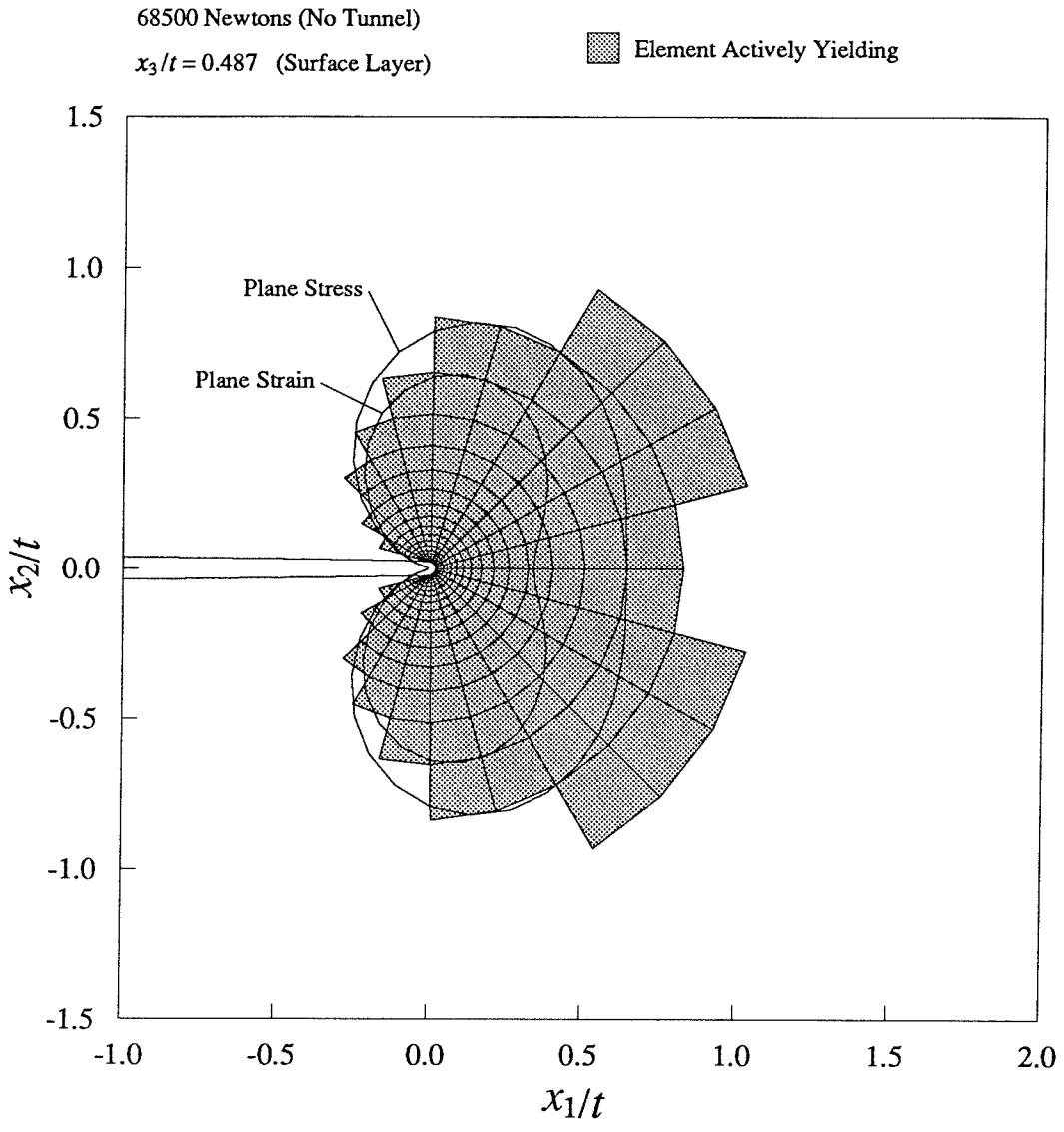


FIGURE 150. The plastic zone of the elements at the free surface of the specimen determined from the finite-element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

the layer of elements nearest the free surface. This zone of unloading occurs well ahead of the tip of the tunnel, as can be seen in Figures 152 and 156.

Figures 143 and 144 also show the plastic zones at the center layer of elements and the free surface layer of elements, in this case at a load of 18500 Newtons, about

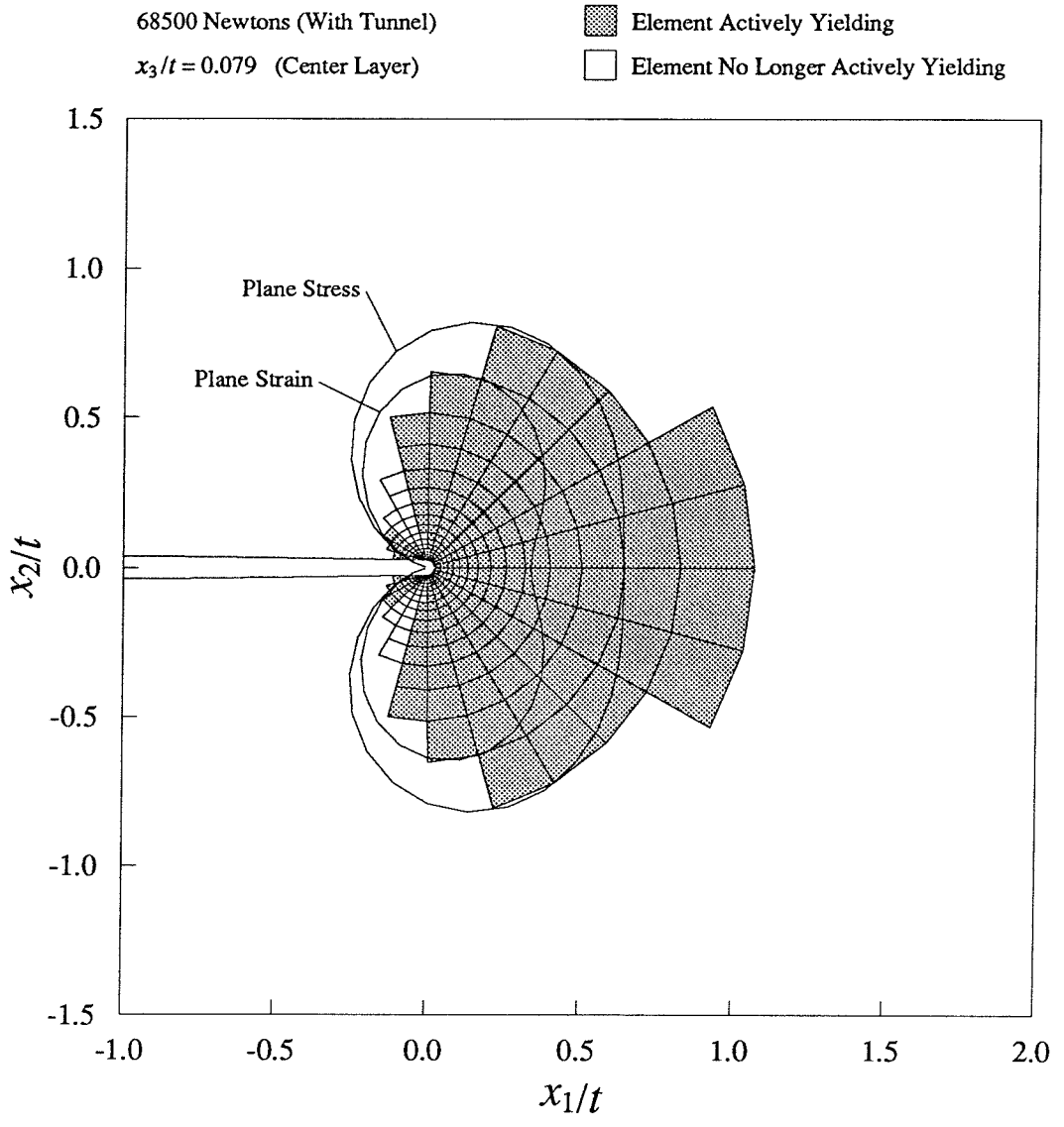


FIGURE 151. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

one quarter of the ultimate failure load.

Tunneling was determined to initiate after a load of 63500 Newtons, as described in Section 4.5. This tunneling was introduced into the finite–element model by releasing the displacement constraints along the crack line for a number of nodes,

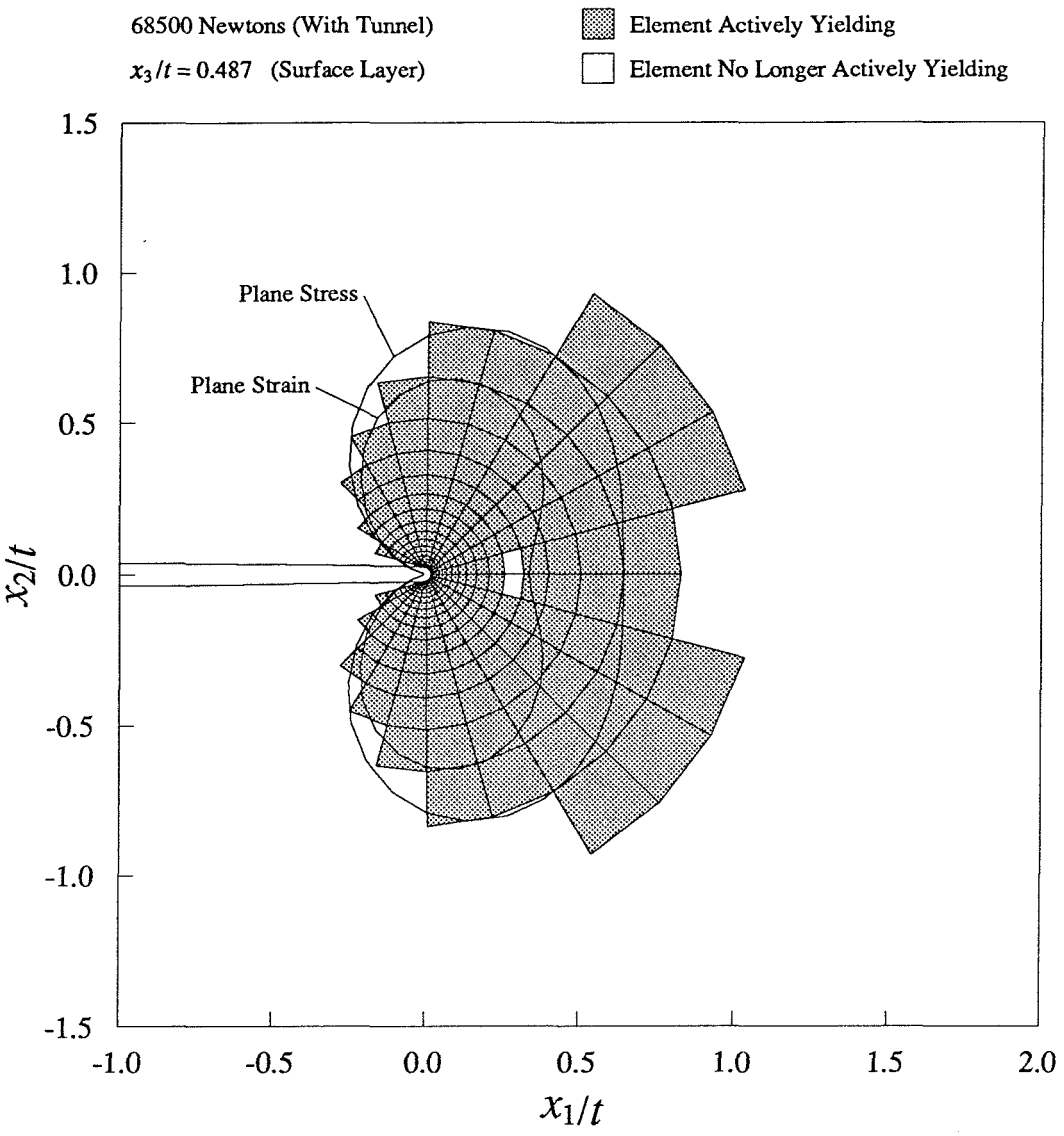


FIGURE 152. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

as discussed in Section 4.3. This nodal–release procedure was carried out in four load steps of 2500 Newtons each from 63500 Newtons to 73500 Newtons, which was the highest load achieved in the interferometric investigation. Plastic zones are shown for loads of 68500 Newtons and 73500 Newtons for two cases, one that

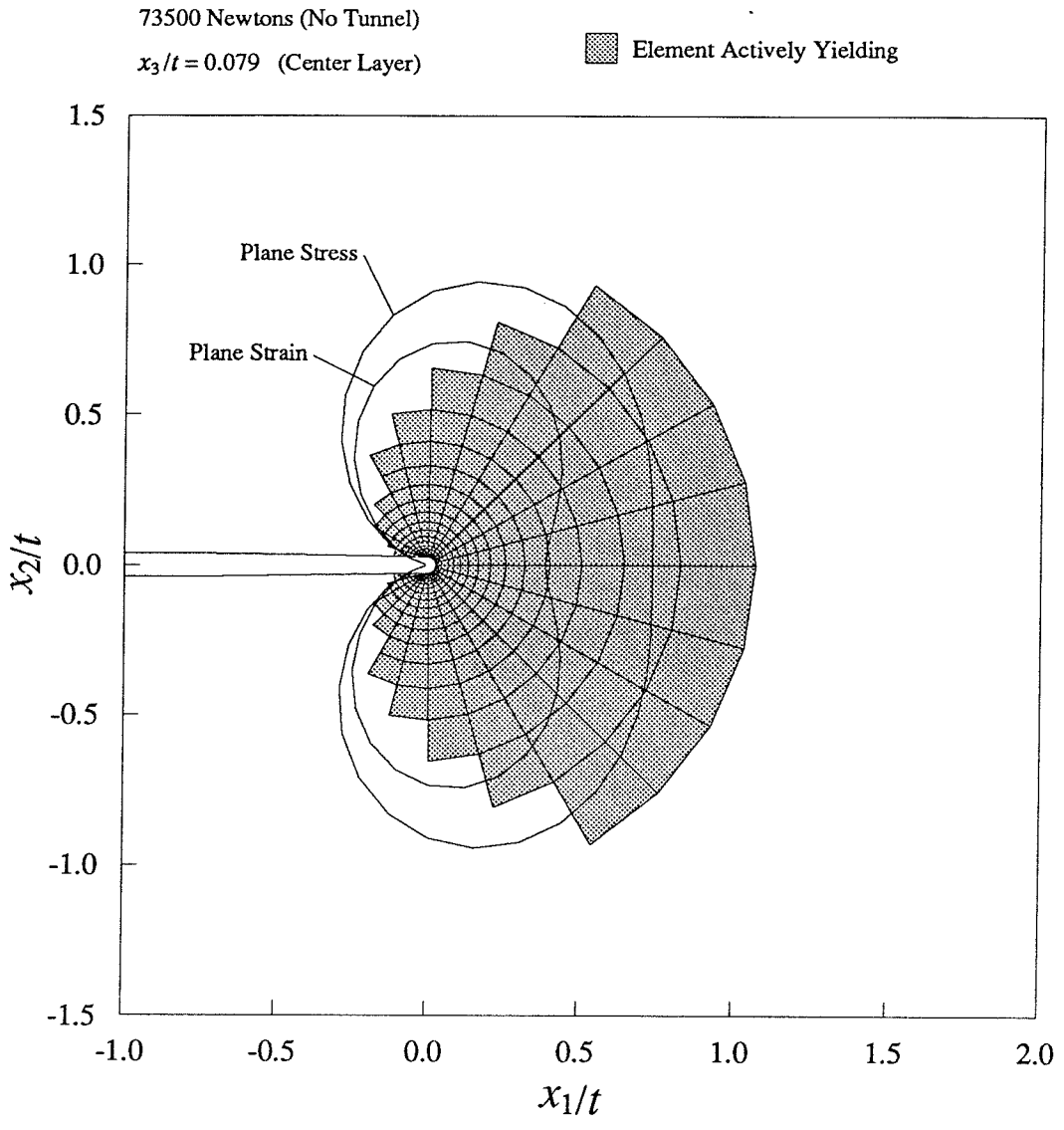


FIGURE 153. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

included tunneling and one that did not. The cases that included tunneling will be magnified in Figures 161 and 162 so that the regions of unloading might be seen more clearly.

Figures 151 and 152 show the plastic zones at 68500 Newtons for the case with

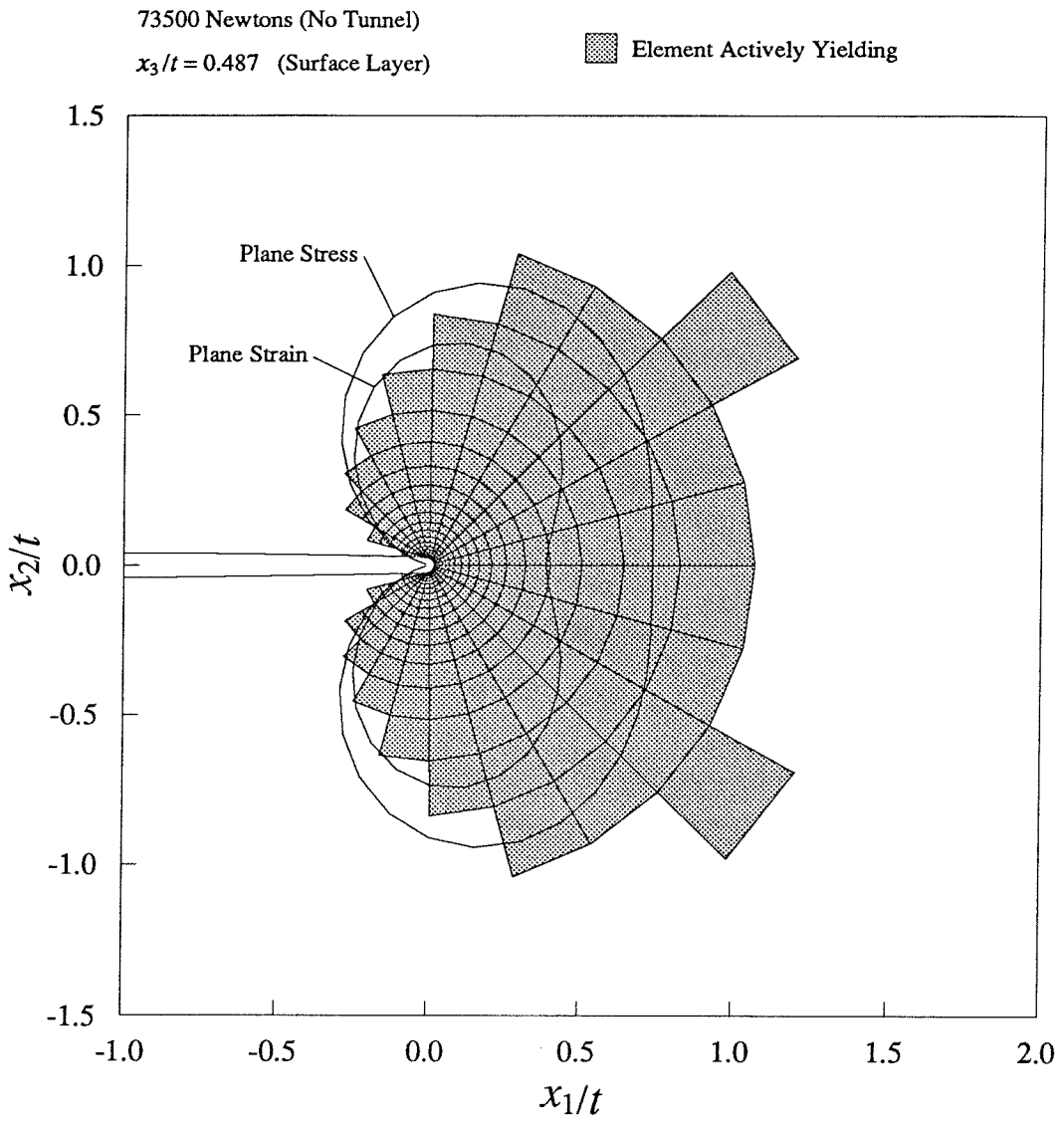


FIGURE 154. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

tunneling included. Some unloading is visible behind the notch tip in the center of the specimen, and a strange region of unloading is visible ahead of the notch tip on the free surface.

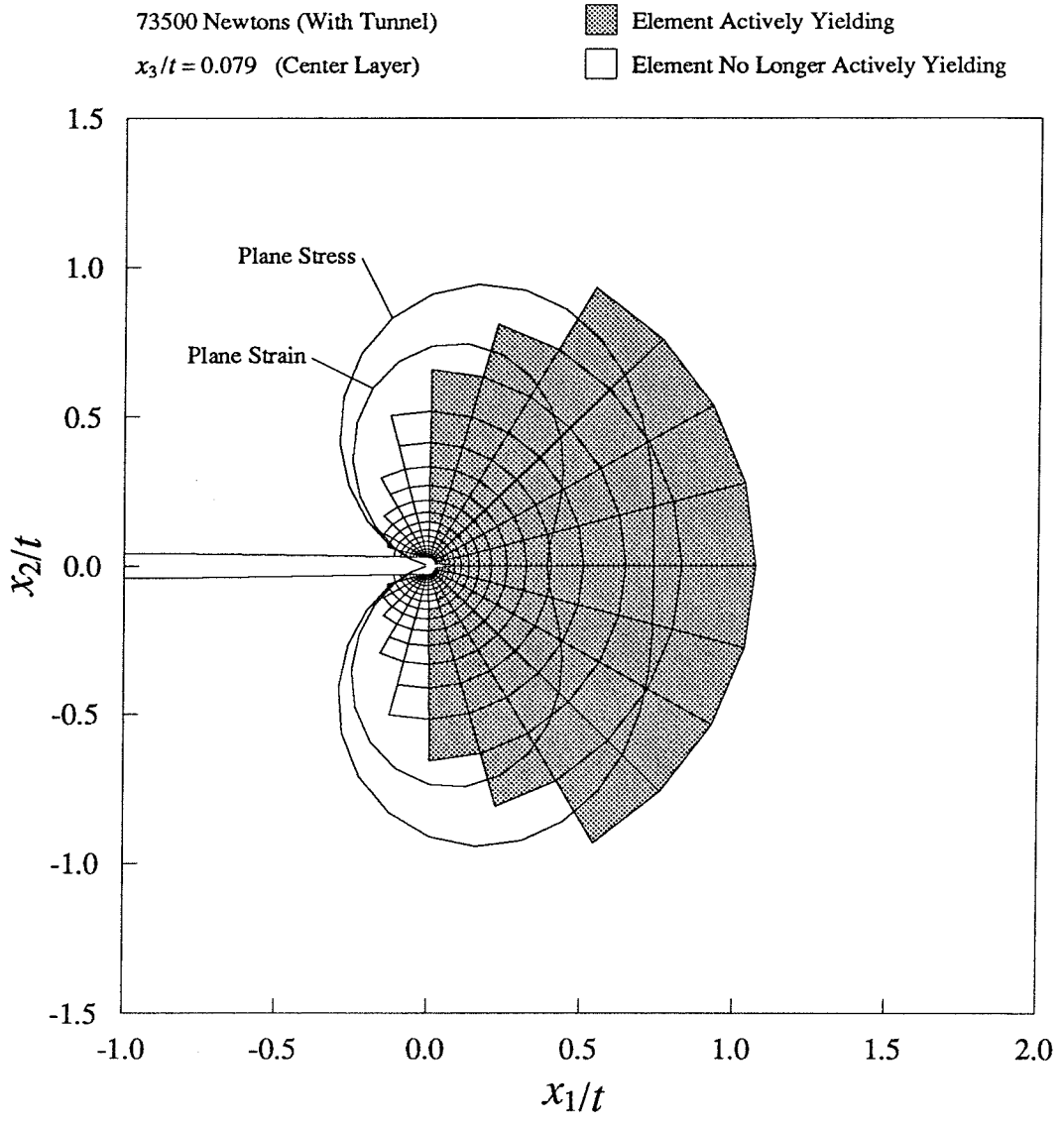


FIGURE 155. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

Figures 155 and 156 show the plastic zones at 73500 Newtons for the case with tunneling included. More unloading is visible behind the notch tip in the center of the specimen, and the region of unloading ahead of the notch tip at the free surface has grown.

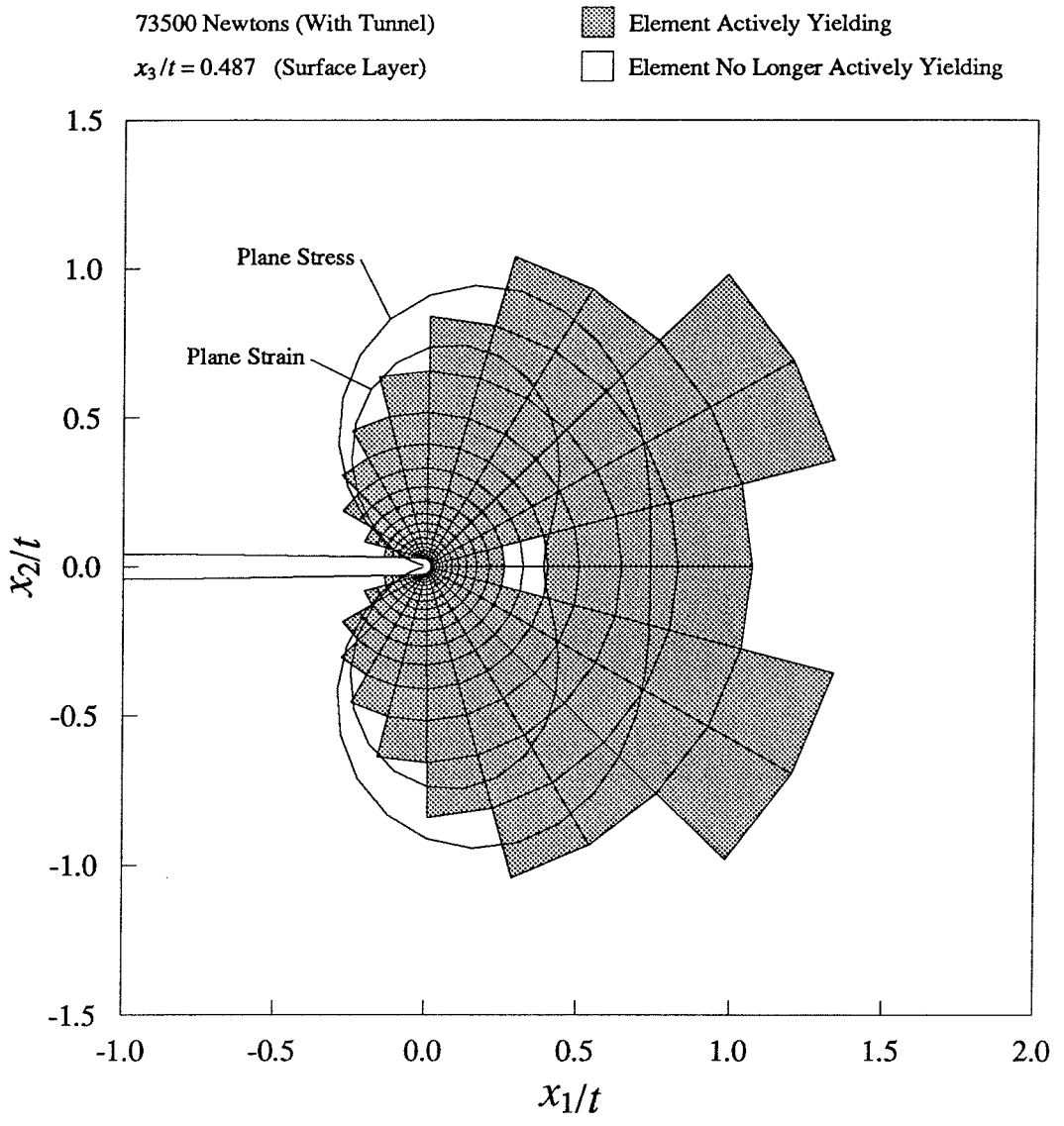


FIGURE 156. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is four times actual size. The plastic zone is shown on a deformed mesh. Also shown are the analytical estimates of equation (5.3.1).

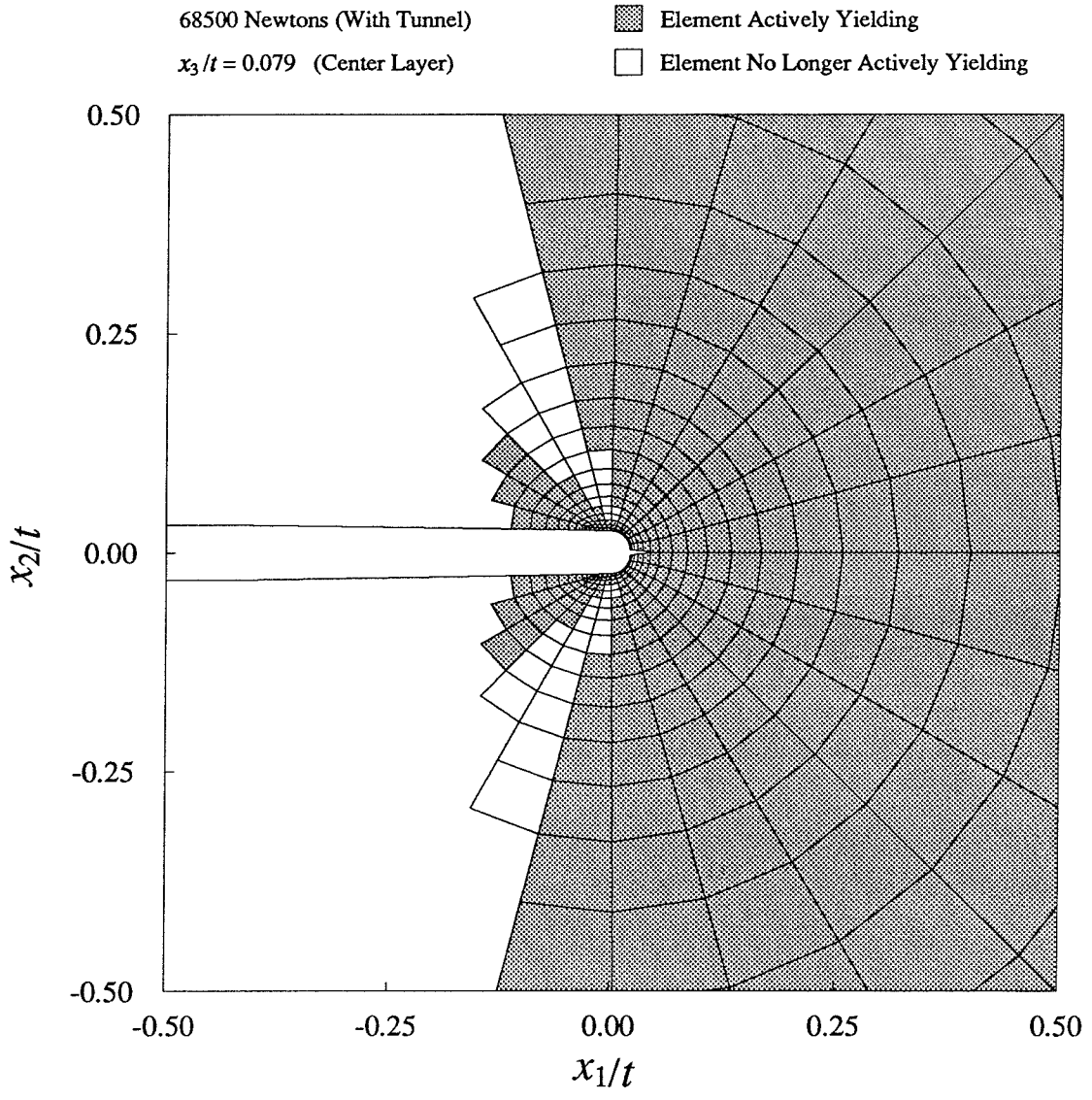


FIGURE 157. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is twelve times actual size. The plastic zone is shown on a deformed mesh.

In Figures 157 and 158 are shown magnified views of the notch tip regions at 68500 Newtons for the case where tunneling was included. As expected, there is unloading occurring behind the notch in the center of the specimen. The tunnel profile is a bell–shaped curve, deep in the center and dropping off towards the free surface, so that no nodes have been released at the free surface for this load. Thus,

there is no unloading behind the notch in the elements at the free surface. However, there is a region of unloading ahead of the notch at the free surface, which was not anticipated.

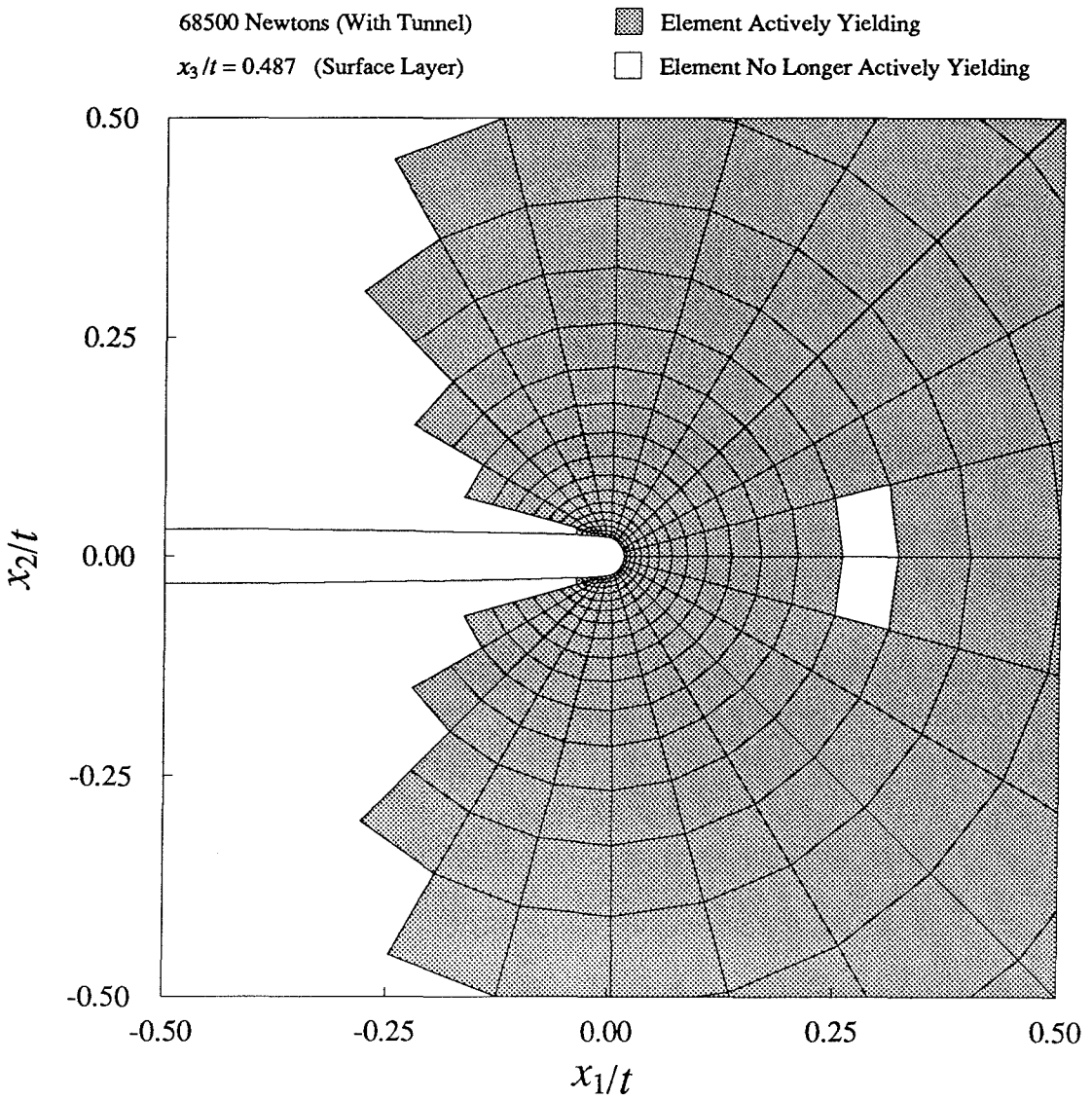


FIGURE 158. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is twelve times actual size. The plastic zone is shown on a deformed mesh.

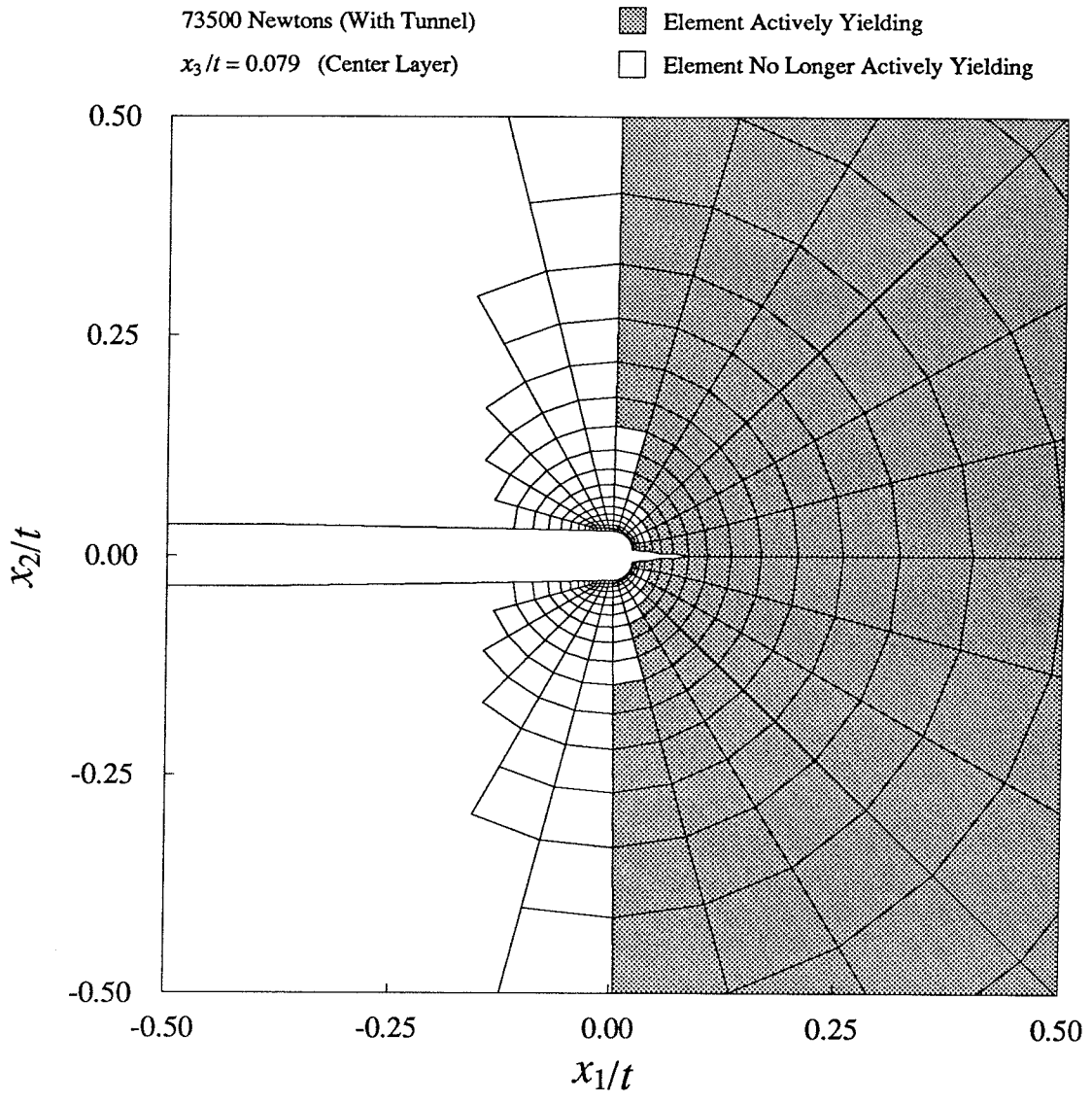


FIGURE 159. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is twelve times actual size. The plastic zone is shown on a deformed mesh.

In Figures 159 and 160 are shown similar magnified views of the notch tip regions at 73500 Newtons, also for the case with tunneling. The material behind the notch tip in the center of the specimen is completely unloaded. At this load, one node has been released at the free surface, but there is still no evidence of unloading behind the notch tip in the elements at the free surface. The region of unloading

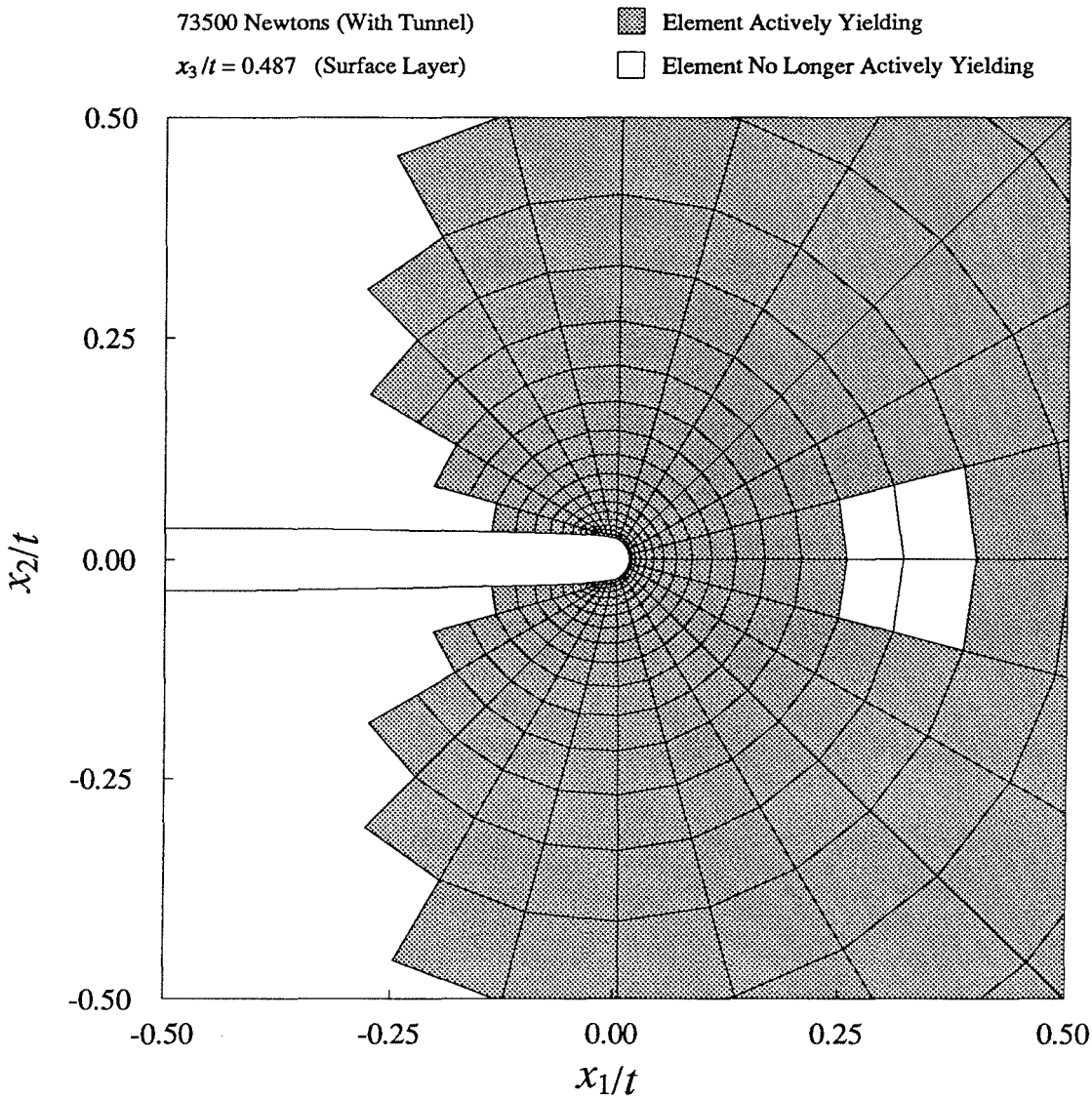


FIGURE 160. The plastic zone of the elements at the free surface of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is twelve times actual size. The plastic zone is shown on a deformed mesh.

ahead of the crack at the free surface has grown.

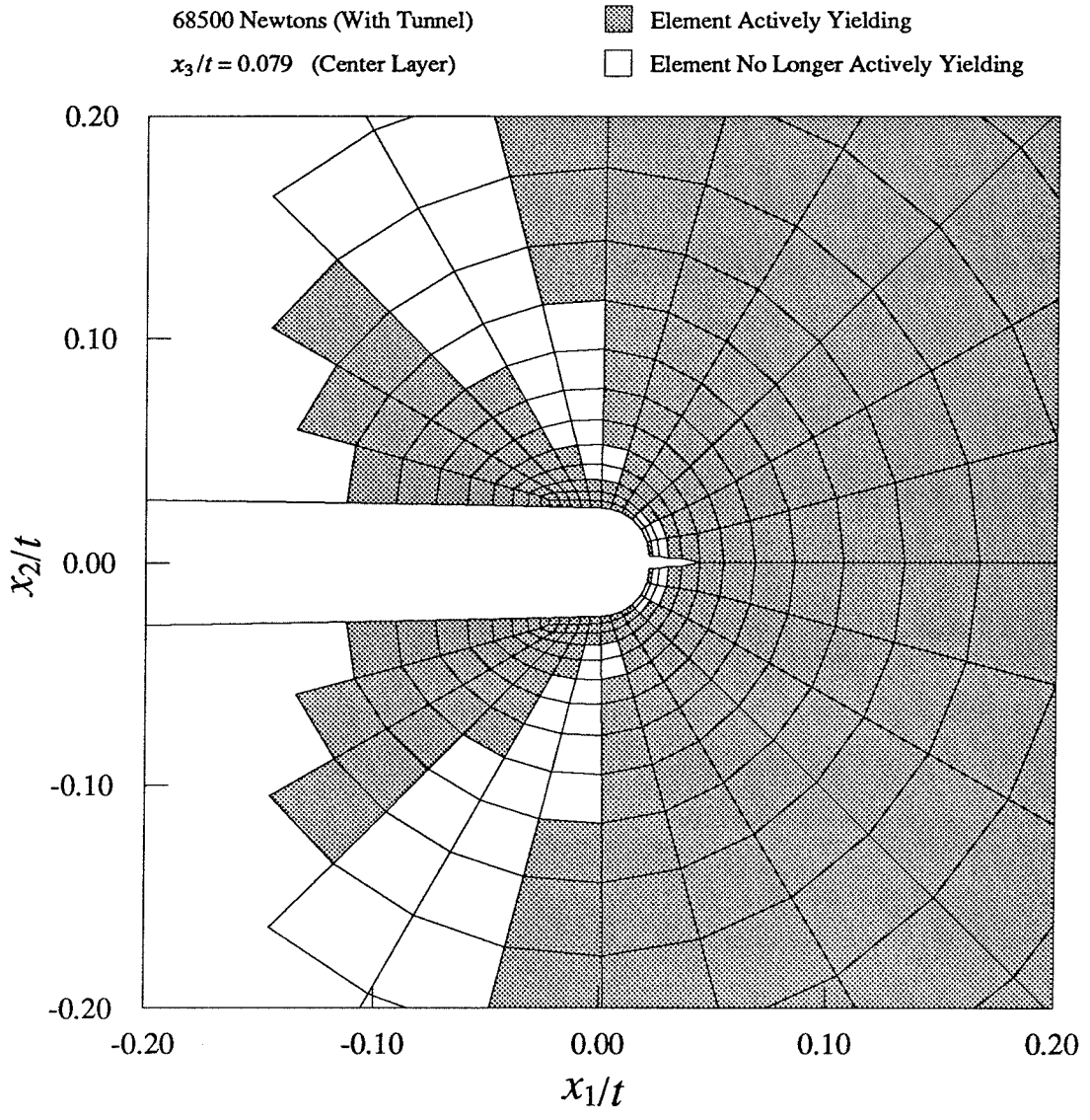


FIGURE 161. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is twelve times actual size. The plastic zone is shown on a deformed mesh.

Figures 161 and 162 are higher magnifications of the notch tip region in the center of the specimen at 68500 and 73500 Newtons, both including tunneling. The progress of the tunnel is clearly visible at this magnification. It is interesting that some of the material very close to the notch tip which was unloaded at 68500 Newtons is again actively yielding at 73500 Newtons.

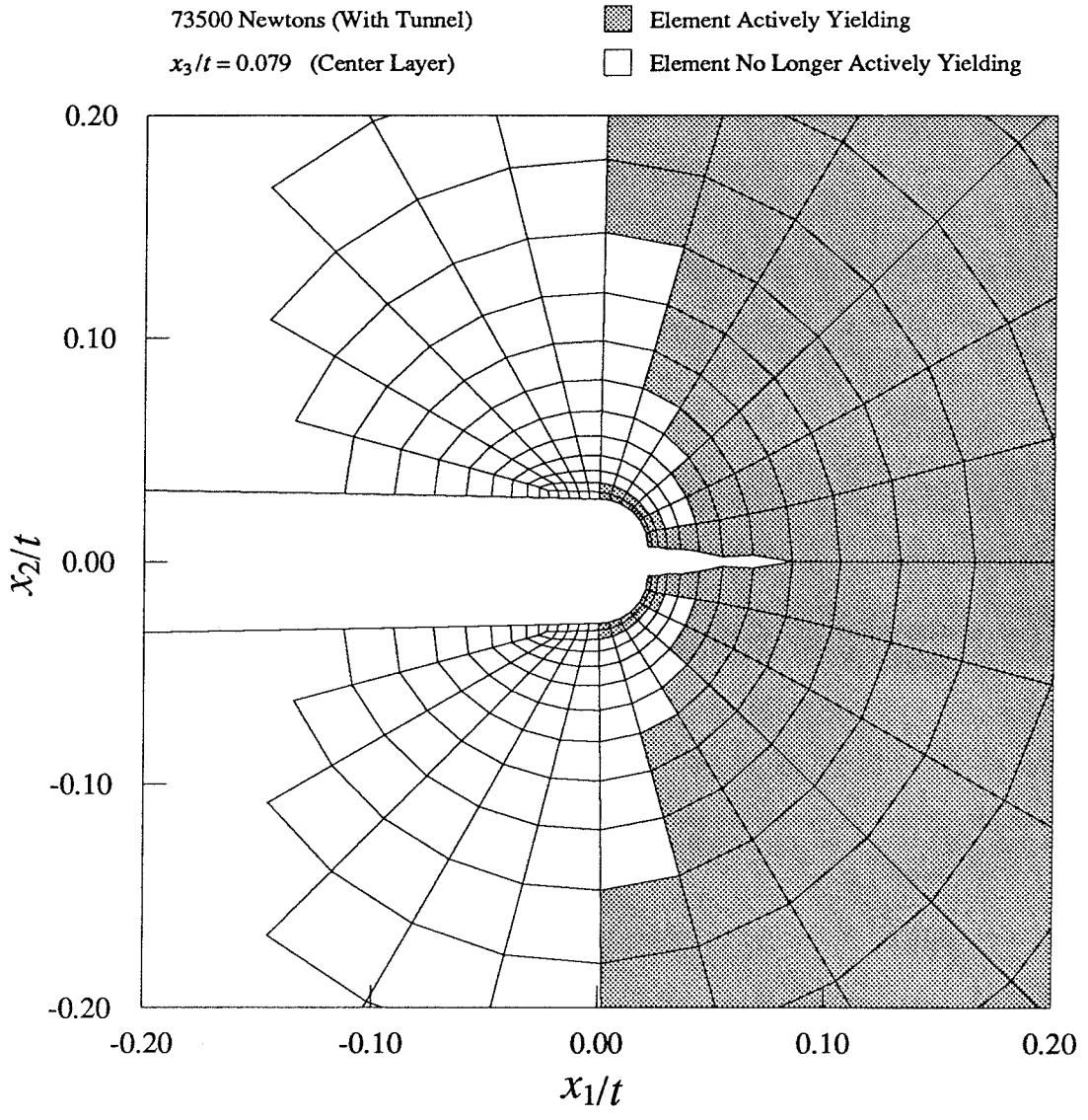


FIGURE 162. The plastic zone of the elements at the center of the specimen determined from the finite–element model. The specimen thickness is 1 cm, and the figure is twelve times actual size. The plastic zone is shown on a deformed mesh.

5.4 Stress Distribution

As seen in Section 4.2.4, the finite notch tip requires a correction to the K field stress distribution in order to accommodate the traction–free condition inside the notch tip itself. Along $\theta = 0$, σ_{11} must go to zero at the edge of the notch, so that the correction term is of opposite sign to the K field σ_{11} stress. This produces a local maximum in the σ_{11} stress distribution slightly ahead of the notch tip. However, the notch tip correction for σ_{22} is of the same sign as the K field σ_{22} stress, so that the σ_{22} stress distribution found using the K field with correction reaches a maximum at the notch tip itself.

A sample comparison between the K field with notch correction and the numerical model is shown in Figure 163. A larger sample of the radial and angular stress distributions from the finite–element code is shown with the corresponding K field with correction in SM 91-2 [73]. The numerical results are shown for loads beginning at 4.0 kN, the lowest load from the experiment; at this load the numerical model indicates that there is no plastic deformation. It can be seen that the agreement between the K field with correction and the numerical model is quite good.

With the onset of plastic deformation, however, there is a change in the stress distribution, particularly in the σ_{22} stress. The combination of the finite notch tip and the plastic flow produces a local maximum in the σ_{22} stress ahead of the notch tip, similar to the local maximum of the σ_{11} stress in the elastic case. Thus, σ_{22} is no longer a maximum at the tip of the notch. This change is illustrated in Figures 164 and 165, which show the radial distribution of σ_{22} along $\theta = 7.5°$ on $x_3/t = 0.079$ and $x_3/t = 0.487$. $\theta = 7.5°$ is the line through the centroids of the elements closest to the crack line of the specimen; $x_3/t = 0.079$ is the plane through the centroids of the elements on the layer nearest the center of the specimen, and $x_3/t = 0.487$ is the plane through the centroids of the elements nearest the free surface of the specimen. The stress is normalized by the yield stress, σ_0, which was measured as 960 MPa. The data are given as a function of r/t, where t is the specimen thickness, $t = 10$mm.

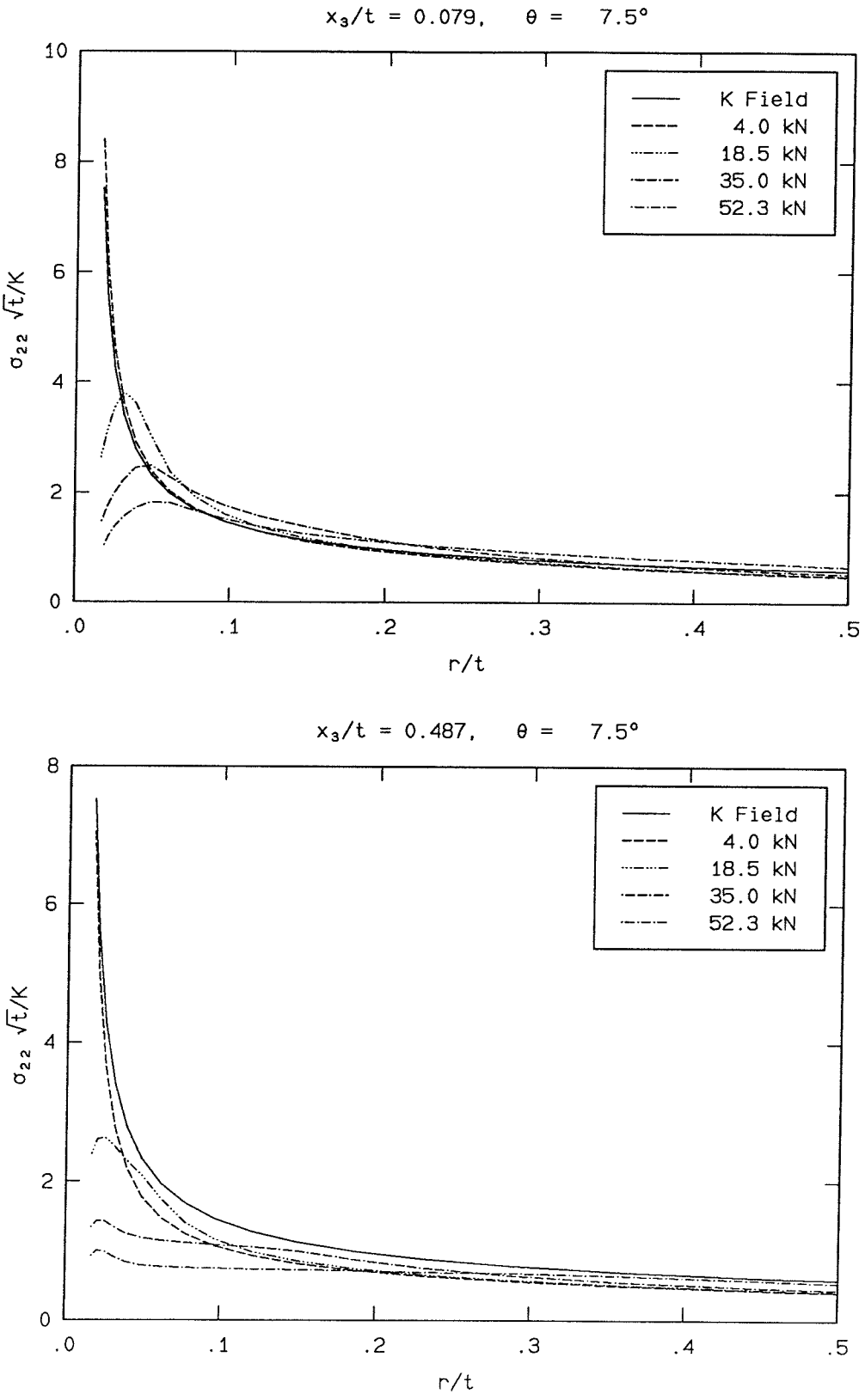


FIGURE 163. Radial variation of σ_{22}. K normalization.

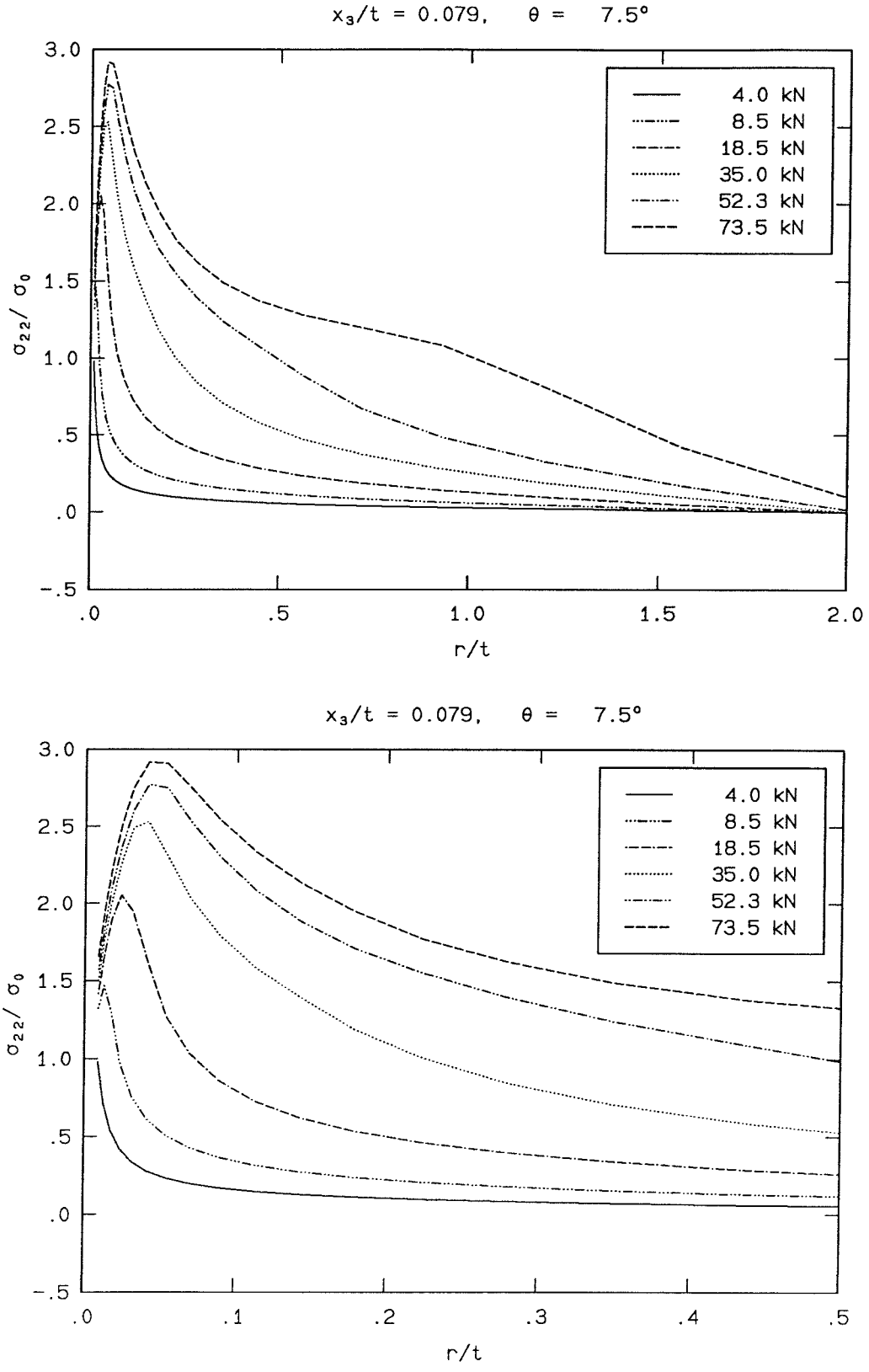


FIGURE 164. Radial variation of σ_{22}/σ_0. Center Layer.

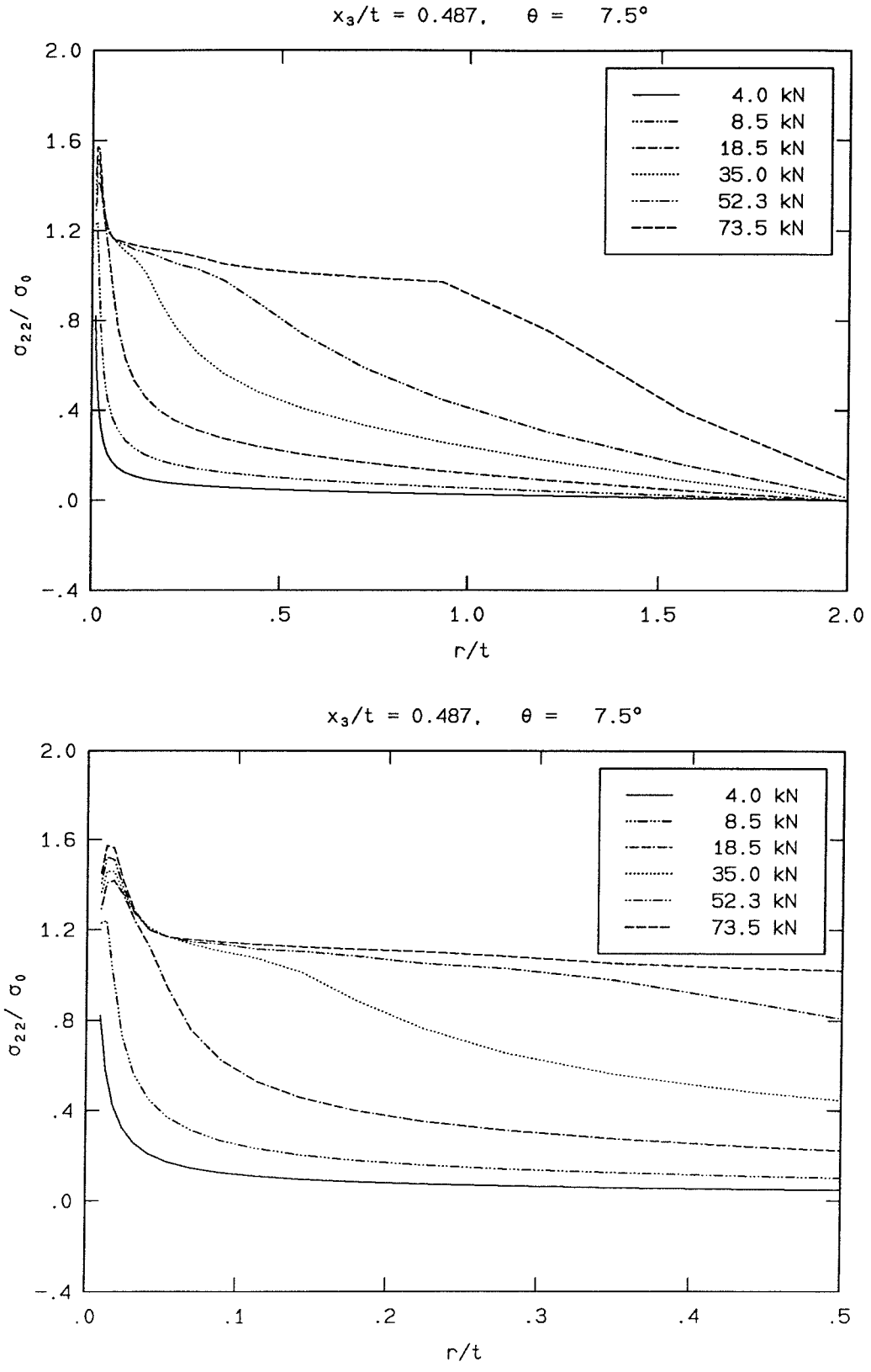


FIGURE 165. Radial variation of σ_{22}/σ_0. Surface Layer.

The local maximum in the σ_{22} distribution has already been described by Rice and Johnson [67], where large goemetry changes in plane–strain fracture were investigated using slip–line theory; and by McMeeking [45], where a finite deformation analysis of plane–strain crack–tip opening was investigated numerically. McMeeking attributes this local maximum ahead of the notch tip to the triaxiality of the deformation, and to the fact that the hydrostatic stress cannot be maintained at the surface of the notch tip. This effect is similar to that seen in the analysis of Bridgman [12,13] for necked tensile specimens, with the finite notch acting as a severely curved neck. Bridgman's analysis also indicated that the maximum tensile across the necked cross section occurs in the center of the specimen, away from the curved free surfaces that would lead to stress concentrations in an elastic medium. Rice and Johnson conclude that if the actual large geometry changes are included, the maximum stress that may be achieved over any reasonable material size scale is limited, contrary to the prediction of the singular HRR field.

The effect of the onset of plasticity in the σ_{11} distribution near the free surface is also interesting. The local maximum in the σ_{11} field is broadened and pushed farther ahead of the notch, and as the load increases, a second maximum is created in the field closer to the notch tip. This second maximum is very sharp, and on the order of one-half the magnitude of the maximum stress farther away from the notch tip. This distribution is shown in Figure 166, and the corresponding σ_{11} distribution in the center of the specimen is shown in Figure 167.

Similar radial hydrostatic stress distributions are displayed in Figures 168 and 169, and the radial variation of σ_{33} on the center layer of the specimen is shown in Figure 170 following.

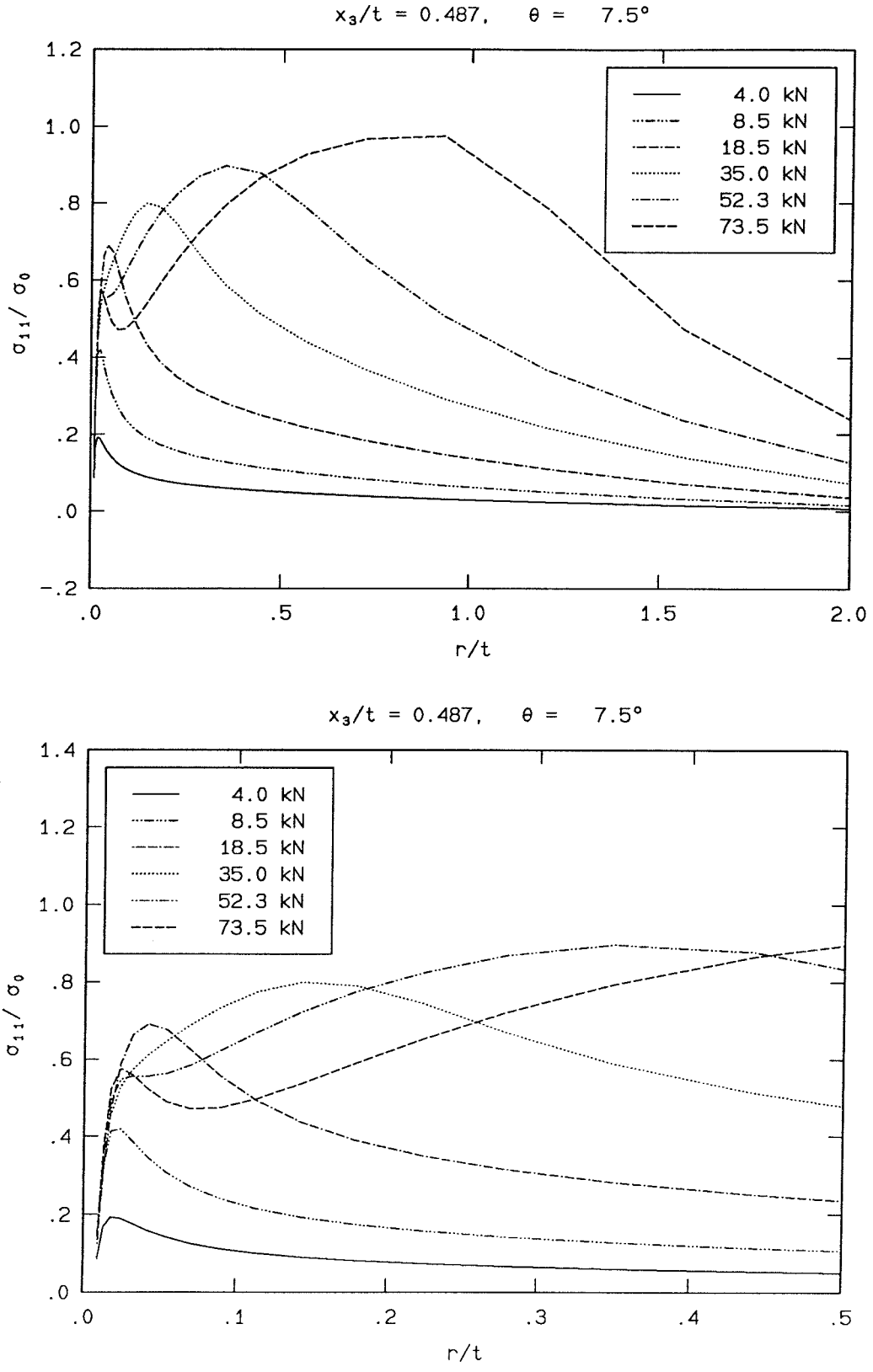


FIGURE 166. Radial variation of σ_{11}/σ_0. Surface Layer.

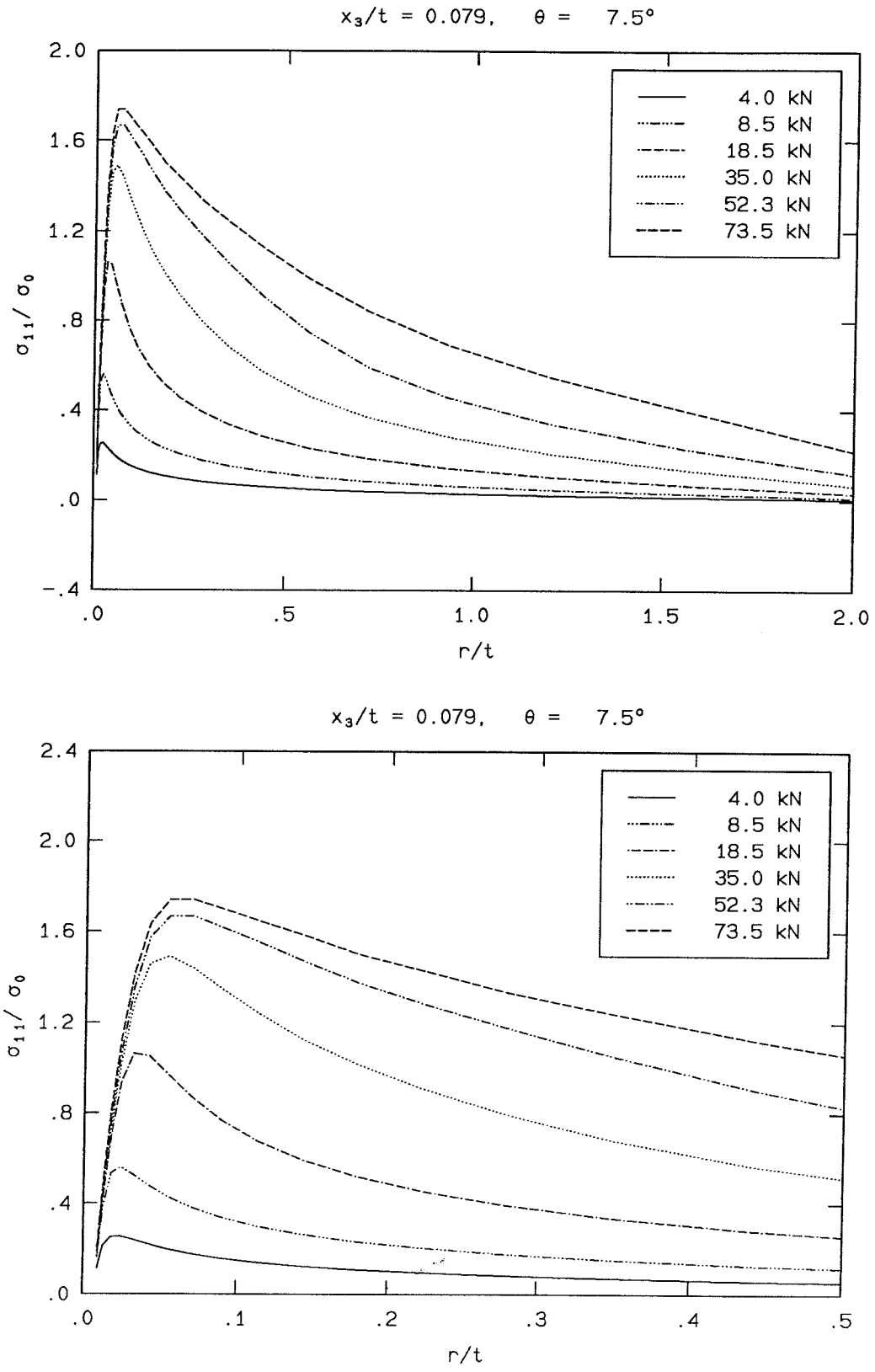


FIGURE 167. Radial variation of σ_{11}/σ_0. Center Layer.

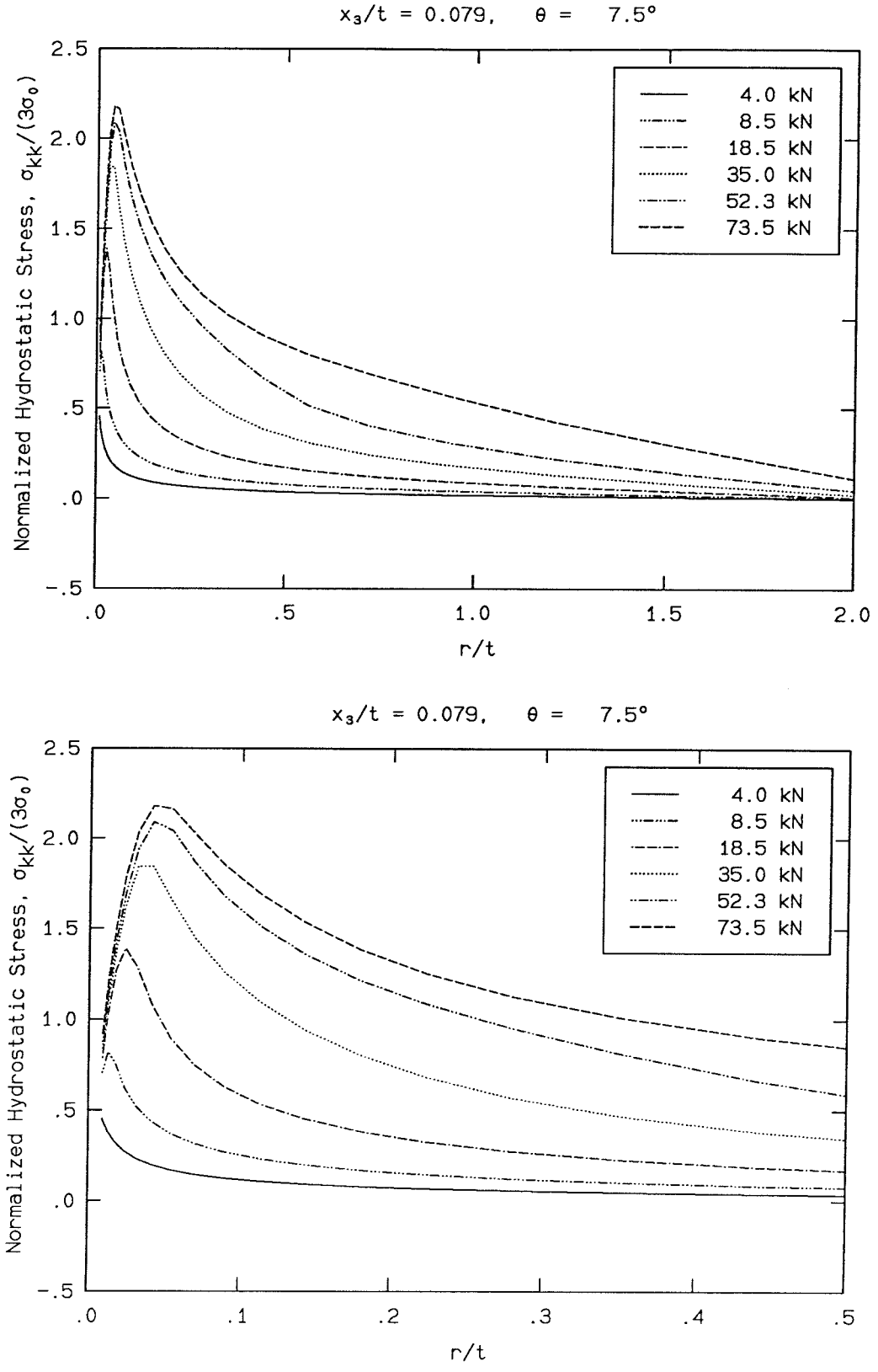


FIGURE 168. Radial variation of $\sigma_{kk}/(3\sigma_0)$. Center Layer.

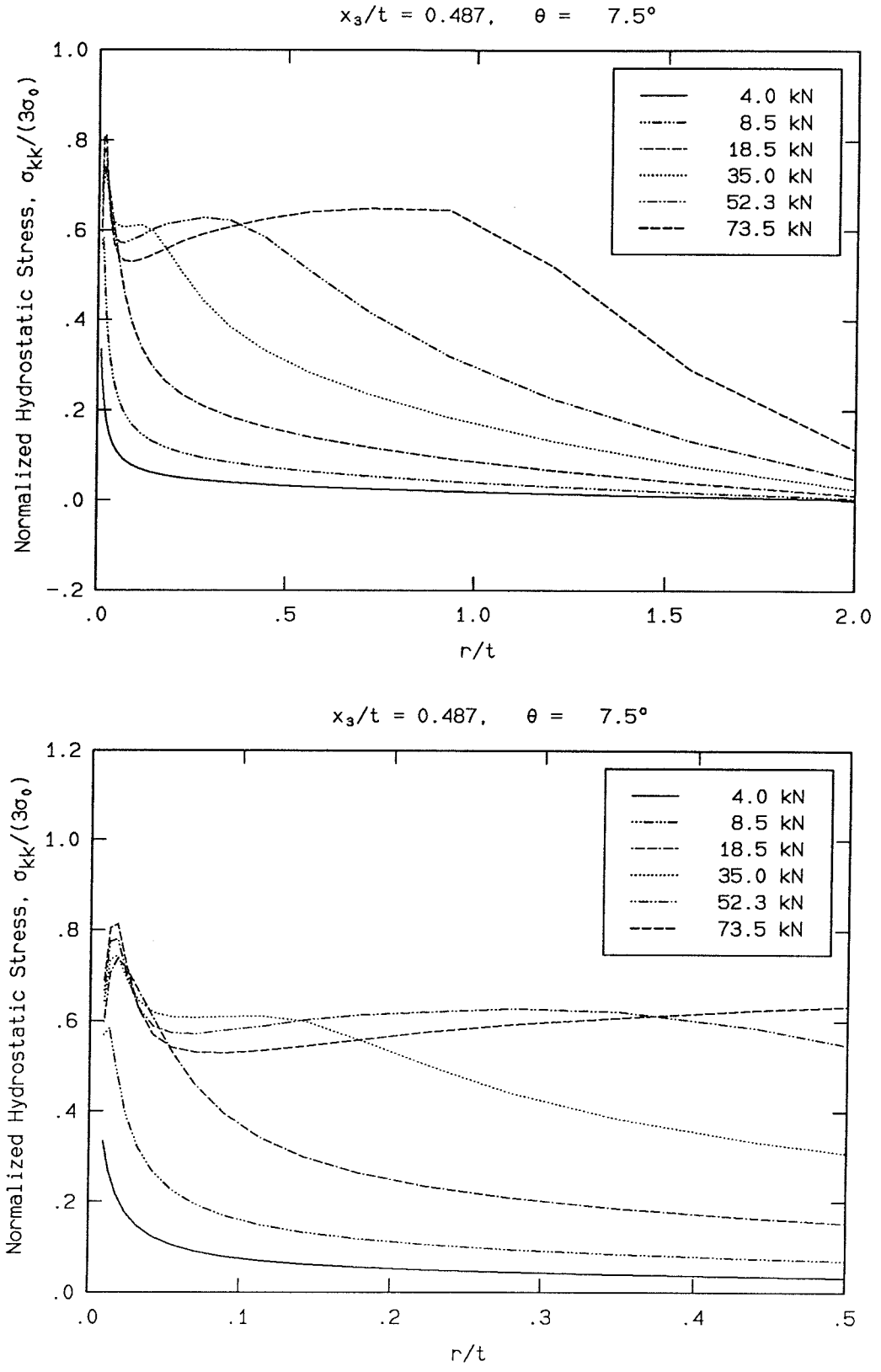


FIGURE 169. Radial variation of $\sigma_{kk}/(3\sigma_0)$. Surface Layer.

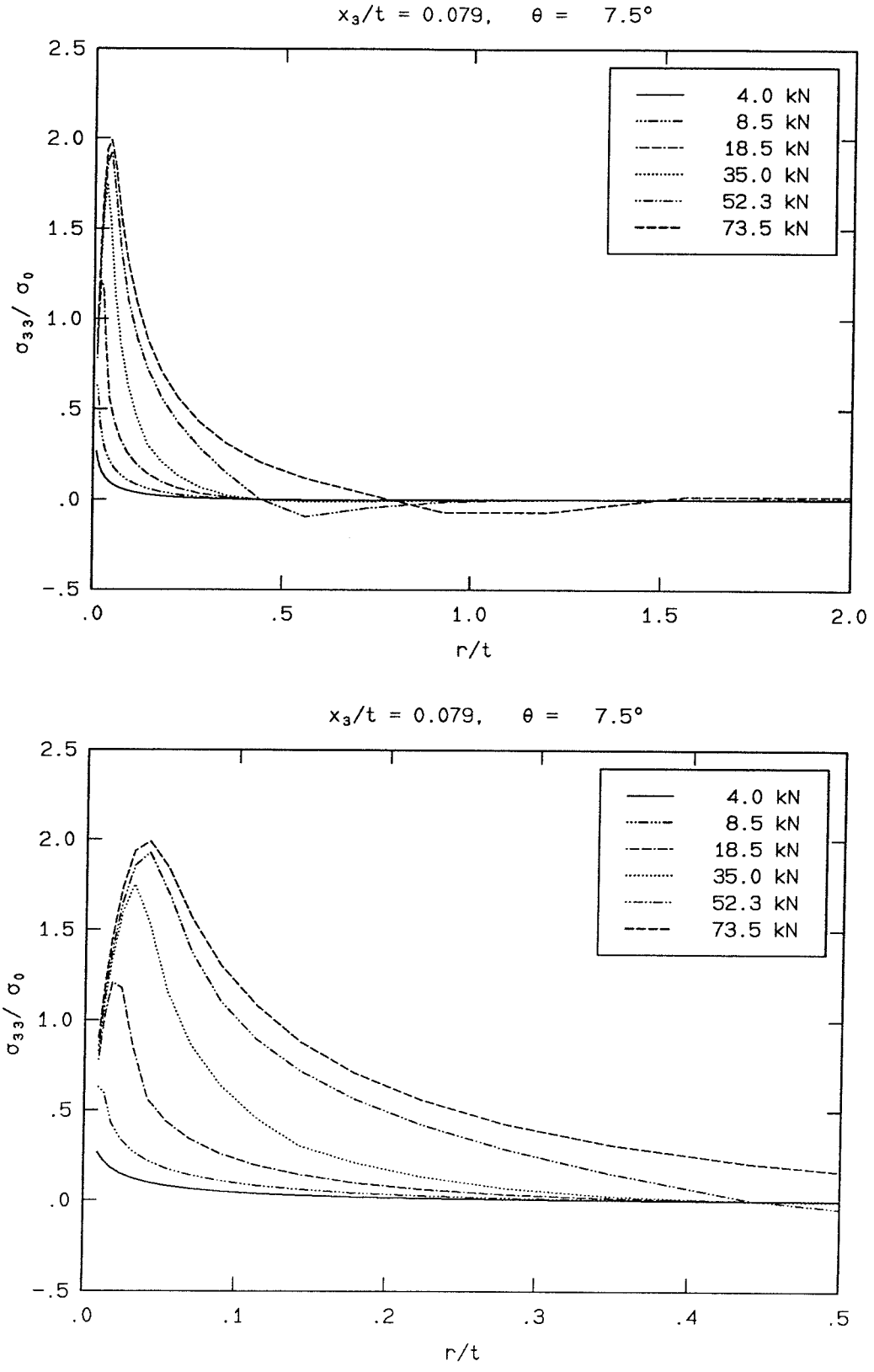


FIGURE 170. Radial variation of σ_{33}/σ_0. Center Layer.

A sample comparison between the HRR field and the numerical model is shown in Figure 171. A more complete comparison between the numerically calculated stresses and the stresses given by the HRR field analysis is made in SM 91-2 [73]. There seems to be almost no region in which the HRR field would provide a good estimate of the calculated stresses. The best agreement seems to be the plane–stress HRR σ_{22} along $\theta = 7.5°$ near the free surface at $x_3/t = 0.479$.

In assessing the applicability of the HRR field, it might have been guessed ahead of time that there would be at most a very small region in which the HRR field is dominant, since the specimen does not satisfy all of the assumptions made in the derivation of the HRR field. The specimen is three–dimensional, contains a notch of finite radius and employs an incremental plasticity constitutive formulation, whereas the HRR field assumes a two–dimensional body (in plane stress or plane strain), a sharp crack and a deformation plasticity law.

An objection might be made to using a specimen with an initial blunt notch, rather than a specimen with a sharper crack. However, the notch in this study is highly repeatable, while a sharper crack would have to be produced through fatigue, which is more difficult to control repeatably. Further, one could follow the rationale used by McMeeking [45] in his investigation of the blunting of a notch using a finite–strain numerical model. McMeeking begins his calculation with an initially blunt notch, but he concludes that if the loaded crack or notch eventually blunts open to a width several times its initial width, the crack or notch may be considered as having been initially sharp. The notch used in the present study meets this criterion, as its width under the highest load (neglecting tunneling) is on the order of two or three times its initial width (as calculated by the finite–element model). In other words, at some distance from the notch, the effect of the finite notch tip is equivalent to the effect of the blunting that would have occurred with a sharper crack.

The important point is that the combination of a finite notch tip and plasticity moves the peak in the stress field away from the edge of the notch tip, contrary to the HRR prediction of a singularity. Further, the work by Rice and Johnson [67] and McMeeking [45] implies that if the effects of finite deformation are included

with the plasticity, then even an initially sharp crack will evolve into a blunted notch, thereby moving the stress maximum away from the tip. The HRR field is intended to describe the deformation asymptotically in the region near the sharp tip; however, the effects of finite deformation conspire to change the geometry of the crack tip so that the assumptions underlying the HRR field are no longer valid.

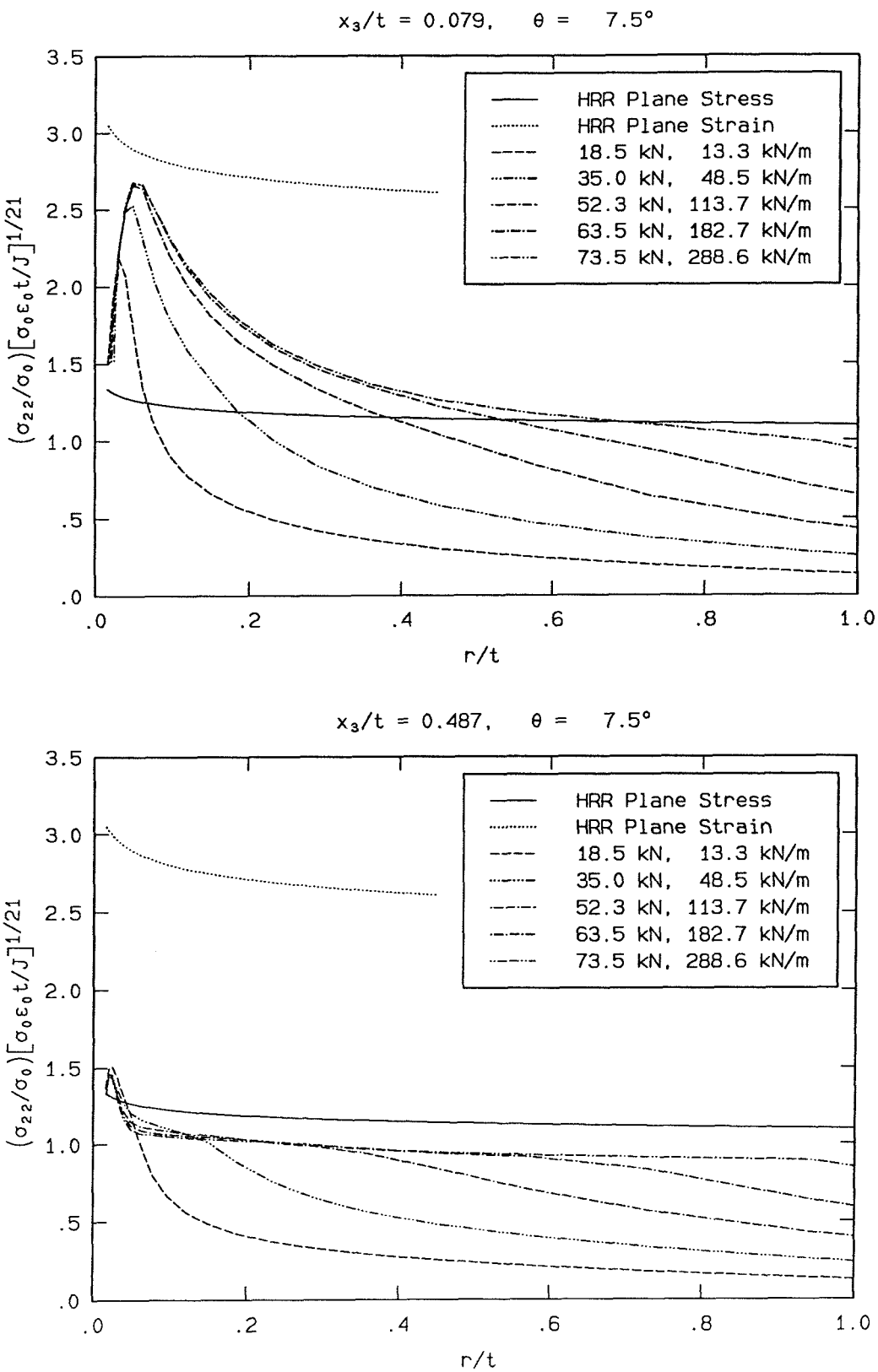


FIGURE 171. Radial variation of σ_{22}. J normalization.

Shih and German [79] assess the size scale over which the J-integral may be used as a criterion for fracture initiation. This assessment is made through a comparison between numerical calculations for several specimen geometries and the HRR field. The numerical calculations employ a plane strain, small strain, J_2 deformation–theory plasticity law, in which the plastic stress and strain are related through a power law. They assume that the J-integral will characterize the deformation at the crack tip and the onset of crack propagation if the crack–tip field is governed by the HRR field. Shih and German [79] estimate the size scale over which the J-integral characterizes the crack–tip deformation in three specimen geometries, a cracked bend bar, a center–cracked panel and a single–edge cracked panel. They employ this estimate to determine the necessary physical dimensions for each specimen geometry so that the fracture process zone is fully encompassed within the zone characterized by J.

For each of the three geometries, Shih and German [79] estimate the length scale over which the J-integral characterizes the deformation as a function of the value of J (which may be calculated from the work done by the applied loading), the initial yield stress of the material, σ_0, and the hardening behavior of the material determined by the hardening exponent, n. This length scale also depends on whether the plasticity may be described as contained or not; the plasticity is contained if the uncracked ligament is large compared to the parameter J/σ_0, which has units of length.

The case they consider that is closest to the present study is the cracked bend bar with $n = 10$, $\sigma_0/\mathrm{E} = 0.002$ and a crack that is one half the width of the bar. For contained plasticity, in which the uncracked ligament c is given as $c = 1200J/\sigma_0$, they estimate that the numerical stresses match the HRR field over a length X of $X < 3J/\sigma_0$. They also estimate that for the cracked bend bar under large plasticity, with $c = 120J/\sigma_0$, the numerical and HRR stresses match within $X < 3J/\sigma_0$; the match for the other geometries under large plasticity is considerably worse.

While it seems clear that the finite notch tip in the specimen of the present study negates the singularity at the notch tip, making the assumptions underlying

the HRR field invalid, it also appears that the stress and strain at the notch tip may be characterized by the same function of J used to characterize the HRR singular fields. This is demonstrated in Figures 171, 179and 180. The estimate of Shih and German of J dominance over a length of $3J/\sigma_0$ may be compared with the notch tip radius and the region over which the stresses and strains seem to be characterized by J for this geometry with a finite notch.

In the current study, the uncracked ligament is 46 mm, and the yield stress is σ_0 = 960 MPa, with σ_0/E = 0.005. For $c = 1200J/\sigma_0$, the case contained plasticity, the corresponding value of J is 36.8 kN/m, which corresponds to a load of approximately 30kN. For large scale plasticity, with $c = 120J/\sigma_0$, the corresponding value of J is 368 kN/m, which corresponds to a load of approximately 80 kN. (This load is beyond the failure load of the experimental specimen.) Thus, the length scale over which Shih and German estimate the J–integral to characterize the deformation is on the order of 0.1 mm for the case of contained plasticity and 1 mm for the case of large scale plasticity. Note that the initial notch–tip radius is 0.15 mm.

The length scale over which the HRR normalization collapses the data from the current study can be estimated from Figure 171. The length chosen for this estimate will be the distance from the tip of the notch to the location of the peak in the σ_{22} distribution in the center of the specimen, which is shown in the upper plot in Figure 171. The center layer of elements should more nearly match the plane–strain assumption of Shih and German [79]. This distance will be denoted l_c. This data is shown in Table 6.

This particular choice of a length over which the J–integral may be said to dominate the deformation does not match well with the estimate of $3J/\sigma_0$. It is difficult to tell from this limited set of data whether there is some direct correlation between l_c and J or the applied load.

At the free surface, the J–integral seems to characterize σ_{22} over a larger length scale, as indicated in the lower plot in Figure 171. For this case, the best length scale for comparison seems to be the position at which the normalized σ_{22} curve begins to drop away from the plateau near unity, which will be denoted l_s. These

Extent of J Characterization of σ_{22} Stress (center layer)				
Load kN	J kN/m	J/σ_0 mm	l_c mm	$3J/\sigma_0$ mm
18.5	13.3	0.0139	0.125	0.0417
35.0	48.5	0.0505	0.292	0.152
52.3	113.7	0.118	0.375	0.354
63.5	182.7	0.190	0.375	0.570
73.5	288.6	0.301	0.375	0.903

Table 6. The distance from the notch edge to the peak in the σ_{22} stress in the center layer of elements, l_c, for various loads and J values. Also shown is the estimate of the length over which the J–integral should dominate from Reference [79], $3J/\sigma_0$.

valuesare listed in Table 7.

If this length parameter is a good choice for comparison, it would seem to indicate that the J–integral characterizes the stress field over a length that is on the order of $30J/\sigma_0$ in the elements at the free surface.

It is very difficult to make a comparison that exactly parallels the comparison made by Shih and German between their numerical calculation and the HRR field, as the finite notch tip radius removes the singularity, which renders the assumptions used in deriving the HRR field invalid. There is as yet no analytical solution or correction for the HRR field that accounts for the effects of the finite notch. Also, it is clear that there will be additional difficulties introduced by the finite thickness of the specimen. Some related work done by Pfaff [56] indicates that the out-of-plane displacement field at the surface scales reasonably well with J in a region between 20 and 60 kN. Note that the out-of-plane displacements at the surface are related to the thickness average of the trace of the stress tensor. Pfaff explains the upper and lower bounds on the loads at which the J–integral will characterize the deformation as being caused by the thickness of the specimen. At lower loads,

Extent of J Characterization of σ_{22} Stress (surface layer)				
Load kN	J kN/m	J/σ_0 mm	l_s mm	$3J/\sigma_0$ mm
18.5	13.3	0.0139	0.339	0.0417
35.0	48.5	0.0505	1.19	0.152
52.3	113.7	0.118	2.88	0.354
63.5	182.7	0.190	6.53	0.570
73.5	288.6	0.301	9.15	0.903

Table 7. The distance from the notch edge to the fall–off in the σ_{22} stress in the surface layer of elements, l_s, for various loads and J values. Also shown is the estimate of the length over which the J–integral should dominate from Reference [79], $3J/\sigma_0$.

the surface deformation will resemble a semi–infinite body, and the plasticity will be dominated by the shape of the notch itself. At higher loads, the plasticity will smooth out the local effects of the notch shape, and the surface deformation will still approximate a semi–infinite body. In this loading regime, the J–integral will be a good choice for characterizing the deformation. At the highest loads, when the plastic zone is on the order of the thickness of the plate, the effects of the finite thickness will dominate the nature of the deformation, and the J–integral may no longer be an effective characterizing parameter.

5.5 Strain Levels

Although the numerical model used in this study allowed for the inclusion of nonlinearity in the constitutive description, it did not allow for any geometrical nonlinearity, *i.e.*, finite strains. It is therefore important to determine whether the calculated strain values remain within the region where the linearized strain formulation is valid. As can be seen in Figure 172, the strain levels above 52.3 kN are on the order of 30% to 90% in the region near the notch tip in the layer near the center of the specimen, which is certainly outside the linearized strain theory. The inclusion of tunneling does decrease the strain level when compared to the case that did not include tunneling, but the addition of the tunnel seemed to create difficulty for the code to converge to a smooth solution in the stress distribution, as discussed in Section 4.3.

Figure 172 shows the radial distribution of ϵ_{11} and ϵ_{22} along $\theta = 7.5°$ on $x_3/t = 0.079$, the midplane of the layer of elements in the center of the specimen. Note that ϵ_{11} and ϵ_{22} are almost equal and opposite (and ϵ_{33} is very nearly zero) near the notch tip, following the assumption of plastic incompressibility used in the constitutive model.

Figure 173 shows the radial distribution of ϵ_{11} and ϵ_{22} along $\theta = 7.5°$ on $x_3/t = 0.487$, the midplane of the layer of elements nearest the free surface of the specimen. In this case, ϵ_{11} is approximately two thirds as large as ϵ_{22}, while ϵ_{33} is approximately one third as large as ϵ_{22}, again demonstrating plastic incompressibility.

The magnitude of the equivalent plastic strain, $\bar{\epsilon}^p$, the scalar quantity that serves to record the history of the plastic deformation, is shown in Figure 174. Again, these data are shown along $\theta = 7.5°$ for $x_3/t = 0.079$ and $x_3/t = 0.487$.

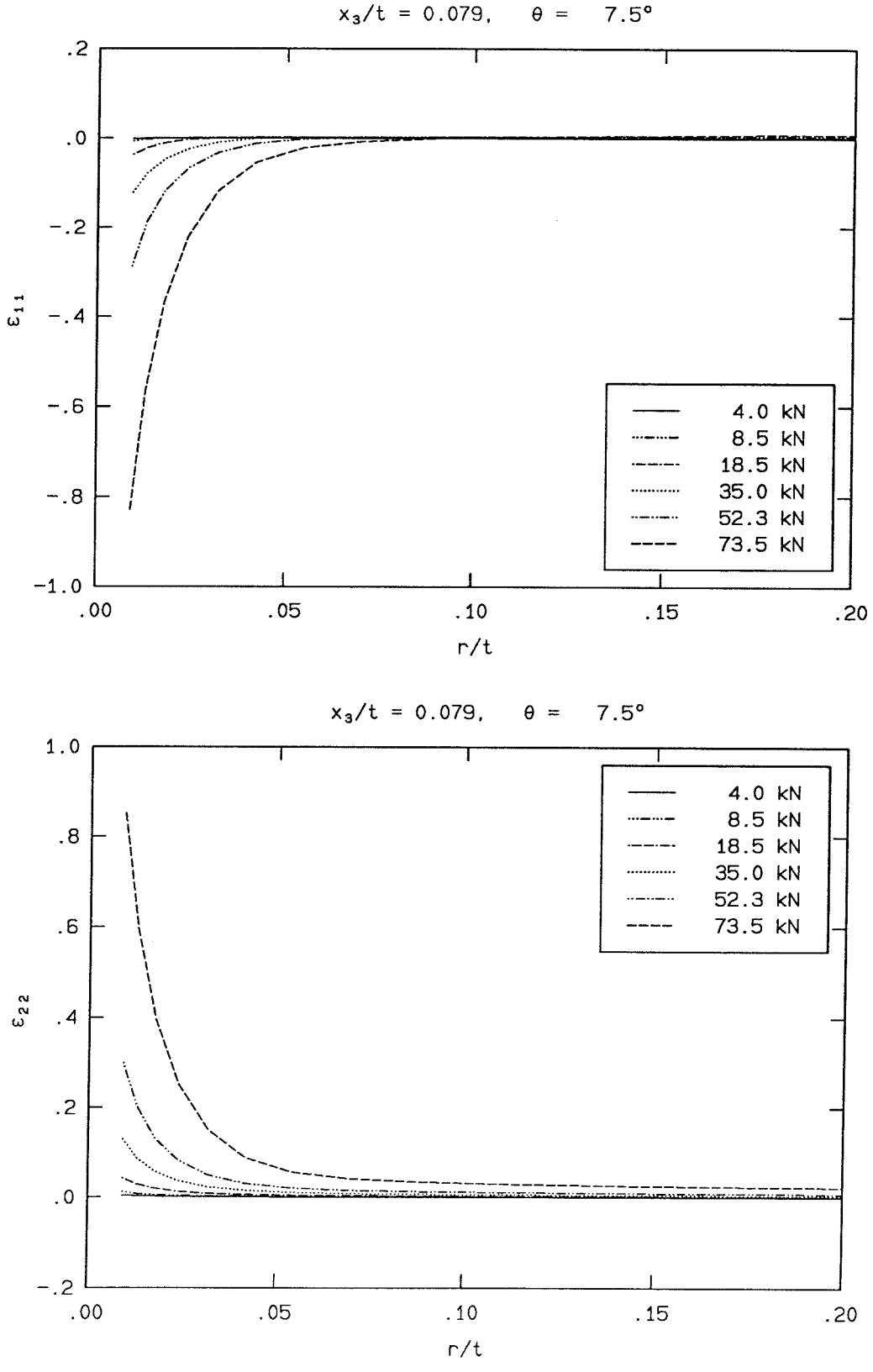


FIGURE 172. Radial variation of ϵ_{11} and ϵ_{22}. Center Layer.

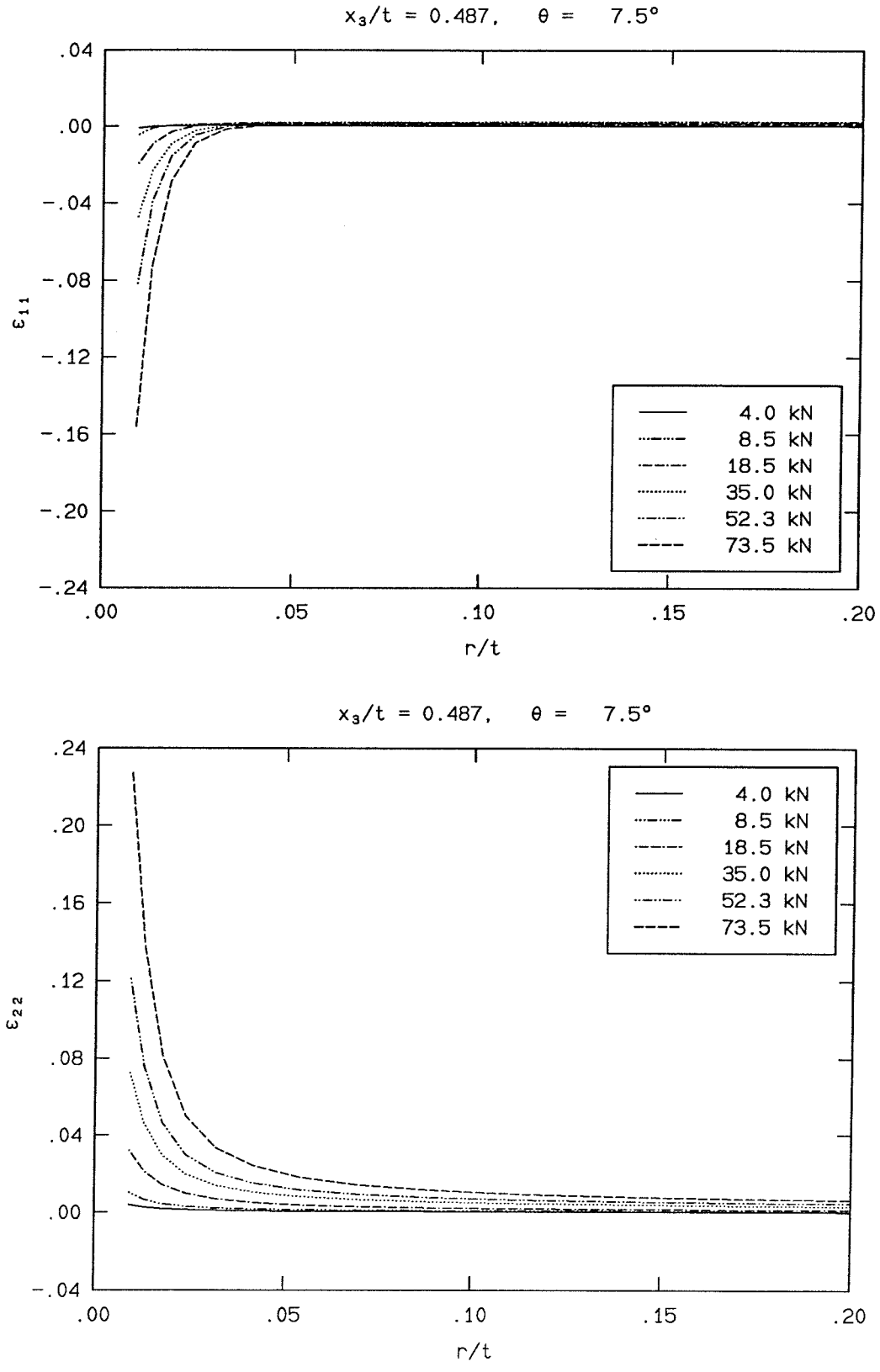


FIGURE 173. Radial variation of ϵ_{11} and ϵ_{22}. Surface Layer.

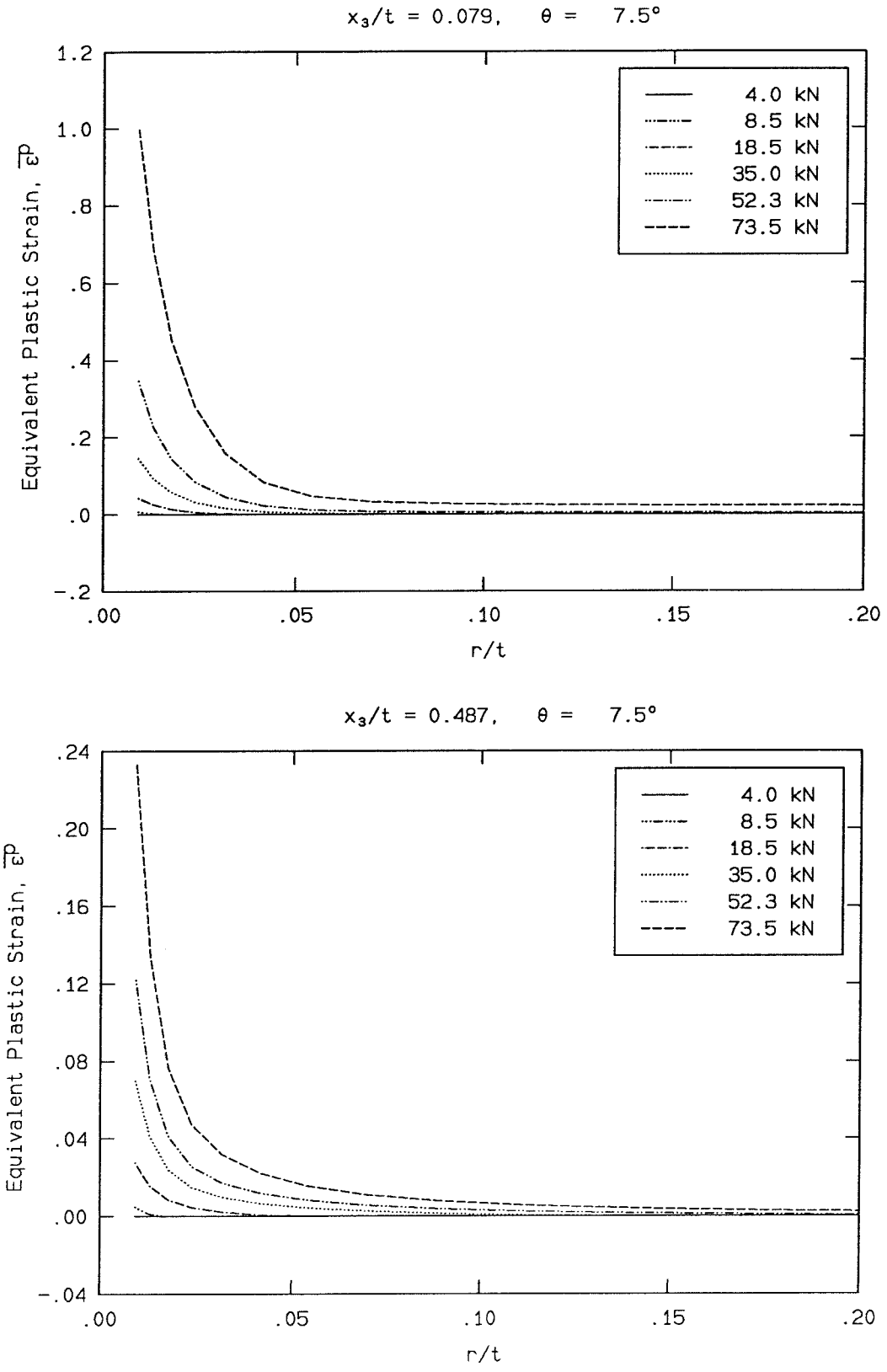


FIGURE 174. Radial variation of $\bar{\epsilon}^p$. Center and Surface Layers.

5.6 Fracture Criterion

It was hoped that some simple criterion for crack propagation might be extracted from a comparison between the results from the numerical model that included tunneling and the results of the equivalent numerical model without tunneling. An example of such a criterion would be a limit on the stresses or strains observed in the case with tunneling that does not appear in the case without tunneling; this limit would correspond to a critical stress or strain for the material. However, there seems not to be a clear trend in the data to make such a critical stress or strain evident.

Tunneling was initiated at a load of 63.5 kN, and comparisons can be made between the case with tunneling and the case without, at loads of 68.5 kN and 73.5 kN. The extent of the tunnel included in the numerical model was about 0.29 mm deep in the center of the specimen at 68.5 kN and about 0.67 mm deep in the center of the specimen at 73.5 kN. (The plate thickness of the specimen was 10.0 mm.)

Figure 175 shows the effect on σ_{22} at $\theta = 7.5°$ in the center of the specimen. One obvious feature is the noise in the data where the tunnel has been included. Part of this is likely due to the complicated geometry of the added tunnel, and part to the fact that the stress is measured in the interior of each element, away from the (now traction–free) boundary. It might be possible to reduce this noise by decreasing the tolerance on the calculation, but adding the tunnel already causes convergence difficulties. It is also clear that the maximum calculated σ_{22} is higher for each case with tunneling than for the corresponding case without. A similar result is seen for the hydrostatic stress in Figure 176.

A similar lack of a clear fracture criterion is found in comparing the strain. The effect on ϵ_{22} is shown in Figure 177. The case that includes tunneling indicates the effect of the incremental plasticity law, whereby the specimen unloads elastically, so that the strain at the tip remains nearly constant. There seems not to be any appreciable increase in strain ahead of the notch tip because of the tunnel, however.

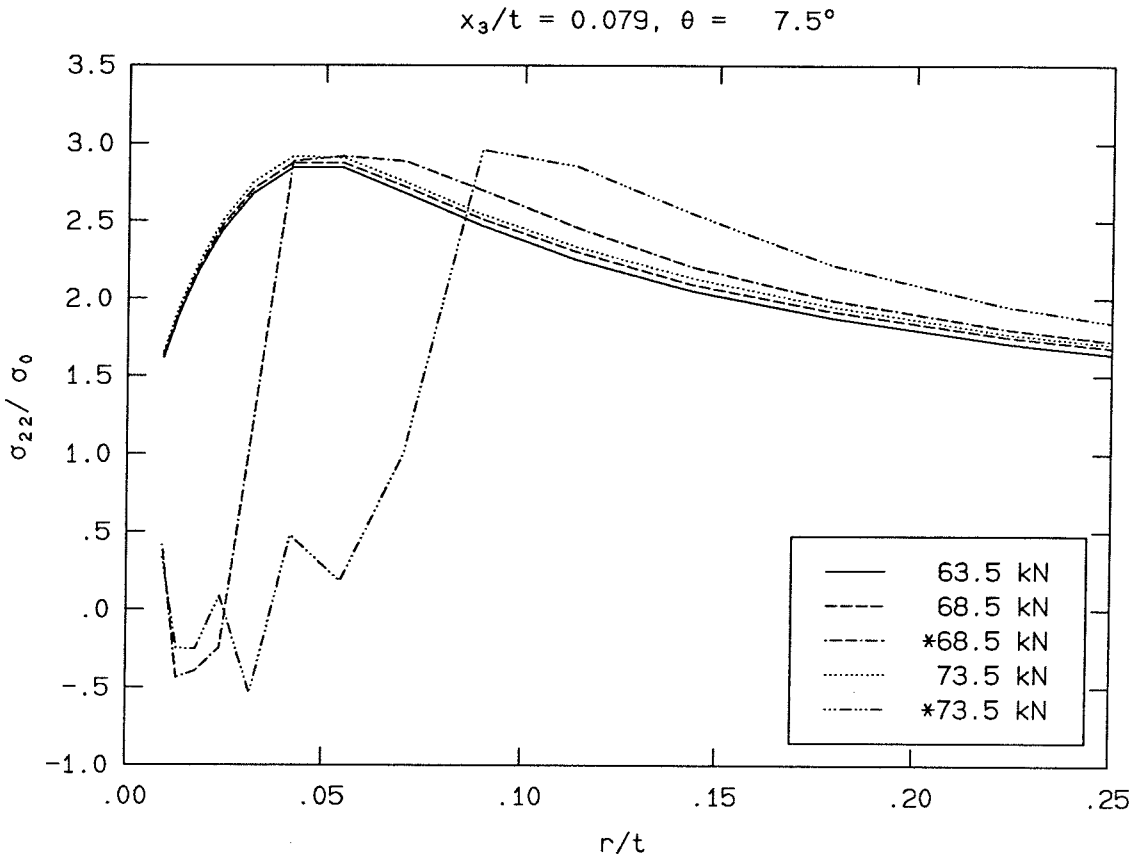


FIGURE 175. The effect on σ_{22} of including artificial tunneling. The loads marked with an asterisk are those that included tunneling.

At 63.5 kN, before tunneling was initiated, the equivalent plastic strain, $\bar{\epsilon}^p$, is approximately 60% in the center of the specimen at the tip of the notch, as shown in Figure 178. The equivalent plastic strain is a scalar quantity that represents the history of the loading, as it is the integration over time of the magnitude of the plastic strain–rate tensor. In the case where no tunneling is included, the equivalent plastic strain reaches 100% at 73.5 kN. The case that includes tunneling indicates the effect of the incremental plasticity law, whereby the specimen unloads elastically, so that the magnitude of the equivalent plastic strain remains unchanged. However, it is interesting that the inclusion of the tunnel does not introduce any large strain peaks in the vicinity of the tip of the tunnel. This means that the peaks in the stress field are due mainly to hydrostatic stress, which does not contribute to the

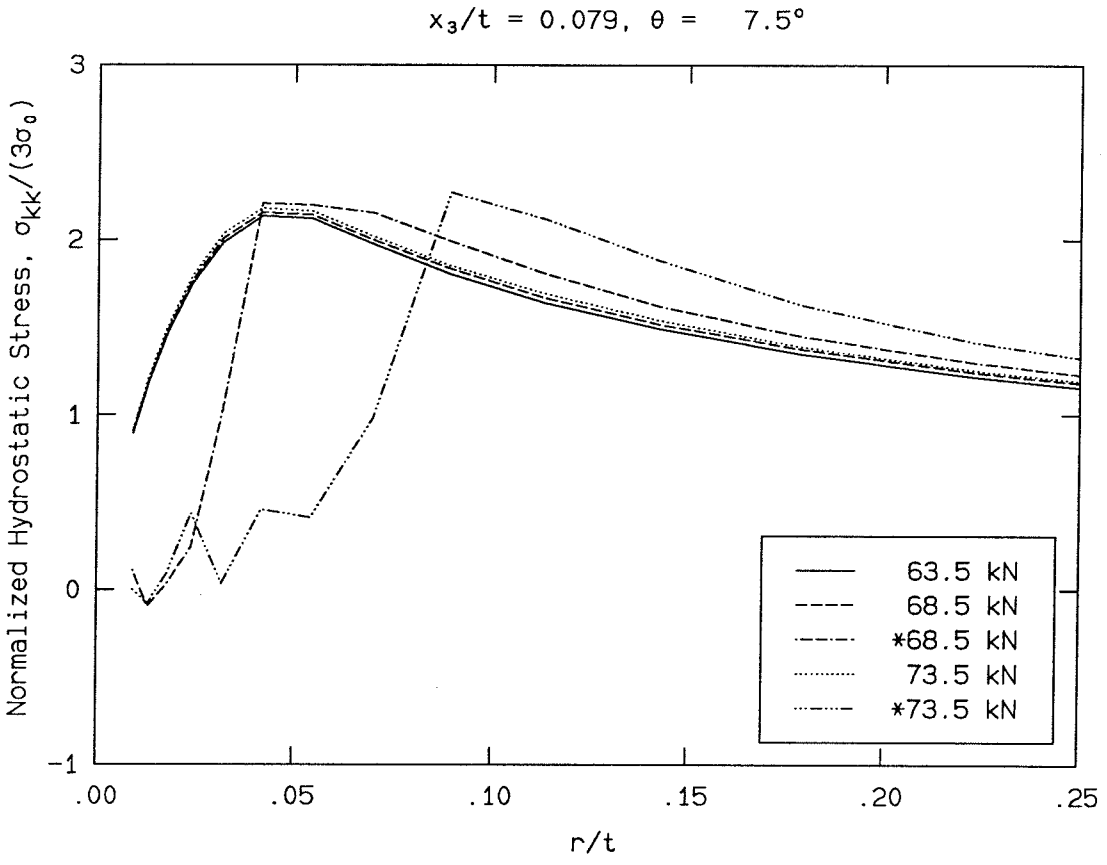


FIGURE 176. The effect on the hydrostatic stress of including artificial tunneling. The loads marked with an asterisk are those that included tunneling.

plastic strain. At 68.5 kN, the tunnel tip is at about $r/t = 0.04$, and at 73.5 kN, the tunnel tip is at about $r/t = 0.08$.

At best, one might make a case for some criterion that includes the effects of both plastic strain and hydrostatic stress, the result being a failure surface similar to the yield surface used in continuum plasticity. Initiation of tunneling seems to be brought about by the high plastic strains at the notch tip, but progression of the tunnel seems to be driven by the hydrostatic stress, since the plastic strains do not achieve the same high values that initiated the tunnel. The Gurson model used by Narasimhan, Rosakis and Moran [52] includes the effects of plastic strain and hydrostatic stress in a model of void nucleation and growth. However, this model had difficulty in converging to a solution as element failure began. Perhaps the

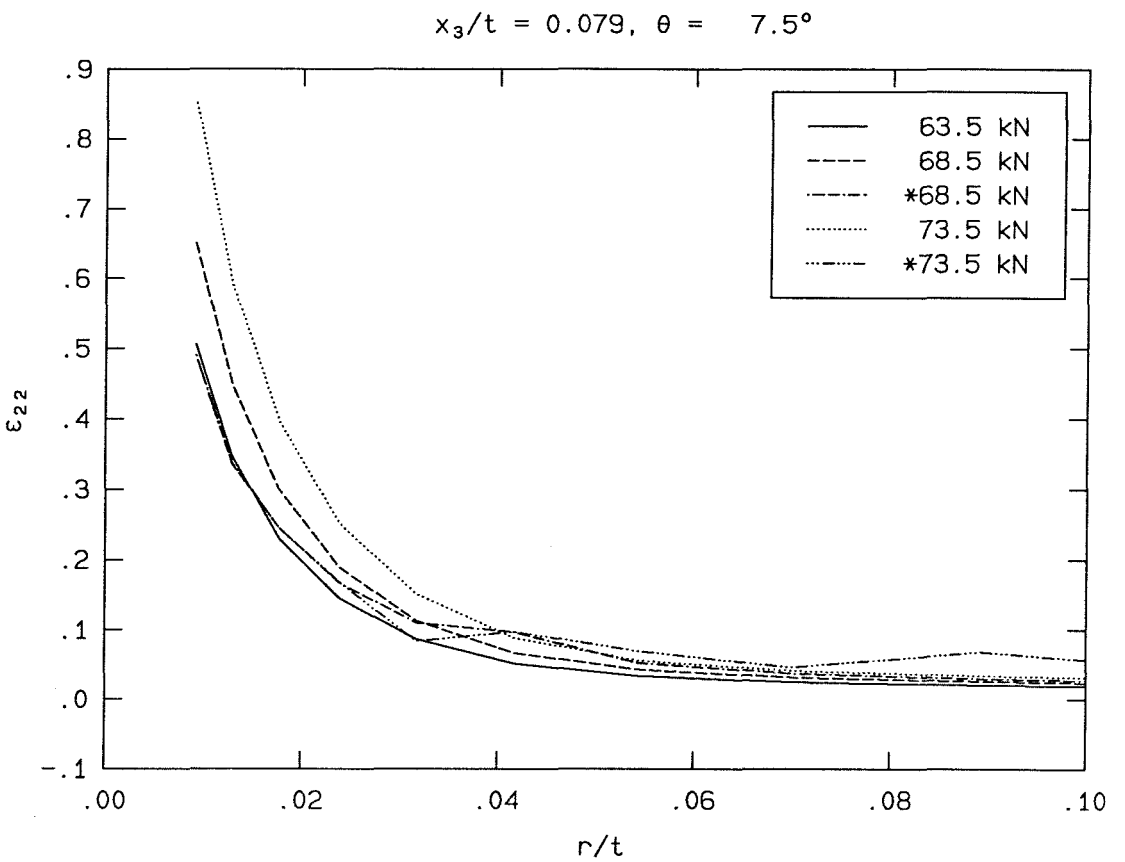


FIGURE 177. The effect on ϵ_{22} of including artificial tunneling. The loads marked with an asterisk are those that included tunneling.

complexity of the void growth model could be reduced to something as simple as a failure surface and the model validated at larger deformation levels. It is important to note that the ultimate failure of the body was through shear failure from the corners of the notch, rather than through fracture along the initial crack plane, which is followed by the tunnel. Thus, an appropriate failure criterion should be able to capture the transition from flat fracture to shear fracture after some tunneling.

The radial variations of the hydrostatic stress and the equivalent plastic strain are shown in Figures 179 and 180 along $\theta = 7.5°$ for the center and surface layers of the specimen. The normalization is taken from the functional representation for the HRR field stresses and strains used by Shih [77]. The normalization appears to work well in the near–tip region, although the distributions are not well described by the

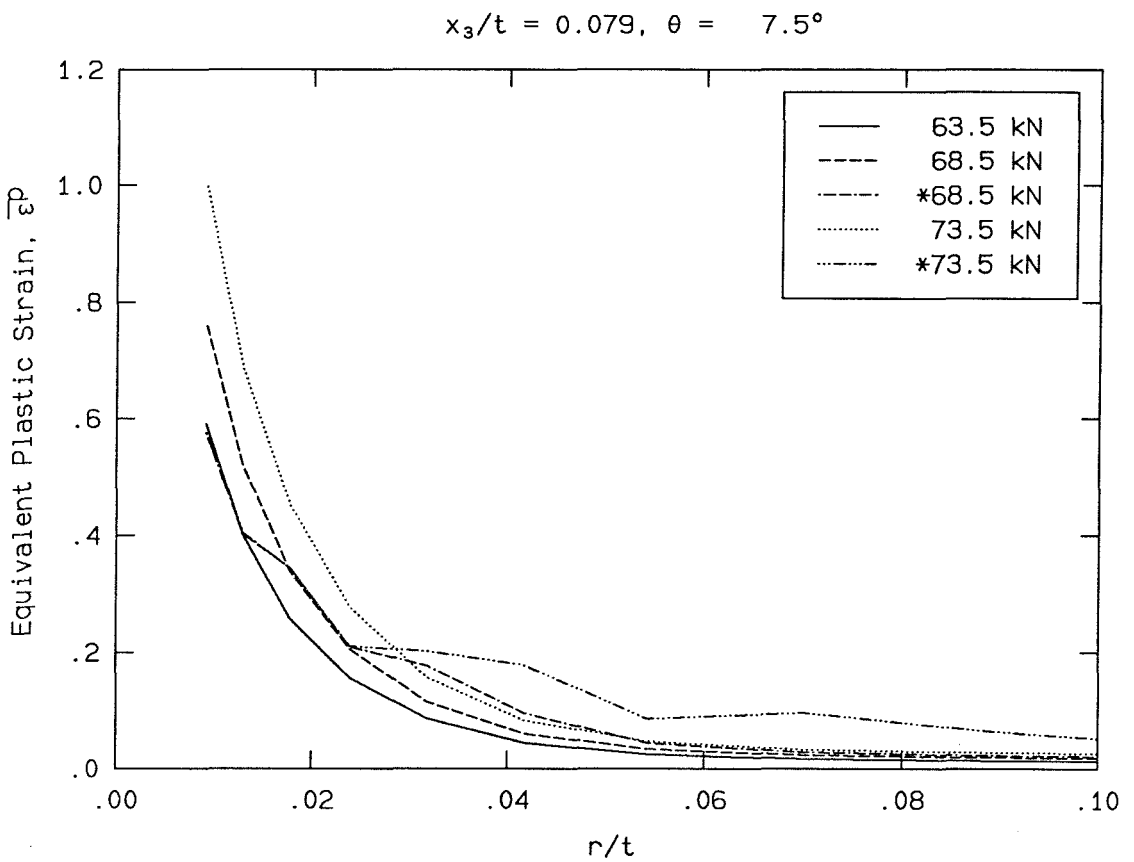


FIGURE 178. The effect on the equivalent plastic strain of including artificial tunneling. The loads marked with an asterisk are those that included tunneling.

HRR field. The J value used for the normalization is the thickness average of the local $\hat{J}(x_3)$ calculated by the numerical model. It appears that the normalization might be improved if the appropriate local value of J were used at each layer, particularly in the equivalent plastic strain, as the dependence on J is much stronger in the strain–normalization function. Thus, it seems that the local value of J is a good predictor of fracture initiation, even though the HRR field is not a good representation of the state of the deformation.

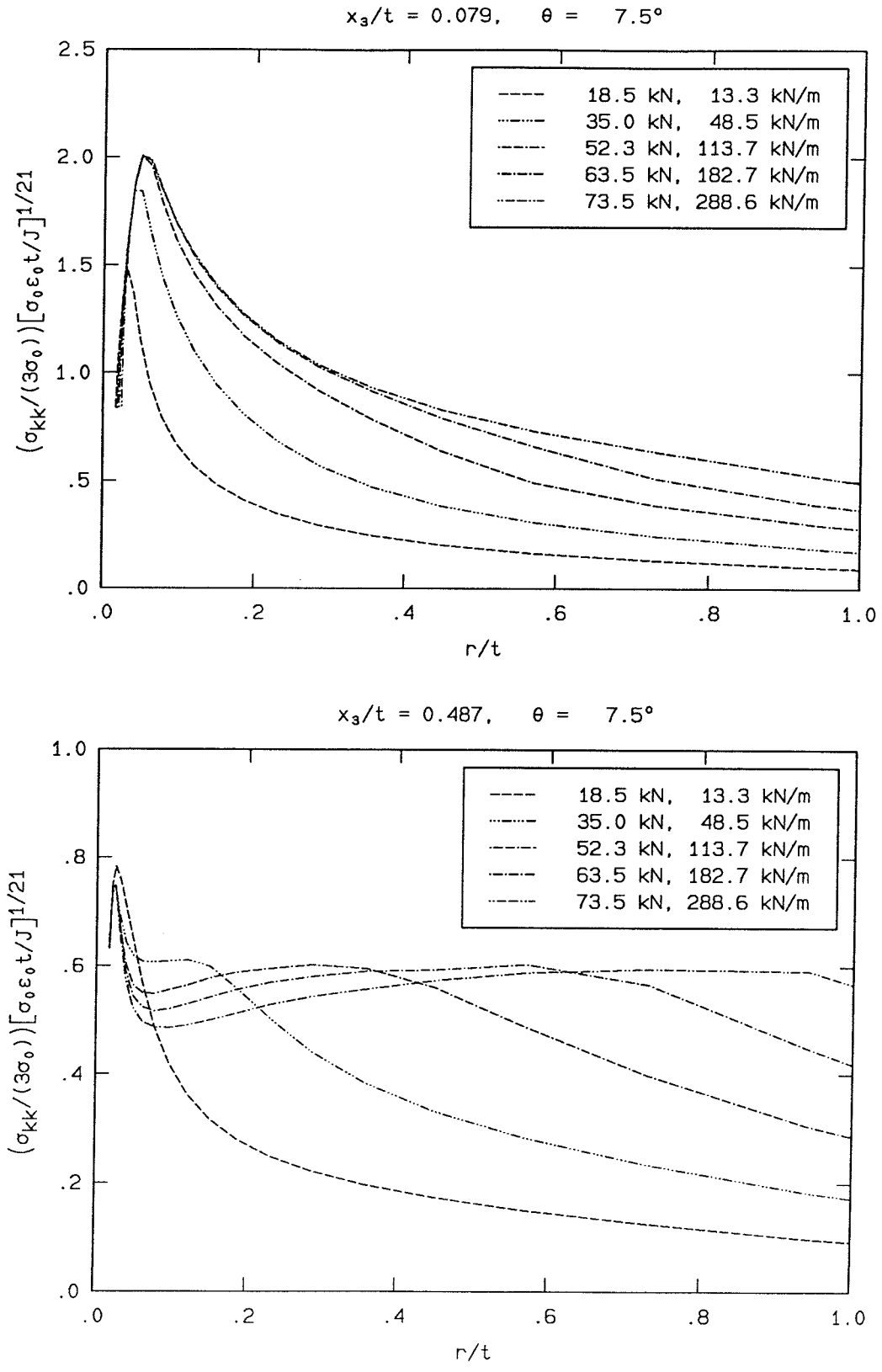


FIGURE 179. Radial variation of σ_{kk}. J normalization.

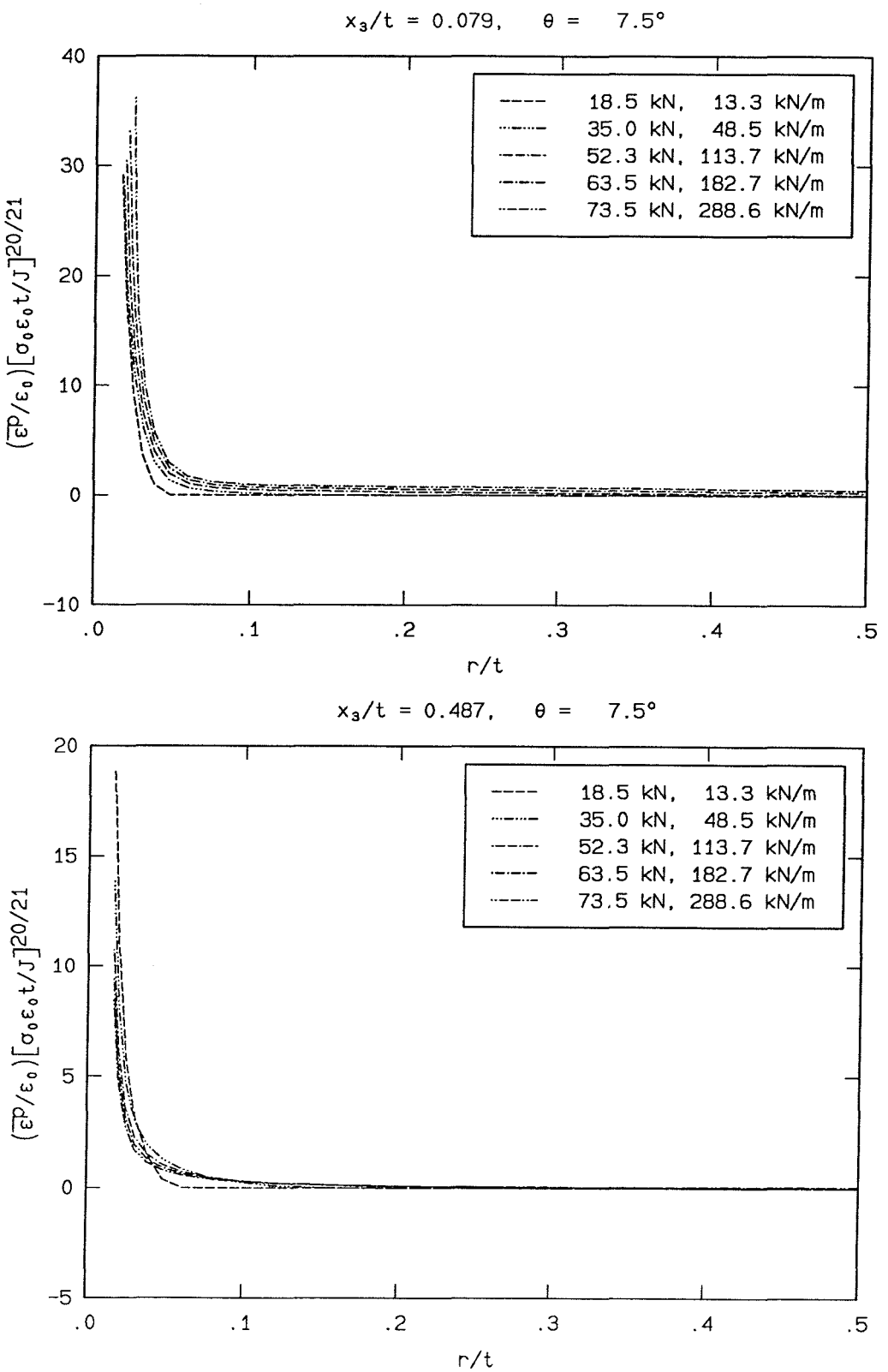


FIGURE 180. Radial variation of $\bar{\epsilon}^p$. J normalization.

5.7 J–Integral Comparison

The load and load–point displacement records from the experiment and the finite–element model match quite well, as shown in Figure 101. As a result, the variation of the J–integral with load as calculated from the work of the indenter using equation (4.4.1) also shows good agreement between the experiment and the finite–element model, as seen in Figure 102. However, the comparison between J values calculated using equation (4.4.1) and J values calculated from the domain integral implemented in the numerical code do not agree very well at higher loads, as shown in Figure 103. The results of the domain integral shown in Figure 103 represent the thickness average of the local, two–dimensional value of the J–integral, $\hat{J}(x_3)$, which is calculated at each layer of nodes.

The domain–integral calculation employed in the numerical model may be used to estimate a critical value of the local value of $\hat{J}$ taken at $x_3 = 0$, where the local $\hat{J}$ is always highest. Tunneling is estimated to initiate at a load between 63.5 kN and 68.5 kN. The values of $\hat{J}(0)$ from the finite–element model at these loads are 219 kN/m and 279 kN/m.

A critical value of J may also be estimated from the tunneling tests, as described in Section 4.5.3 and shown in Figures 107 and 108. This critical J value is estimated to lie between 218 kN/m and 265 kN/m.

The estimate of the critical J value from the experiments agrees quite well with the estimate of the critical value of $\hat{J}(0)$ from the finite–element model. This good agreement suggests that the J value calculated from the load and load–point displacement record of the experiment (or the numerical model) should be compared with the value of $\hat{J}$ taken at the center of the specimen, at $x_3 = 0$. This comparison is shown in Figure 181. The agreement between $\hat{J}(x_3)$ taken at $x_3 = 0$ and J calculated from the experiment appears better than the agreement between the thickness average of $\hat{J}$ and the experiment at the higher loads. One possible explanation is that the local $\hat{J}$ at the center of the specimen more nearly matches the plane–strain condition for which equation (4.4.1) is derived.

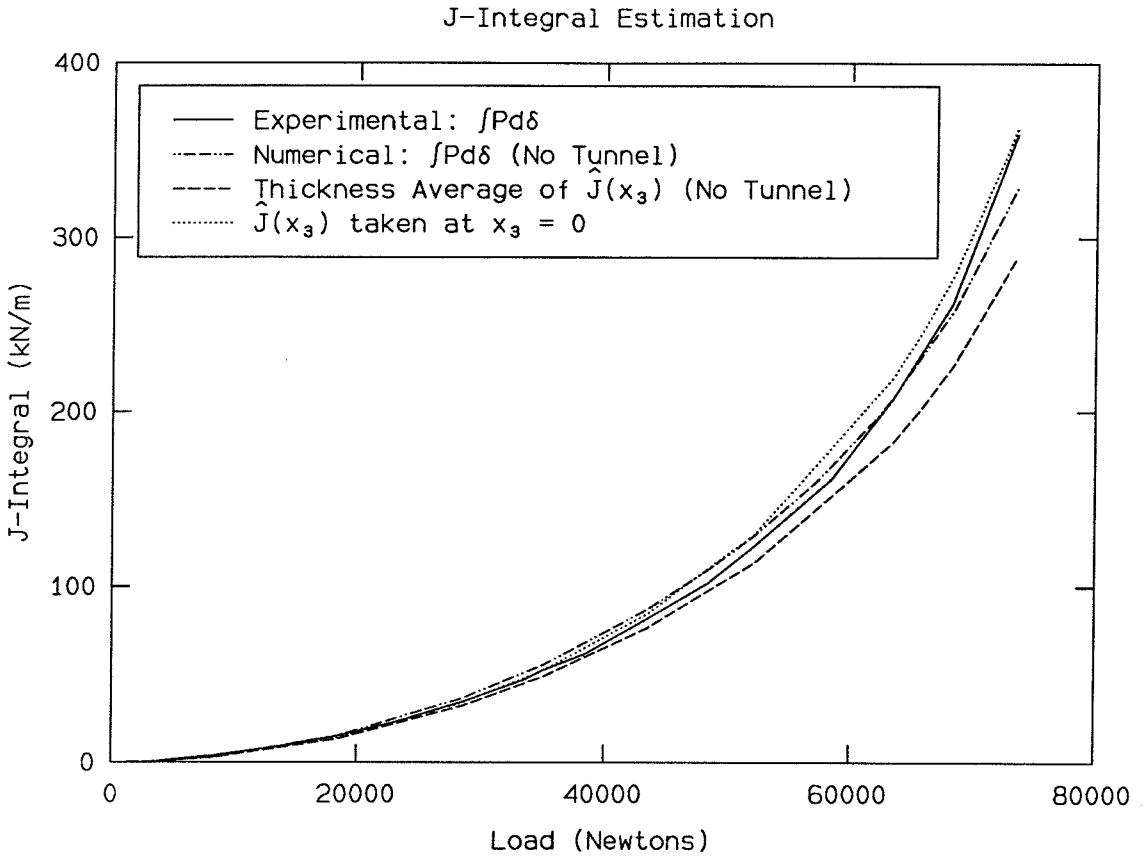


FIGURE 181. Comparison of J values from the experiment and the finite–element model. The first two curves show values of the J–integral calculated from equation (4.4.1), using the experimental and numerical load and load–point displacement records. The dashed line represents the thickness average of the local value of $\hat{J}(x_3)$ and the dotted line represents the local $\hat{J}$ at the center of the specimen.

The plane–strain fracture toughness of a ductile material is always less than the plane–stress fracture toughness, as plane–strain conditions constrain the material so that little shear deformation can occur. Consequently, use of the plane–strain fracture toughness provides an additional margin of safety in the design of structures. However, simulating plane–strain conditions requires a very large and thick specimen, and the specimen geometry used in this study does not satisfy the size requirements to approximate plane strain. It would be interesting to compare the critical values of J calculated with this specimen geometry, particularly the critical

$\hat{J}$ at the center of the specimen, with the plane–strain critical value of J determined for the same material, using both experiment and a numerical model. If the center of the thinner specimen does approach plane–strain conditions, then the results from a thinner specimen might be used to estimate the plane–strain fracture toughness. The use of a thinner specimen is desirable because it requires less material and is thus less expensive, and because the loading could be accomplished using a smaller and less expensive load frame. It would also be interesting to investigate whether the value of J calculated from the work of the indenter correlates well with the local value of $\hat{J}$ at the center of specimens of various thicknesses.

CHAPTER 6

Conclusion

The experimental investigation into the static fracture of a tough material using interferometry is intended to bridge the gap between the behavior away from the effects of the crack tip, where the material response is linearly elastic and assumed to be well understood, and the behavior near the crack tip, where the material is undergoing large plastic deformation. Interferometry is a useful tool in this regard, because it combines a high sensitivity with applicability over a large field of view, and with the proper recording optics, large deformation gradients may be measured. The experiments performed thus far have concentrated on the investigation of the external part of the deformation field. Further investigations closer to the crack tip can use these external measurements as a starting point, so that the field of view may be decreased, the magnification increased and the lens aperture moved closer to the object (reducing the shadows that were due to the aperture limitation). In addition, the grating profile could be etched into the surface of the specimen, removing the possibility that the epoxy layer does not transfer the deformation from the specimen to grating exactly.

The experiment was also compared to a finite–element model, in order to determine how well the finite–element calculation could capture the deformation. The material properties of the material used in the experiment were measured for input to the numerical code so that the finite–element model would match the experiment as closely as possible.

With the experimental apparatus and numerical code available, it is desirable to pursue this line of investigation for different materials (such as for different heat treatment of 4340 steel), and for specimens of varying thickness, to get some idea of the effects of varying the parameters of the problem.

The comparison between the displacements measured in the experiment and the displacements calculated by the numerical model is quite good for loads up to 52.3 kN, in a relatively large region surrounding the notch tip. Near the notch tip, as r decreases, the numerically calculated displacements show an interesting change in direction that should be the focus of further experimental measurements. At 73.5 kN, the highest load achieved in the experiment before continuous slow fracture began, the numerical model and the experiment compared less well. There are several possible sources for this discrepancy that could be addressed. The first source is tunneling of the crack; an attempt was made to eliminate this problem by including a tunnel through releasing nodes at the higher loads. The second source of discrepancy is the lack of a finite–strain formulation in the numerical model; at 52.3 kN the strains are already on the order of 30% in the center of the specimen. A third source of discrepancy may be due to the growth of the plastic zone into a region where the mesh is too coarse to model the deformation adequately. A comparison with the u_3 displacement by Pfaff [56] indicates that more elements are needed through the thickness, the ratio of the out-of-plane mesh dimension to the in-plane mesh dimension is 6:1 at the notch tip. *

The numerical model predicts an interesting change in the slope of both the u_1 and u_2 displacements as a function of r near the crack tip. This feature has yet to be confirmed in the experiment; validating such behavior would be a good task for a future experiment.

The added tunnel did move the displacement in the proper direction, and it resulted in a region of unloading on the surface of the specimen that corresponds fairly well with the regon of lowered fringe density seen in the u_2 displacement fringe pattern.

One source of discrepancy that seemed not to play a role in the comparison is the effect of Lüder's bands (attributed to the release of pinned dislocations) seen in the uniaxial tests. This effect should have appeared in the experiment at the

* The San Diego Supercomputer Center has recently upgraded from a Cray X-MP to a Cray Y-MP that is considerably faster. It might be possible to further double the number of elements in the model without requiring an out-of-core solution procedure.

elastic–plastic boundary, as it appears in the uniaxial test at the yield point. It was anticipated that this effect might appear in the photographs as a discontinuity in the fringe patterns of some sort, but no discontinuities are visible in the fringe photographs in the vicinity of the elastic–plastic boundary. It may be that this is a small local effect that appears large only in the uniaxial test because the entire uniaxial test specimen is in a state of homogeneous stress and strain, so that a very large number of pinned dislocations are released at once. (It would be interesting to study the local nature of Lüder's bands using the moire interferometer.)

The finite notch tip negates any singularity in the stress and strain fields. In particular, the combination of the finite notch tip and plastic deformation result in a local maximum in the σ_{22} stress away from the notch tip. The paper by Rice and Johnson [67] indicates that if finite deformation is included, even an initially sharp crack will blunt to a notch, resulting in nonsingular stresses and strains.

The notch correction from [83] to the K field [90] appears to model adequately the stresses in the vicinity of the notch tip in the elastic regime; it would probably be best to use a thickness average of the numerical stresses for comparison, as in generalized plane stress. The singular HRR field stresses [77] are not very good predictors of the stress calculated using the finite–element model. Perhaps an equivalent notch correction could be produced for the HRR field that would improve the comparison.

Three–dimensional effects seem to be contained within a region that is on the order of the plastic zone size plus one-half the plate thickness.

In comparing the stresses and strains from the cases with and without tunneling, it appears that a fracture criterion must include the effects of both the plastic strain and the hydrostatic stress, such as the Gurson model used by Narasimhan, Rosakis and Moran [52]. † However, a simpler model might be possible, one that would predict material failure when some function of the current stress and strain

† The Gurson model of [52] encountered convergence difficulties as element failure initiated. If the convergence difficulty was simply a matter of waiting for more iterations, it might be interesting to try to run the Gurson model again on the faster Cray Y-MP machine at the San Diego Supercomputer Center.

reaches a limit, similar to the yield surface employed in continuum plasticity. It is interesting to note that although the HRR field does not model the deformation well, the J normalization used for the HRR field by Shih [77] does seem to collapse the data reasonably well in the region close to the notch tip, and it appears that using the local J value would further improve the match. Thus a fracture criterion based on J would still be useful.

One of the biggest difficulties with both the interferometric experimental technique and the finite–element model is that they result in extremely large amounts of raw data. A better theoretical framework is still needed to reduce the data sensibly. It would be possible to fit the data to some function using some large number of parameters, but the choice of the form of that function should have some basis in the physics of the problem.

In order to make use of the information available from the interior of the numerical model, and assuming that the displacements have compared well with the experiment on the exterior, the solution must be unique. This is a nonstandard uniqueness problem, as both the tractions (=0) and the displacements are specified on the same region of the boundary. Such a uniquness theorem is available for linear elasticity [35]; an extension to a case involving plasticity is necessary.

As computers become faster, it will be possible to perform detailed finite–element calculations (including finite strains and automatic remeshing) of just about any problem, effectively eliminating the need to make experimental models. However, the accuracy of such calculations will depend on the accuracy of the constitutive model input to the code, which will still require experimental measurement. Thus, it seems that the best use of the experimental methods would be in simpler tests that concentrate on the constitutive behavior of materials, such as the uniaxial test (and such effects as Lüder's bands). The fracture specimen has some interesting attributes in this sense, as it produces high stresses in a localized area with relatively small forces, and the initial stages of fracture are stable, but the complicated stress and strain fields make it difficult to glean any information about the material constitutive behavior.

References

1. Anderson, T.L., McHenry, H.I. and Dawes, M.G., "Elastic–Plastic Fracture Toughness Tests with Single–Edge Notched Bend Specimens," *Elastic–Plastic Fracture Test Methods: The User's Experience*, ASTM Special Technical Publication 856, pp. 210-229, E.T. Wessel and F.J. Loss, Editors, American Society for Testing and Materials, Philadelphia, 1985.

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

3. Ashby, M.F., and Jones, R.H.J., *Engineering Materials 2, An Introduction to Microstructures, Processing and Design*, Pergamon Press, Oxford, 1986.

4. ASTM E 18–79, "Standard Test Method for Rockwell Hardness and Rockwell Superficial Hardness of Metallic Materials," Annual Book of ASTM Standards, Part 10, pp. 243-266, American Society for Testing and Materials, Philadelphia, 1982.

5. ASTM E 399–81, "Standard Test Method for Plane–Strain Fracture Toughness of Metallic Materials," Annual Book of ASTM Standards, Part 10, pp. 592-622, American Society for Testing and Materials, Philadelphia, 1982.

6. ASTM E 646–78, "Standard Test Method for Tensile Strain–Hardening Exponents (n–Values) of Metallic Sheet Materials," Annual Book of ASTM Standards, Part 10, pp. 762-771, American Society for Testing and Materials, Philadelphia, 1982.

7. ASTM E 813–81, "Standard Test Method for J_{Ic}, A Measure of Fracture Toughness," Annual Book of ASTM Standards, Part 10, pp. 822-840, American Society for Testing and Materials, Philadelphia, 1982.

8. Atkins, A.G., and Mai, Y-W., *Elastic and Plastic Fracture Metals, Polymers, Ceramics, Composites, Biological Materials*, Ellis Horwood Limited, Chichester, 1988.

9. Bathe, K-J., *Finite Element Procedures in Engineering Analysis*, Prentice–Hall,

Inc., Englewood Cliffs, N.J., 1982.

10. Begley, J.A. and Landes, J.D., "The J Integral as a Fracture Criterion," *Fracture Toughness, Proceedings of the 1971 National Symposium on Fracture Mechanics, Part II*, ASTM Special Technical Publication 514, pp. 1-20 American Society for Testing and Materials, Philadelphia, 1972.

11. Born, M. and Wolf, E., *Principles of Optics*, Sixth Edition, Pergamon Press, Oxford, 1980.

12. Bridgman, P.W., "The Stress Distribution at the Neck of a Tension Specimen," *Transactions of the American Society of Metals*, Vol. 32, 553-574, 1944.

13. Bridgman, P.W., *Studies in Large Plastic Flow and Fracture*, Metallurgy and Metallurgical Engineering Series, R.F. Mehl, Consulting Editor, McGraw–Hill Book Company, Inc., New York, 1952.

14. Broek, D., *Elementary Engineering Fracture Mechanics*, Noordhoff International Publishing, Leyden, 1974.

15. Bucci, R.J., Paris, P.C., Landes, J.D. and Rice, J.R., "J Integral Estimation Procedures," *Fracture Toughness, Proceedings of the 1971 National Symposium on Fracture Mechanics, Part II*, ASTM Special Technical Publication 514, pp. 40-69 American Society for Testing and Materials, Philadelphia, 1972.

16. Budiansky, B., "A Reassessment of Deformation Theories of Plasticity," *Journal of Applied Mechanics*, Vol. 26, No. 2, 259-264, June, 1959.

17. Durelli, A.J. and Parks, V.J., *Moire Analysis of Strain*, Prentice–Hall, Inc., Englewood Cliffs, N.J., 1970.

18. Durelli, A.J. and Parks, V.J., "Moire Fringes as Parametric Curves," *Experimental Mechanics*, Vol. 7, 97-104, March, 1967.

19. Fowles, G.R., *Introduction to Modern Optics*, Holt, Rinehart and Winston, Inc., New York, 1968.

20. Futato, R.J., Aadland, J.D., Van Der Sluys, W.D. and Lowe, A.L., "A Sensitiv-

ity Study of the Unloading Compliance Single–Specimen J–Test Technique," *Elastic–Plastic Fracture Test Methods: The User's Experience*, ASTM Special Technical Publication 856, pp. 84-103, E.T. Wessel and F.J. Loss, Editors, American Society for Testing and Materials, Philadelphia, 1985.

21. Goodman, J.W., *Introduction to Fourier Optics*, McGraw–Hill, San Francisco, 1968.

22. Gross, B. and Srawley, J.E., "Stress–Intensity Factors for Three–Point Bend Specimens by Boundary Collocation," *NASA Technical Note TN D-3092*, National Aeronautics and Space Administration, Washington, D.C., December, 1965.

23. Guild, J., *The Interference Systems of Crossed Diffraction Gratings*, Clarendon Press, Oxford, 1956.

24. Guild, J., *Diffraction Gratings as Measuring Scales*, Oxford University Press, London, 1960.

25. *Handbook on Experimental Mechanics*, A.S. Kobayashi, Editor, Society for Experimental Mechanics, Inc., Prentice–Hall, Inc., Englewood Cliffs, New Jersey, 1987.

26. Hertzberg, R.W., *Deformation and Fracture Mechanics of Engineering Materials*, Third Edition, John Wiley and Sons, New York, 1989.

27. Hettrick, M.C., "Aberrations of Varied Line-Spacing Grazing Incidence Gratings in Converging Light Beams," *Applied Optics*, Vol. 23, No. 18, 3221-3235, September, 1984.

28. Hill, R., *The Mathematical Theory of Plasticity*, Clarendon Press, Oxford, 1950.

29. Hollstein, T., Blauel, J.G., and Voss, B., "On the Determination of Elastic–Plastic Fracture Material Parameters: A Comparison of Different Test Methods," *Elastic–Plastic Fracture Test Methods: The User's Experience*, ASTM Special Technical Publication 856, pp. 104-116, E.T. Wessel and F.J. Loss,

Editors, American Society for Testing and Materials, Philadelphia, 1985.

30. Hughes, T.J.R., "Generalization of Selective Integration Procedures to Anisotropic and Nonlinear Media," *International Journal for Numerical Methods in Engineering*, Vol. 15, 1413-1418, 1980.

31. Hutchinson, J.W., "Singular Behaviour at the End of a Tensile Crack in a Hardening Material," *Journal of Mechanics and Physics of Solids*, Vol. 16, 13-31, 1968.

32. Hutchinson, J.W., "Plastic Stresses and Strain Fields at a Crack Tip," *Journal of Mechanics and Physics of Solids*, Vol. 16, 337-347, 1968.

33. Inglis, C.E., "Stresses in a Plate Due to the Presence of Cracks and Sharp Corners," Transactions of the Institution of Naval Architects, London, England, Vol. 60, p. 219, 1913.

34. Kachanov, L.M., *Foundations of the Theory of Plasticity*, North–Holland Publishing Company, Amsterdam, 1971.

35. Knops, R.J. and Payne, L.E., *Uniqueness Theorems in Linear Elasticity*, Published by Springer–Verlag, Berlin, 1971.

36. Kock, W.E., *Lasers and Holography*, Dover Publications, Inc., New York, 1981.

37. KODAK Pamphlet P-110, *KODAK Materials for Holography*, Eastman Kodak Company, Rochester, New York, 1976.

38. KODAK Pamphlet P-255, *KODAK Technical Pan Film 2415/6415*, Eastman Kodak Company, Rochester, New York, 1985.

39. Landes, J.D. and Begley, J.A., "The Effect of Specimen Geometry on J_{1c}," *Fracture Toughness, Proceedings of the 1971 National Symposium on Fracture Mechanics, Part II*, ASTM Special Technical Publication 514, pp. 24-39 American Society for Testing and Materials, Philadelphia, 1972.

40. Li, F.Z., Shih, C.F. and Needleman, A., "A Comparison of Methods for Calculating Energy Release Rates," *Engineering Fracture Mechanics*, Vol. 21, No.

2, 405-421, 1985.

41. Livnat, A. and Post, D., "The Governing Equations of Moire Interferometry and Their Identity to Equations of Geometrical Moire," *Experimental Mechanics*, Vol. 25, No. 4, 1985.

42. Loewen, E.G., "Diffraction Gratings, Ruled and Holographic," *Applied Optics and Optical Engineering*, Volume IX, pp. 33-72, Shannon, R.R. and Wyant, J.C., Editors, Academic Press, Inc., Orlando, 1983.

43. Matthies, H. and Strang, G., "The Solution of Nonlinear Finite Element Equations," *International Journal for Numerical Methods in Engineering*, Vol. 14, 1613-1626, 1979.

44. McDonach, A., McKelvie, J. and Walker, C.A., "Stress Analysis of Fibrous Composites Using Moire Interferometry," *Optics and Lasers in Engineering*, Vol. 1, 85-105, 1980.

45. McMeeking, R.M., "Finite Deformation Analysis of Crack–Tip Opening in Elastic–Plastic Materials and Implications for Fracture," *Journal of Mechanics and Physics of Solids*, Vol. 25, 357-381, 1977

46. *Metals Handbook*, Ninth Edition, Volume 1, "Properties and Selection: Irons and Steels," American Society for Metals, Ohio, 1978.

47. Nakamura, T., Shih, C.F. and Freund, L.B., "Elastic–Plastic Analysis of a Dynamically Loaded Circumferentially Notched Round Bar," *Engineering Fracture Mechanics*, Vol. 22, No. 3, 437-452, 1985.

48. Nakamura, T., Shih, C.F. and Freund, L.B., "Analysis of a Dynamically Loaded Three-point-bend Ductile Fracture Specimen," *Engineering Fracture Mechanics*, Vol. 25, 323-339, 1986.

49. Narasimhan, R. and Rosakis, A.J., *Influence of Hardening on Quasi-Static Extension of Mode I, Plane Stress Cracks in Isotropic Elastic-Plastic Materials*, SM Report 86-35, California Institute of Technology Report, Pasadena, 1986.

50. Narasimhan, R. and Rosakis, A.J., "A Finite Element Analysis of Small–Scale Yielding Near a Stationary Crack Under Plane Stress," *Journal of Mechanics and Physics of Solids*, Vol. 36, No. 1, 77-117, 1988.

51. Narasimhan, R. and Rosakis, A.J., *Three Dimensional Effects Near a Crack Tip in a Ductile Three Point Bend Specimen, Part I: A Numerical Investigation*, SM Report 88-6, California Institute of Technology Report, Pasadena, January, 1988.

52. Narasimhan, R., Rosakis, A.J. and Moran, B., *A Three Dimensional Numerical Investigation of Fracture Initiation by Ductile Failure Mechanisms in a 4340 Steel*, SM Report 89-5, California Institute of Technology Report, Pasadena, September, 1989.

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

54. Parmerter, R.R. and Mukerji, B., "Stress Intensity Factors for an Edge–Cracked Strip in Bending," *International Journal of Fracture*, Vol. 10, 441-444, 1974.

55. Petit, R., Editor, *Electromagnetic Theory of Gratings*, Topics in Current Physics 22, Springer-Verlag, Berlin, 1980.

56. Pfaff, R., *The Continuum Concept in Fracture Mechanics*, Doctoral Thesis, California Institute of Technology, Pasadena, California, 1991.

57. Pirodda, L., "Strain Analysis by Grating Interferometry," *Optics and Lasers in Engineering*, Vol. 5, No. 1, 1984.

58. Polakowski, N.H., and Ripling, E.J., *Strength and Structure of Engineering Materials*, Prentice–Hall, Inc., Englewood Cliffs, New Jersey, 1966.

59. Post, D., "Sharpening and Multiplication of Moire Fringes," *Experimental Mechanics*, Vol. 7, No. 4, April, 1967.

60. Post, D., "Moire Fringe Multiplication with a Nonsymmetrical Doubly Blazed

Reference Grating," *Applied Optics*, Vol. 10, No. 4, 901-907, April, 1971.

61. Post, D., "Optical Interference for Deformation Measurements – Classical, Holographic and Moire Interferometry," *Mechanics of Nondestructive Testing*, Stinchcomb, W.W., Editor, Plenum Press, New York, 1980.

62. Post, D., "Moire Interferometry at VPI & SU," *Experimental Mechanics*, Vol. 23, No. 2, 203-210, June, 1983.

63. Post, D. and Baracat, W.A., "High Sensitivity Moire Interferometry — A Simplified Approach," *Experimental Mechanics*, Vol. 21, No. 3, March, 1981.

64. Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T., *Numerical Recipes*, Cambridge University Press, Cambridge, 1986.

65. Rice, J.R., "A Path Independent Integral and the Approximate Analysis of Strain Concentration by Notches and Cracks," *Journal of Applied Mechanics*, Vol. 35, 379-386, June, 1968.

66. Rice, J.R., "Mathematical Analysis in the Mechanics of Fracture," *Fracture, An Advanced Treatise*, H. Liebowitz, Editor, Volume II, Mathematical Fundamentals, pp. 191-311, Academic Press, New York, 1968.

67. Rice, J.R. and Johnson, M.A., "The Role of Large Crack Tip Geometry Changes in Plane Strain Fracture," *Inelastic Behavior of Solids*, Kanninen, M.F., Adler, W.F., Rosenfield, A.R., Jaffee, R.I., Editors, pp. 641-672, McGraw–Hill, New York, 1970.

68. Rice, J.R., Paris, P.C. and Merkle, J.G., "Some Further Results of J–Integral Analysis and Estimates," *Progress in Flaw Growth and Fracture Toughness Testing*, ASTM Special Technical Publication 536, pp. 231-245, American Society for Testing and Materials, Philadelphia, 1973.

69. Rice, J.R. and Rosengren, G.F., "Plane Strain Deformation Near a Crack Tip in a Power-law Hardening Material," *Journal of Mechanics and Physics of Solids*, Vol. 16, 1-12, 1968.

70. Rooke, D.P. and Cartwright, D.J., *Compendium of Stress Intensity Factors*, Her Majesty's Stationery Office, London, 1976.

71. Ross, B.E., Sciammarella, C.A., and Sturgeon, D., "Basic Optical Law in the Interpretation of Moire Patterns Applied to the Analysis of Strains – Part 2," *Experimental Mechanics*, Vol. 5, 161-166, June, 1965.

72. Schultheisz, C., *Moire Interferometry*, SM Report 84-4, California Institute of Technology Report, Pasadena, 1984.

73. Schultheisz, C., *Comparison of Numerical and Analytical Crack–Tip Deformations*, SM Report 91-2, California Institute of Technology Report, Pasadena, 1991.

74. Sciammarella, C.A., "The Moire Method – A Review," *Experimental Mechanics*, Vol. 22, 418-433, November, 1982.

75. Sciammarella, C.A., "Basic Optical Law in the Interpretation of Moire Patterns Applied to the Analysis of Strains – Part 1," *Experimental Mechanics*, Vol. 5, 154-160, May, 1965.

76. Sciammarella, C.A., "Techniques of Fringe Interpolation in Moire Patterns," *Experimental Mechanics*, Vol. 7, 19A-30A, November, 1967.

77. Shih, C.F., *Tables of Hutchinson–Rice–Rosengren Singular Field Quantities*, MRL E-147, Materials Research Laboratory Report, Brown University, Providence, 1983.

78. 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, 73-84, 1985.

79. Shih, C.F. and German, M.D., "Requirements for a one parameter characterization of crack tip fields by the HRR singularity," *International Journal of Fracture*, Vol. 17, 27-43, 1981.

80. Shih, C.F., Moran, B. and Nakamura, T., "Energy Release Rate Along a Three–

Dimensional Crack Front in a Thermally Stressed Body," *International Journal of Fracture*, Vol. 30, 79-102, 1985.

81. Stroke, G.W., *An Introduction to Coherent Optics and Holography*, Academic Press, New York, 1969.

82. Swedlow, J.L., *On the Elastic Stress Singularities in a Cracked Plate of Finite Thickness*, SM Report 64-12, California Institute of Technology Report, Pasadena, April, 1964.

83. Tada, H., Paris, P.C., and Irwin, G.R., *The Stress Analysis of Cracks Handbook*, Del Research Corporation, Hellertown, Pennsylvania, 1973.

84. Taylor, J.R., *An Introduction to Error Analysis*, University Science Books, Oxford University Press, 1982.

85. Theocaris, P.S., *Moire Fringes in Strain Analysis*, Pergamon Press, Oxford, 1969.

86. Vest, Charles M., *Holographic Interferometry*, John Wiley & Sons, New York, 1979.

87. Walker, C.A., and McKelvie, J., "A Practical Multiplied–Moire System," *Experimental Mechanics*, Vol. 18, 316-320, August, 1978.

88. Westergaard, H.M., "Stresses at a Crack, Size of the Crack, and the Bending of Reinforced Concrete," *Proceedings of the American Concrete Institute*, Vol. 30, 93-102, 1934.

89. Williams, C.S., and Becklund, O.A., *Optics: A Short Course for Engineers and Scientists*, Wiley–Interscience, New York, 1972.

90. Williams, M.L., "On the Stress distribution at the Base of a Stationary Crack," *Journal of Applied Mechanics*, Vol. 24, 109-114, 1957.

91. Zehnder, A.T., *Dynamic Fracture Initiation and Propagation in Metals: Experimental Results and Techniques*, Doctoral Thesis, California Institute of Technology, 1987.

92. Zehnder, A.T. and Rosakis, A.J., *Three Dimensional Effects Near a Crack Tip in a Ductile Three Point Bend Specimen, Part II: An Experimental Investigation Using Interferometry and Caustics*, SM Report 88-7, California Institute of Technology Report, Pasadena, March, 1988.

93. Zhang, Y., and Ravi–Chandar, K., *A Finite Element Investigation into the Foundations of Elastic–Plastic Fracture Mechanics, Part I: Dominance of the HRR Field*, Department of Mechanical Engineering, University of Houston, Houston, Texas, January, 1990.

APPENDIX A

The Three–Beam Moire Interferometer

A.1 Orthogonal Displacements with the Three–Beam Interferometer

Initially, the moire interferometer intended for use in this study was designed to use three laser beams to measure orthogonal in-plane displacements instead of four laser beams. This type of interferometer is described by McDonach, McKelvie and Walker [44]. The configuration of optical components to make such measurements is shown in Figure 182 and can be compared with the configuration of the four–beam interferometer in Figure 183. Ultimately, the drawbacks of the three–beam interferometer were found to outweigh its advantages over the four–beam interferometer, and it was replaced by the four–beam interferometer.

The three–beam interferometer has several advantages over the four–beam interferometer; it uses two fewer mirrors and only one laser source. However, there are problems due to unequal beam intensity and polarization mismatch, and there is considerably less light as the single laser beam must be divided twice, necessitating longer exposure times and making vibrations more of a problem. Figure 184 shows the three–beam interferometer with the addition of the beam expander to produce collimated beams of 50 mm diameter.

The three–beam interferometer uses two crossed gratings at $\pm 45°$ to the globally fixed X_1 and X_2 axes to measure components of displacement along those axes. As in the four–beam interferometer, the three incoming laser beams are aligned so that the first diffraction order exits along the normal to the specimen surface (the X_3 axis). Interference between any pair of diffracted laser beams will produce fringe patterns which are related to the deformation of the grating.

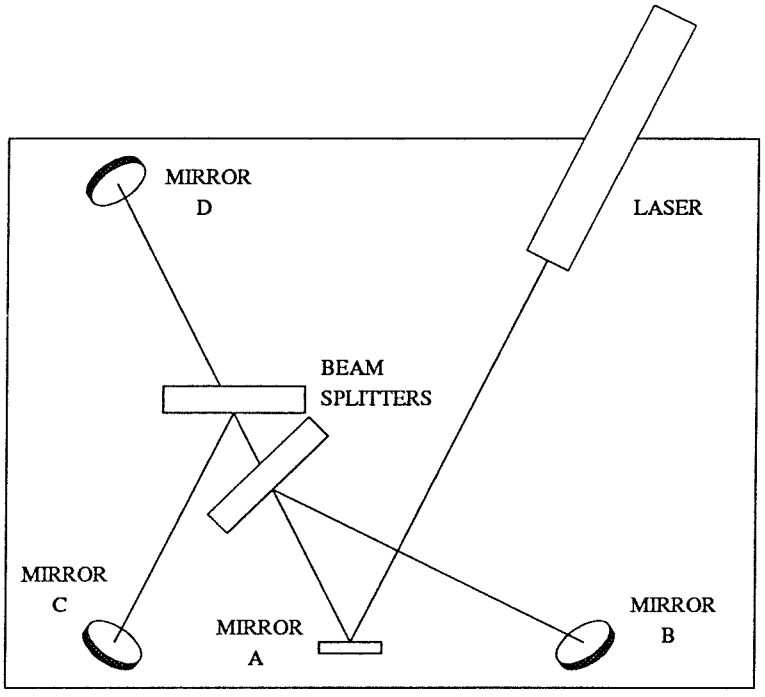


FIGURE 182. Diagram of the three–beam interferometer configuration to measure both horizontal and vertical displacement components. Beams from mirrors B and C yield the horizontal component, and beams from mirrors C and D yield the vertical.

Referring to Figure 182, interference between the beams from mirrors B and C produces a fringe pattern of contours of horizontal displacement; interference between the beams from mirrors C and D produces a fringe pattern of contours of vertical displacement; and interference between the beams from mirrors B and D produces a fringe pattern of contours of displacement parallel to the line between mirrors B and D.

Analysis of the governing equations of fringe formation for the three–beam interferometer will follow the lines of the derivation for the four–beam interferometer. However, only the case of planar deformations will be analyzed. The effects of rotations about axes in the plane can be determined in a manner analogous to that in Section 2.3, which describes the fringes found using a four–beam interferometer with arbitrary changes in the grating orientation.

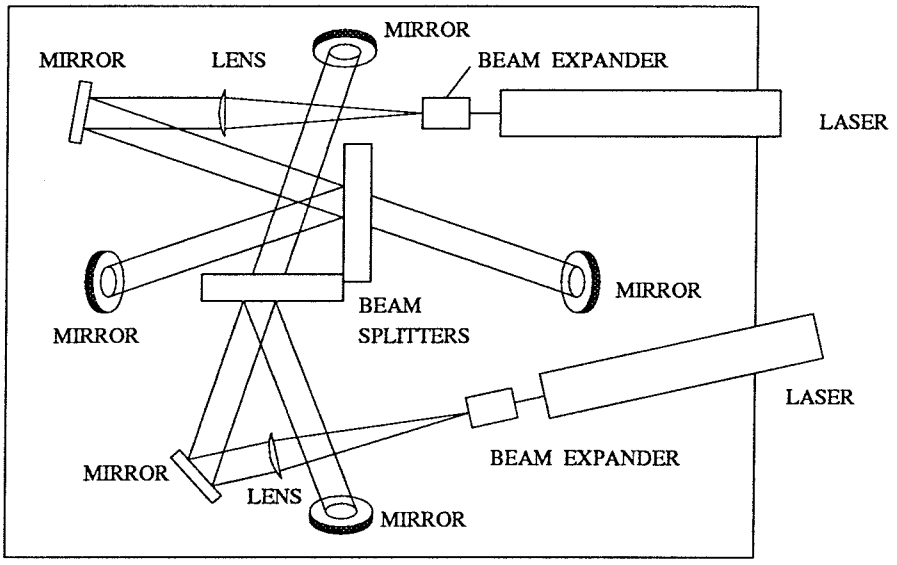


FIGURE 183. Diagram of the four–beam interferometer configuration to measure both horizontal and vertical displacement components.

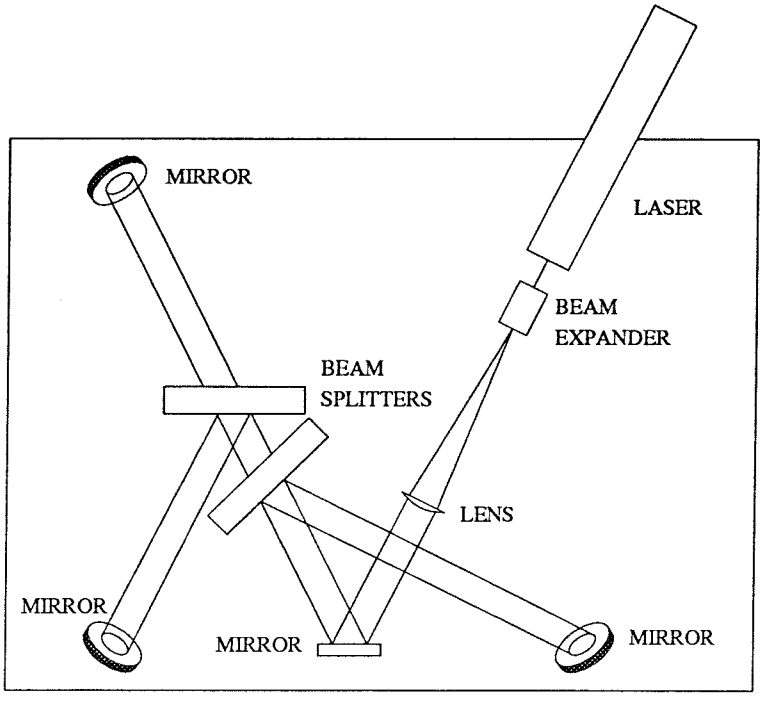


FIGURE 184. Diagram of the three–beam interferometer configuration to measure both horizontal and vertical displacement components, including components to expand the beam.

A.1.1 Geometry Describing the Crossed Gratings

The interference patterns observed will come from two light beams, each diffracted from one of the crossed gratings. The mean grating surface coincides with the $X_1 - X_2$ plane in a globally fixed coordinate frame. Initially, the gratings have identical wavelengths of w_0 and are perpendicular to each other, but at angles of $\pm 45°$ to the X_1 axis.

Let $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$ be unit vectors aligned with the X_1, X_2 and X_3 axes, respectively. Let the unit normal to the ridges of grating #1 be $\mathbf{n}_1$, and let $\mathbf{t}_1$ be a unit vector parallel to the ridges of grating #1 such that $\mathbf{n}_1$, $\mathbf{t}_1$ and $\mathbf{E}_3$ form a right–handed orthogonal coordinate system. Further, let the unit normal to the ridges of grating #2 be $\mathbf{n}_2$, and let $\mathbf{t}_2$ be a unit vector parallel to the ridges of grating #2 such that $\mathbf{n}_2$, $\mathbf{t}_2$ and $\mathbf{E}_3$ form another right–handed coordinate system. Initially, $\mathbf{n}_1$ is at an angle of $\frac{\pi}{4}$ radian to the X_1 axis, and $\mathbf{n}_2$ is at an angle of $\frac{3\pi}{4}$ radian to the X_1 axis. The gratings are allowed to deform independently in the sense that the deformed wavelength and orientation of each grating may be specified separately. If the deformation field is known, the parameters describing the two gratings will be related to that field. Grating #1 stretches or contracts to a new wavelength w_1, and the grating lines are rotated so that the normal $\mathbf{n}_1$ is at an angle of $(\frac{\pi}{4} + \beta_1)$ radian to the X_1 axis. Grating #2 stretches or contracts to a new wavelength w_2, and these grating lines rotate so that the normal $\mathbf{n}_2$ is at an angle of $(\frac{3\pi}{4} + \beta_2)$ radian to the X_1 axis. See Figures 185 and 186. Figure 185 shows the overlap of the crossed gratings, and Figure 186 shows the geometry of the two gratings separately.

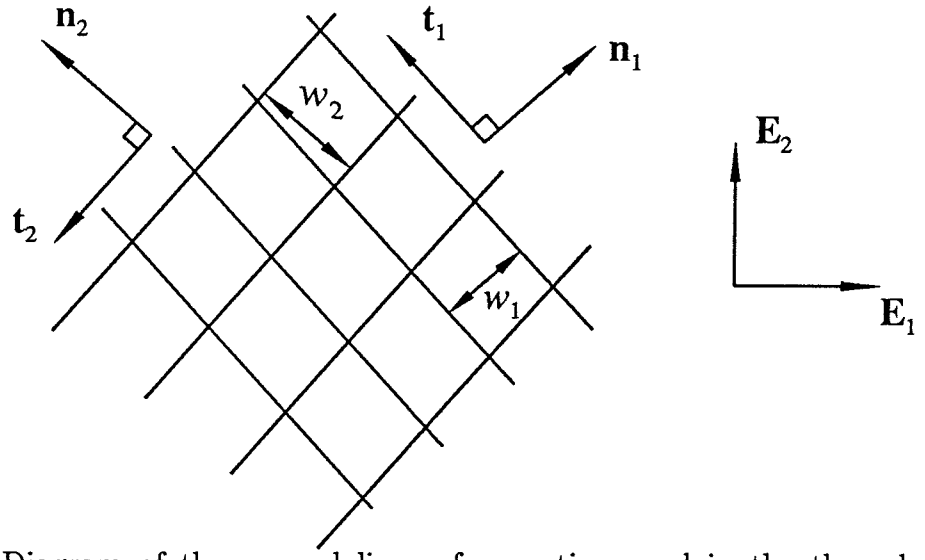


FIGURE 185. Diagram of the crossed lines of a grating used in the three–beam interferometer. Vector $\mathbf{n}_1$ is normal to ridges of grating #1, which are spaced w_1 apart. Vector $\mathbf{n}_2$ is normal to ridges of grating #2, which are spaced w_2 apart. Unit vectors $\mathbf{E}_1$ and $\mathbf{E}_2$ lie along the globally fixed coordinate axes.

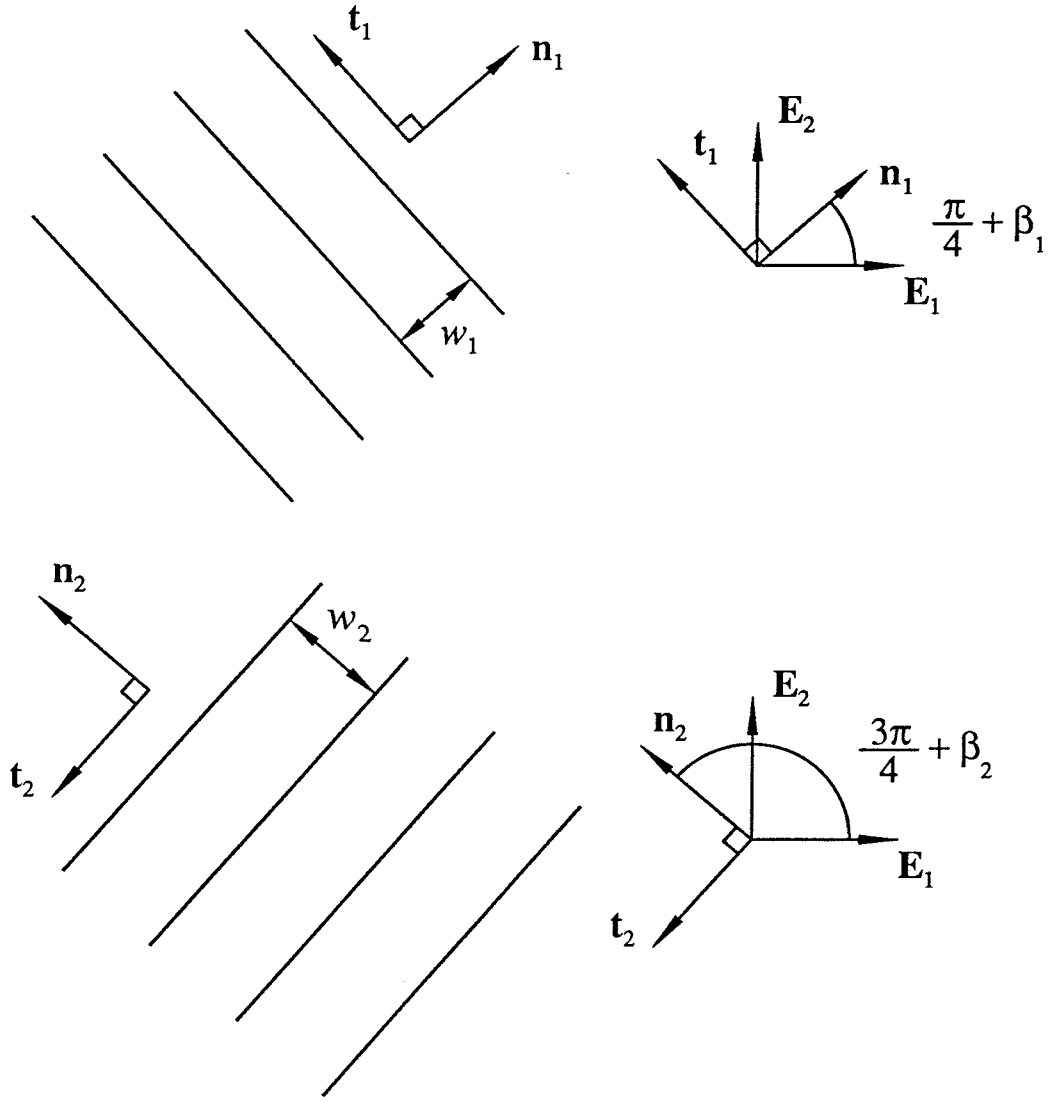


FIGURE 186. Diagram showing the geometry of gratings #1 and #2 separately. Grating #1 has wavelength w_1. Unit vector $\mathbf{n}_1$ is normal to the ridges of grating #1, and unit vector $\mathbf{t}_1$ is parallel to the ridges of grating #1. The angle between the globally fixed unit vector $\mathbf{E}_1$ and the unit vector $\mathbf{n}_1$ is $\frac{\pi}{4} + \beta_1$. Grating #2 has wavelength w_2. Unit vector $\mathbf{n}_2$ is normal to the ridges of grating #2, and unit vector $\mathbf{t}_2$ is parallel to the ridges of grating #2. The angle between the globally fixed unit vector $\mathbf{E}_1$ and the unit vector $\mathbf{n}_2$ is $\frac{3\pi}{4} + \beta_2$.

A.1.2 Equations of Diffraction

To calculate the direction of propagation of diffracted plane waves of any diffraction order from incident, plane light waves that are directed at the grating from an arbitrary angle, the unit normal to the incoming light waves must be resolved into components with respect to the coordinate frame defined by the ridges of the grating. For instance, if the unit vector $\mathbf{j}$ represents the normal of the incident plane waves and the unit vector $\mathbf{u}^{(n)}$ represents the normal to the diffracted plane waves of order n, then the equations relating the components of $\mathbf{j}$ to the components of $\mathbf{u}^{(n)}$ are:

$$\begin{aligned} u_1^{(n)} &= \frac{n\lambda}{w} + j_1 \\ u_2^{(n)} &= j_2 \end{aligned} \qquad (A.1.1)$$

where j_1 and $u_1^{(n)}$, are the components of $\mathbf{j}$ and $\mathbf{u}^{(n)}$ that lie in the grating plane and are normal to the ridges of the grating, and j_2 and $u_2^{(n)}$, are the components of $\mathbf{j}$ and $\mathbf{u}^{(n)}$ that lie in the grating plane and are parallel to the ridges of the grating. The diffraction order is n, and the light has wavelength λ.

The component of the unit vector $\mathbf{u}^{(n)}$ that is normal to the grating plane, namely, $u_3^{(n)}$, is nonnegative, and since $\mathbf{u}^{(n)}$ is a unit vector, equations (A.1.1) uniquely determine $\mathbf{u}^{(n)}$ from any given $\mathbf{j}$. Since the waves defined by $\mathbf{j}$ are directed toward the grating, the component j_3 is negative.

Let the diffracted light waves of order -1 from grating #1 be represented by the unit normal $\mathbf{u}$, and let the diffracted light waves of order -1 from grating #2 be represented by unit normal $\mathbf{v}$. Let $\mathbf{u}$ have components u_1, u_2 and U_3 in the $\mathbf{n}_1\mathbf{t}_1\mathbf{E}_3$ coordinate system and components U_1, U_2 and U_3 in the $\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3$ coordinate system. Let $\mathbf{v}$ have components v_1, v_2 and V_3 in the $\mathbf{n}_2\mathbf{t}_2\mathbf{E}_3$ coordinate system and components V_1, V_2 and V_3 in the $\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3$ coordinate system.

The fixed, unit normal vector representing the incident light ray to grating #1 is denoted $\mathbf{j}$, and it has components J_i $(i = 1, 2, 3)$ in the $\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3$ coordinate system

given by

$$\begin{aligned} J_1 &= \sin\alpha \cos\frac{\pi}{4} \\ J_2 &= \sin\alpha \sin\frac{\pi}{4} \\ J_3 &= -\cos\alpha \end{aligned} \tag{A.1.2}$$

In the coordinate frame defined by the deformed grating #1, the $\mathbf{n}_1\mathbf{t}_1\mathbf{E}_3$ coordinate system, which is at an angle of $\frac{\pi}{4}$ to the fixed coordinate system, the vector $\mathbf{j}$ has components j_1, j_2 and J_3 given by

$$\begin{aligned} j_1 &= \sin\alpha \cos\beta_1 \\ j_2 &= -\sin\alpha \sin\beta_1 \\ J_3 &= -\cos\alpha \end{aligned} \tag{A.1.3}$$

The vector $\mathbf{u}$ is the unit normal to the diffracted waves of order -1 from grating #1. From equations (A.1.1), the components of $\mathbf{u}$ in the grating frame are

$$\begin{aligned} u_1 &= \frac{-\lambda}{w_1} + \sin\alpha \cos\beta_1 \\ u_2 &= -\sin\alpha \sin\beta_1 \\ U_3 &= \left[1 - (u_1)^2 - (u_2)^2\right]^{\frac{1}{2}} \end{aligned} \tag{A.1.4}$$

In the globally fixed coordinate frame, the components of $\mathbf{u}$ are found as

$$\begin{aligned} U_1 &= \frac{1}{\sqrt{2}}\left[\frac{-\lambda}{w_1}(\cos\beta_1 - \sin\beta_1) + \sin\alpha\right] \\ U_2 &= \frac{1}{\sqrt{2}}\left[\frac{-\lambda}{w_1}(\cos\beta_1 + \sin\beta_1) + \sin\alpha\right] \\ U_3 &= \left[1 - (U_1)^2 - (U_2)^2\right]^{\frac{1}{2}} \end{aligned} \tag{A.1.5}$$

Similarly, the fixed, unit normal vector representing the incident light ray to grating #2 is denoted $\mathbf{k}$, and it has components K_i $(i = 1, 2, 3)$ in the $\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3$ coordinate system given by

$$\begin{aligned} K_1 &= \sin\alpha \cos\frac{3\pi}{4} \\ K_2 &= \sin\alpha \sin\frac{3\pi}{4} \\ K_3 &= -\cos\alpha \end{aligned} \tag{A.1.6}$$

In the coordinate frame defined by the deformed grating #2, the $\mathbf{n_2 t_2 E_3}$ coordinate system, which is at an angle of $\frac{3\pi}{4}$ to the fixed system, the vector $\mathbf{k}$ has components k_1, k_2 and K_3 given by

$$\begin{aligned} k_1 &= \sin\alpha\cos\beta_2 \\ k_2 &= -\sin\alpha\sin\beta_2 \\ K_3 &= -\cos\alpha \end{aligned} \tag{A.1.7}$$

The vector $\mathbf{v}$ is the unit normal to the diffracted waves of order -1 from grating #2. From equations (A.1.1), the components of $\mathbf{v}$ in the grating frame are

$$\begin{aligned} v_1 &= \frac{-\lambda}{w_2} + \sin\alpha\cos\beta_2 \\ v_2 &= -\sin\alpha\sin\beta_2 \\ V_3 &= \left[1-(v_1)^2-(v_2)^2\right]^{\frac{1}{2}} \end{aligned} \tag{A.1.8}$$

In the globally fixed coordinate frame, the components of $\mathbf{v}$ are found as

$$\begin{aligned} V_1 &= \frac{1}{\sqrt{2}}\left[\frac{\lambda}{w_2}(\cos\beta_2+\sin\beta_2)-\sin\alpha\right] \\ V_2 &= \frac{1}{\sqrt{2}}\left[\frac{-\lambda}{w_2}(\cos\beta_2-\sin\beta_2)+\sin\alpha\right] \\ V_3 &= \left[1-(V_1)^2-(V_2)^2\right]^{\frac{1}{2}} \end{aligned} \tag{A.1.9}$$

A.1.3 Initial Alignment of the Laser Beams

Initially, *i.e.*, before any deformation occurs, the incoming laser beams must be aligned so that the diffracted beams are parallel to the X_3 axis. The initial conditions are that $\beta_1 = \beta_2 = 0$, $w_1 = w_2 = w_0$ and $u_1 = u_2 = v_1 = v_3 = 0$. This leads to the conclusion that

$$\sin\alpha = \frac{\lambda}{w_0} \tag{A.1.10}$$

This relation does not change during the deformation.

Making the substitution from equation (A.1.10) into equations (A.1.5) and (A.1.9) yields

$$\begin{aligned} U_1 &= \frac{1}{\sqrt{2}}\left[\frac{-\lambda}{w_1}(\cos\beta_1 - \sin\beta_1) + \frac{\lambda}{w_0}\right] \\ U_2 &= \frac{1}{\sqrt{2}}\left[\frac{-\lambda}{w_1}(\cos\beta_1 + \sin\beta_1) + \frac{\lambda}{w_0}\right] \\ U_3 &= \left[1 - (U_1)^2 - (U_2)^2\right]^{\frac{1}{2}} \end{aligned} \qquad (A.1.11)$$

$$\begin{aligned} V_1 &= \frac{1}{\sqrt{2}}\left[\frac{\lambda}{w_2}(\cos\beta_2 + \sin\beta_2) - \frac{\lambda}{w_0}\right] \\ V_2 &= \frac{1}{\sqrt{2}}\left[\frac{-\lambda}{w_2}(\cos\beta_2 - \sin\beta_2) + \frac{\lambda}{w_0}\right] \\ V_3 &= \left[1 - (V_1)^2 - (V_2)^2\right]^{\frac{1}{2}} \end{aligned} \qquad (A.1.12)$$

A.1.4 Observed Fringe Spacing and Inclination

As described in Section 2.3, the observed fringe spacing, δ^*, and the angle between the fringes and the X_2 axis, ϕ^*, are determined from the vector difference $\mathbf{z} = \mathbf{u} - \mathbf{v}$. These relations are found to be

$$\delta^* = \frac{\lambda}{[\mathbf{z}\cdot\mathbf{z} - (\mathbf{z}\cdot\mathbf{E}_3)^2]^{\frac{1}{2}}} \qquad (2.3.38)$$

$$\tan\phi^* = \frac{\mathbf{z}\cdot\mathbf{E}_2}{\mathbf{z}\cdot\mathbf{E}_1} \qquad (2.3.39)$$

These equations indicate that only the components Z_1 and Z_2 in the globally fixed $\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3$ coordinate system are needed to find the apparent fringe spacing and inclination. In this case, δ^* and ϕ^* are found from

$$\delta^* = \frac{\lambda}{\left[(Z_1)^2 + (Z_2)^2\right]^{\frac{1}{2}}} \qquad (A.1.13)$$

$$\tan\phi^* = \frac{Z_2}{Z_1} \qquad (A.1.14)$$

From equations (A.1.11) and (A.1.12) the components Z_1 and Z_2 are found as

$$Z_1 = U_1 - V_1$$

$$= \frac{1}{\sqrt{2}}\left[\frac{-\lambda}{w_1}(\cos\beta_1 - \sin\beta_1) - \frac{\lambda}{w_2}(\cos\beta_2 + \sin\beta_2) + 2\frac{\lambda}{w_0}\right] \tag{A.1.15}$$

$$Z_2 = U_2 - V_2$$

$$= \frac{1}{\sqrt{2}}\left[\frac{-\lambda}{w_1}(\cos\beta_1 + \sin\beta_1) + \frac{\lambda}{w_2}(\cos\beta_2 - \sin\beta_2)\right]$$

A.1.5 The Third Grating

It is proposed that the interference pattern produced by the light diffracted from gratings #1 and #2, originally with lines at ±45° to the X_1 axis as described above, is equivalent to an interference pattern that is due to the deformation of a third grating (grating #3), which originally had lines parallel to the X_2 axis. From Figure 187, it can be seen that the peaks and valleys of a pair of crossed gratings at ±45° to the X_1 axis with wavelength w_0 make peaks and valleys for a grating at 90° to the X_1 axis with a wavelength of $w_0/\sqrt{2}$.

As seen in Figure 188, and from Durelli and Parks [17], or Theocaris [85], it can be shown that after deformation, grating #3 will have a wavelength w that is related to the wavelengths of the deformed gratings #1 and #2 (originally at ±45°) by the equation:

$$w = \frac{w_1 w_2}{[w_1^2 + w_2^2 - 2w_1 w_2 \cos\theta]^{\frac{1}{2}}} \tag{A.1.16}$$

where w_1 is the wavelength of grating #1, w_2 is the wavelength of grating #2, and θ is the angle between the ridges of grating #1 and the ridges of grating #2. The angle between the ridges of grating #1 and the ridges of grating #3 is denoted

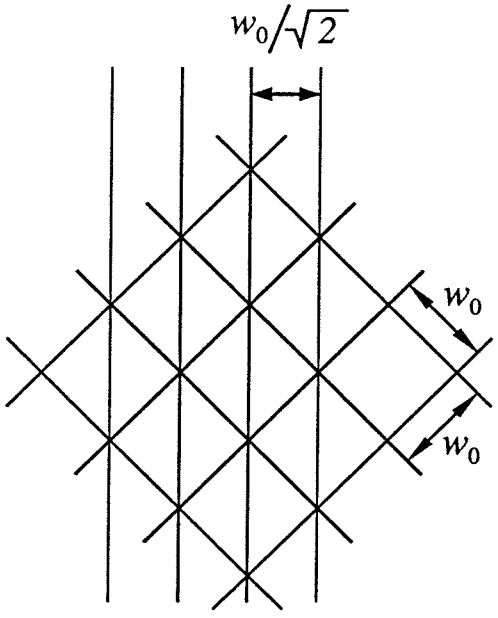


FIGURE 187. The third grating before deformation. The ridges and valleys of gratings that cross at $\pm 45°$ to the X_1 axis and that have identical wavelengths of w_0 form a third grating that has a wavelength of $w_0/\sqrt{2}$.

ψ, and the trigonometric functions of $\sin\psi$ and $\cos\psi$ are related to the variables associated with gratings #1 and #2 by the equations:

$$\sin\psi = \frac{w_1 \sin\theta}{[w_1^2 + w_2^2 - 2w_1 w_2 \cos\theta]^{\frac{1}{2}}} = \left(\frac{w \sin\theta}{w_2}\right) \qquad (A.1.17)$$

$$\cos\psi = \frac{w_1 \cos\theta - w_2}{[w_1^2 + w_2^2 - 2w_1 w_2 \cos\theta]^{\frac{1}{2}}} = \left(\frac{w \cos\theta}{w_2} - \frac{w}{w_1}\right) \qquad (A.1.18)$$

The angle θ can be calculated from the angles β_1 and β_2. The ridges of grating #1 lie at an angle of $(\frac{\pi}{4} + \beta_1)$ radian to the X_2 axis, and the ridges of grating #2 lie at an angle of $(\frac{3\pi}{4} + \beta_2)$ radian to the X_2 axis. The angle θ is then found as

$$\begin{aligned} \theta &= \left(\frac{3\pi}{4} + \beta_2\right) - \left(\frac{\pi}{4} + \beta_1\right) \\ &= \left(\frac{\pi}{2} + \beta_2 - \beta_1\right) \end{aligned} \qquad (A.1.19)$$

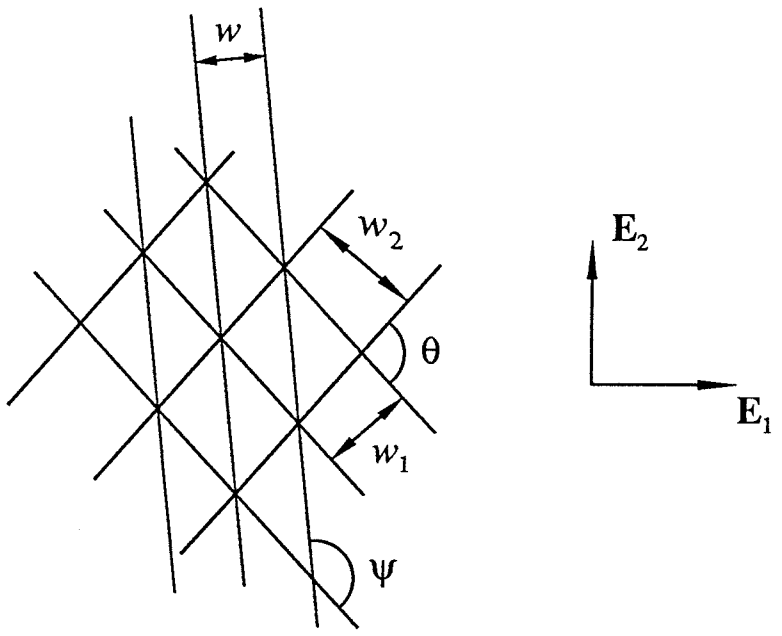


FIGURE 188. The geometry of the third grating after deformation. Grating #1 has wavelength w_1, grating #2 has wavelength w_2, and the angle between the ridges of grating #1 and grating #2 is θ. The third grating has wavelength w, and the angle between the ridges of grating #1 and grating #3 is ψ.

A.1.6 Equations Describing the Fringes in Terms of the Third Grating

In order to show the correspondence between the equations describing the fringes observed with the three–beam interferometer and the equations describing the fringes observed using geometric moire methods, it is necessary to write the equations for the fringe spacing δ^* and the fringe inclination ϕ^*, equations (A.1.13) and (A.1.14), in terms of variables associated with grating #3. These variables are the initial wavelength of grating #3, $\widetilde{w}_0$, the deformed wavelength of grating #3, w, and the angle β, which is the angle between the ridges of grating #3 and the X_2 axis.

The initial wavelength of grating #3, the wavelength of the grating before any deformation occurs, is denoted $\widetilde{w}_0$. The value of $\widetilde{w}_0$ can be related to the initial conditions of gratings #1 and #2. From the geometry shown in Figure 187, or from equation (A.1.16) using the initial conditions of gratings #1 and #2, that

$w_1 = w_2 = w_0$ and $\theta = \pi$, the initial wavelength for grating #3 is found as

$$\widetilde{w}_0 = \frac{w_0}{\sqrt{2}} \qquad (A.1.20)$$

The angle β, which is the angle between the ridges of the third grating and the X_2 axis, is found from the angle ψ between the ridges of grating #1 and the ridges of grating #3 and the angle β_1 between the ridges of grating #1 and the X_1 axis as

$$\beta = \psi - \frac{\pi}{2} - \left(\frac{\pi}{4} - \beta_1\right) \quad = \psi + \beta_1 - \frac{3\pi}{4} \qquad (A.1.21)$$

From the geometry relating the three gratings, it can then be found that

$$\cos\beta = \frac{1}{\sqrt{2}}\left(\frac{w}{w_1}(\cos\beta_1 - \sin\beta_1) + \frac{w}{w_2}(\cos\beta_2 + \sin\beta_2)\right)$$

$$\sin\beta = \frac{1}{\sqrt{2}}\left(\frac{w}{w_1}(\cos\beta_1 + \sin\beta_1) - \frac{w}{w_2}(\cos\beta_2 - \sin\beta_2)\right) \qquad (A.1.22)$$

The equations for the fringe spacing δ^* and inclination ϕ^*, equations (A.1.13) and (A.1.14), can now be rewritten in terms of variables associated with grating #3: w, $\widetilde{w}_0$ and β. Comparing equations (A.1.15) with equations (A.1.22), it is found that

$$Z_1 = \quad U_1 - V_1 \quad = \frac{\lambda}{\widetilde{w}_0} - \frac{\lambda}{w}\cos\beta \qquad (A.1.23)$$

$$Z_2 = \quad U_2 - V_2 \quad = \frac{-\lambda}{w}\sin\beta \qquad (A.1.24)$$

Therefore,

$$Z_1^2 + Z_2^2 = \frac{\lambda^2}{\widetilde{w}_0^2} + \frac{\lambda^2}{w^2} - \frac{2\lambda^2}{w\widetilde{w}_0}\cos\beta \qquad (A.1.25)$$

Finally, the equations for the observed fringe spacing δ^* and the fringe inclination ϕ^* are found as

$$\delta^* = \frac{\lambda}{[Z_1^2 + Z_2^2]^{\frac{1}{2}}} \quad = \frac{w\widetilde{w}_0}{[w^2 + \widetilde{w}_0^2 - 2w\widetilde{w}_0\cos\beta]^{\frac{1}{2}}} \qquad (A.1.26)$$

$$\tan\phi^* = \frac{Z_2}{Z_1} \quad = \frac{\widetilde{w}_0\sin\beta}{\widetilde{w}_0\cos\beta - w} \qquad (A.1.27)$$

A.1.7 Discussion

Equations (A.1.26) and (A.1.27) are exactly the equations for the fringes to represent contours of constant displacement. The measured displacement is parallel to the X_1 axis with an increment of displacement between fringes of $\widetilde{w}_0$, as shown in Durelli and Parks [17] or Theocaris [85]. The equations are identical to the equations for the four–beam interferometer under planar deformation except for a multiplying factor of 2 in the fringe spacing δ^*. This factor can be thought of as being due to the spacing of the virtual reference grating formed by the interference between the two incoming laser beams, as discussed in Section 2.2. With the four–beam interferometer, the incoming laser beams intersect at an angle of 2α, creating an interference pattern spaced at $w_0/2$, which is half the initial wavelength of the grating used of measurement. In the three–beam interferometer, the incoming laser beams intersect at an angle such that the interference pattern is spaced at $\widetilde{w}_0$, which is exactly the same as the initial wavelength of the grating being measured. Fringe multiplication is described in Section 2.2, and more detail can be found in the paper by Post [59].

Though the three–beam interferometer will measure the same quantities as the four–beam interferometer, the disadvantages of the three–beam interferometer as shown in Figure 182 outweigh its advantages compared to the four–beam interferometer shown in Figure 183. In order to produce fringe patterns of high contrast, one requirement is that the laser beams follow paths of similar length to maintain the coherence between the two interfering light wave fronts. Both the three–beam and four–beam interferometers meet this requirement. However, high–contrast fringe patterns also require that the two interfering beams have similar amplitudes and are polarized similarly. Here the three–beam interferometer falls short (at least in the configuration shown in the figure). At the first beam splitter in the path of the laser, the light is divided approximately in half, with half of the light going to one of the mirrors and then to the specimen, and half of the light going to the next beam splitter. At the second beam splitter, the remaining light is again divided approximately in half, with each beam passing to a mirror and then to the specimen. Thus, one of the beams is approximately half of the strength of the original laser

beam, while two of the beams are one quarter of the strength of the original laser beam. In order to match the amplitude of the beams, an additional beam splitter would have to be inserted into the path of the beam that encounters only one beam splitter. (Or the mirror directing the beam to the specimen could be replaced with a half–reflecting mirror.) The intensity of each of the beams would then be one quarter the original laser beam, in contrast to the four–beam interferometer, where each of the beams is one-half the intensity of the original laser beam. The lower light levels in the three–beam case would lead to longer exposure times for the fringe recording system, which makes the experiment more sensitive to vibrations. Also, if an additional beam splitter is used, it reduces the cost and material advantage of the three–beam interferometer over the four–beam interferometer.

The polarization of the diffracted beams must also be similar for the interference patterns to have high contrast. The three–beam interferometer does not match the polarization of the interfering laser beams as well as the four–beam interferometer. The polarization mismatch can be improved with a rotating polarizing filter on the recording camera, which forces the diffracted beams to be similarly polarized, and can be adjusted to find the best polarization angle at which this occurs. However, the addition of a polarizing filter also reduces the available light, increasing exposure time and the sensitivity to vibrations. The polarization mismatch might also be improved by a rotation of the laser mounting about the axis along the laser beam.

Another difficulty with the three–beam interferometer when compared with the four–beam interferometer is that it will lose information at lower displacement gradient levels. For the same fringe spacing, the diffracted light deviates from the normal to the specimen surface at a larger angle for the three–beam interferometer than for the four–beam interferometer. This is because the three–beam interferometer uses light diffracted from two different gratings to infer displacement information about a third grating. Stretching of the two diffracting gratings can occur with only part of the resulting angular deflection of the diffracted light being related to the stretch of the third grating.

APPENDIX B

Elliptical Polarization

B.1 The Interference of Elliptically Polarized Plane Harmonic Monochromatic Light Waves

In Section 2.3.1, it was assumed that the interfering light waves are linearly polarized. Linear polarization is a special case of the more general state of elliptic polarization. For elliptically polarized waves, the electric–field vector may be thought of as rotating as it passes through space, its tip tracing an ellipse when viewed along the direction of propagation of the plane waves. In linearly polarized light, the electric–field vector always remains in a plane, and its tip draws a line when viewed along the direction of propagation of the plane waves. Linear polarization occurs when the ellipse collapses to a straight line. (Circular polarization is, of course, also a special case of elliptical polarization.)

If the interfering waves are assumed to be elliptically polarized, the result of Section 2.3.3 is still found to hold, but the contrast of the fringes may be reduced if the polarization of the two sets of waves is out of phase. The spatial coherence of the waves is still assumed to hold, so that the polarization of one set of waves is the same across the wave front, and the waves are still assumed to be strictly monochromatic.

Returning to the electric–field vector of equation (2.3.26), which describes harmonic plane waves traveling in the direction of unit vector $\mathbf{u}$, the electric field is given by

$$\boldsymbol{\mathcal{E}}^{(\mathrm{u})} = Re\left\{ \mathbf{A}^{(\mathrm{u})} \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{u} - ct) + i\phi^{(\mathrm{u})} \right] \right\} \qquad (2.3.26)$$

where the vector amplitude $\mathbf{A}^{(u)}$ is a complex vector. The effect of elliptical polarization is more easily seen if the electric–field vector is resolved into components with respect to some basis. If we choose $\mathbf{u}$ to be one of the basis vectors with $\mathcal{E}_3^{(u)}$ to be the component along $\mathbf{u}$, then $\mathcal{E}_3^{(u)} = 0$ for plane waves, since the electric–field vector is always orthogonal to the direction of propagation. The remaining two components, $\mathcal{E}_1^{(u)}$ and $\mathcal{E}_2^{(u)}$, can be written as

$$\begin{aligned}
\mathcal{E}_1^{(u)} &= Re\left\{ A_1^{(u)} \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{u} - ct) + i\phi^{(u)}\right]\right\} \\
&= R_1^{(u)}\ Re\left\{ \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{u} - ct) + i\phi^{(u)} + i\psi_1^{(u)}\right]\right\} \\
\mathcal{E}_2^{(u)} &= Re\left\{ A_2^{(u)} \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{u} - ct) + i\phi^{(u)}\right]\right\} \\
&= R_2^{(u)}\ Re\left\{ \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{u} - ct) + i\phi^{(u)} + i\psi_2^{(u)}\right]\right\}
\end{aligned} \qquad (B.1.1)$$

where the complex numbers $A_1^{(u)}$ and $A_1^{(u)}$ have been replaced using

$$A_k^{(u)} = R_k^{(u)} \exp\left(i\psi_k^{(u)}\right), \qquad k = 1, 2 \qquad (B.1.2)$$

with $R_k^{(u)}$ and $\psi_k^{(u)}$ real–valued constants. It is seen from equation (B.1.1) that elliptical polarization causes the two components of $\mathcal{E}^{(u)}$ to have a phase difference between them given by $\psi_2^{(u)} - \psi_1^{(u)}$. This phase difference will have an effect on the interference between the plane waves described in equation (2.3.15). The electric–field vector for the plane waves traveling in the direction of unit vector $\mathbf{v}$ are described as in equation (B.1.1) by

$$\begin{aligned}
\mathcal{E}_1^{(v)} &= R_1^{(v)}\ Re\left\{ \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{v} - ct) + i\phi^{(v)} + i\psi_1^{(v)}\right]\right\} \\
\mathcal{E}_2^{(v)} &= R_2^{(v)}\ Re\left\{ \exp\left[\frac{i2\pi}{\lambda}(\mathbf{r}\cdot\mathbf{v} - ct) + i\phi^{(v)} + i\psi_2^{(v)}\right]\right\}
\end{aligned} \qquad (B.1.3)$$

Since we are interested in the inner product between vectors, and since $\mathcal{E}_3^{(\mathrm{u})} = 0$, we can neglect $\mathcal{E}_3^{(\mathrm{v})}$, although since $\mathbf{u}$ and $\mathbf{v}$ are not in general coincident, $\mathcal{E}_3^{(\mathrm{v})}$ will not necessarily be zero in this coordinate frame. It *would* be zero in a coordinate frame where $\mathcal{E}_3^{(\mathrm{v})}$ is the component along $\mathbf{v}$.

The intensity distribution can be determined using equation (2.3.9), so that

$$\begin{aligned} 2\mathrm{I}^{(total)} = {} & |\mathbf{A}^{(\mathrm{u})}|^2 + |\mathbf{A}^{(\mathrm{v})}|^2 \\ & + 2R_1^{(\mathrm{u})}R_1^{(\mathrm{v})}\cos\left[\frac{2\pi}{\lambda}\left(\mathbf{r}\cdot(\mathbf{u}-\mathbf{v})\right) + \left(\phi^{(\mathrm{u})} - \phi^{(\mathrm{v})}\right) + \left(\psi_1^{(\mathrm{u})} - \psi_1^{(\mathrm{v})}\right)\right] \\ & + 2R_2^{(\mathrm{u})}R_2^{(\mathrm{v})}\cos\left[\frac{2\pi}{\lambda}\left(\mathbf{r}\cdot(\mathbf{u}-\mathbf{v})\right) + \left(\phi^{(\mathrm{u})} - \phi^{(\mathrm{v})}\right) + \left(\psi_2^{(\mathrm{u})} - \psi_2^{(\mathrm{v})}\right)\right] \end{aligned} \tag{B.1.4}$$

Again, the first two terms in equation (B.1.4) represent the intensities of each set of plane waves individually, and the third and fourth terms represent the interference terms.

It can be seen that the third term in equation (B.1.4) is out of phase with the fourth term in equation (B.1.4) by the relative amount of $\Delta\psi = \left(\psi_2^{(\mathrm{u})} - \psi_2^{(\mathrm{v})}\right) - \left(\psi_1^{(\mathrm{u})} - \psi_1^{(\mathrm{v})}\right)$. If $\Delta\psi$ is zero, which may be possible if the waves originate from the same source and then traverse similar routes to the place where they interfere, the interference terms will reinforce each other, resulting in fringes of high contrast. On the other hand, if $\Delta\psi$ approaches π, the interference terms will act to cancel each other out, depending on the relative magnitude of the amplitude terms, and the fringes will have lower contrast.

As mentioned, the laser beams used in the experiment are strongly linearly polarized, so the interference in the experiment is well described by equation (2.3.15). However, linear polarization is a special case that may not always hold. Should the light be elliptically polarized, or should the light be linearly polarized with the electric–field vectors in dissimilar directions, the fringe contrast can be improved by the addition of a polarizing filter, which forces the light to be linearly polarized so that the electric–field vectors point in a single direction.

The most important point brought out by equation (B.1.4) is that the fringes, the loci of constant phase difference between the plane waves, will still be planes

perpendicular to the vector difference $\mathbf{z} = \mathbf{u} - \mathbf{v}$, and they will still have a spacing δ given by $\delta = \lambda/|\mathbf{z}|$, regardless of the state of the polarization of the laser beams.