Structural Simulations Using Multi-resolution Material Models

Daniel C. Hammerand, Samuel W. Key, J. Tinsley Oden, Ivo Babuska, Greg Rodin, Chandrajit Bajaj, Serge Prodhomme, and Kuman Vemaganti

Sandia National Laboratories

Structural Simulations Using Multi-resolution Material Models

Daniel C. Hammerand

Material Mechanics

Samuel W. Key

Computational Solid Mechanics/Structural Dynamics

Sandia National Laboratories

P.O. Box 5800

Albuquerque, NM 87185-0847

J.T. Oden, I. Babuška, G. Rodin, C. Bajaj, S. Prudhomme, and K. Vemaganti

Texas Institute for Computational and Applied Mechanics

The University of Texas at Austin

Austin, TX 78712

ABSTRACT

This report summarizes work on the project "Structural Simulation Using Multi-resolution Material Models" which was a joint LDRD between Sandia and TICAM (Texas Institute for Computational and Applied Mathematics). The project focused on the development of new methods and procedures for developing high-fidelity models of material response and new ideas for assessing modeling error, as well as automatic model adaptation. Major issues studied in this effort included (1) the implementation of Fast-Multipole Methods and Fictitious Domain Methods in the analysis of heterogeneous materials; (2) experimental characterization of epoxy filled with glass beads; (3) estimation of modeling error and adaptive modeling; (4) the treatment of uncertainties in determining properties of models from imaging data; and (5) the use of CT images of heterogeneous materials to automatically generate quadrilateral and hexahedral meshes.

The details of the LDRD supported work on Fast-Multipole Methods and Fictitious Domain Methods are not given in this report. However, much of this work is summarized in [1]. This report

consists of four chapters corresponding to items (2)-(5) listed above.

Complementary portions of the work, also reported here so as to present a more complete picture of the ideas and results, were done at TICAM under contracts from ONR and NSF. In particular, the material collected in Chapter 2 was jointly supported by SNL and ONR and the work presented in Chapter 3 was jointly supported by SNL and NSF.

<u>REFERENCES</u>

[1] Podnos, E.G., Application of Fictitious Domain Method to Analysis of Composite Materials, Ph.D. Dissertation, University of Texas at Austin, 1999.

<u>ACKNOWLEDGMENTS</u>

CONTENTS

CHAPTER 1

Experimental Characterization of Glass-Epoxy Specimens

Greg Rodin

1. Introduction

Characterization of the relationships between the engineering (macroscopic) and micro-structural (microscopic) properties is a central problem in materials science and engineering. Solution of this problem is difficult from both experimental and modeling perspectives, because it may involve massive amounts of data pertaining to the microstructure. Of course, some materials and phenomena may require rather simple experimental data and models. For example, elastic properties of particle-reinforced composites can be accurately predicted using simple models that require only one microstructural parameter -- the volume fraction [1]. In contrast, the prediction of plastic and fracture properties of fiber-reinforced composites may require data pertaining to the microstructural architecture and fiber waviness. In general, the required microstructural data are impossible to define *a priori*, and therefore one may have to go through a long trial-and-error process before obtaining satisfactory results.

With the emergence of three-dimensional imaging methods, such as computerized tomography (CT), and continued growth of available computing power, it becomes possible to characterize the relationships between the microscopic and macroscopic properties without making crude assumptions about the microstructural geometry. However, it would be naïve to believe that three-dimensional imaging methods dramatically simplify the task of material characterization based on microstructural properties. To the contrary, due to non-conventional ways of representing the microstructural geometry, the use of three-dimensional imaging methods gives rise to new challenging problems. Of course, one hopes that by resolving those problems one would be able to bring material characterization to a significantly better level.

9

The principal objective of this work is to learn how to use CT data for mechanical characterization of heterogeneous materials. Toward this objective we considered several glass-epoxy specimens whose microstructural geometry was quantified using CT and optical microscopy (OM), and weighing; OM and weighing data were used for benchmarking CT data. In addition, the specimens were tested under low-stress compression in order to determine their elastic properties and high-stress compression in order to examine their fracture properties.

Eventually, results of this work will be used for benchmarking and improvement of computational modeling and CT data acquisition methods. To this end, it was critical to work with sufficiently simple specimens that lend themselves to detailed and verifiable computational modeling.

Presently, material characterization based on CT data is at its infancy. One can use CT data for observational purposes only, for identification of flaws and other imperfections. This approach has been pursued by Yancey and co-workers [2-6] and Bossi and coworkers [7], who used CT for detection of cracks and microscopic geometry in fiber-reinforced composites, and identified possible applications of CT imaging in manufacturing. Kikuchi and co-workers [8-10] utilized CT data for predictions of mechanical behavior, by integrating CT data with finite element and homogenization methods. The approach that appeals to us the most, is the one due to Huet [11] who integrated CT data with finite element computations and mechanical testing of concrete. The principal difference between our approach and that of Huet [11] is that in his analysis CT data is used to define a set of spheres that approximate the actual microstructure, and we rely on CT data alone, without making any assumptions about the microstructural geometry.

The remainder of this chapter is organized as follows. In Sections 2 and 3 we describe prototype specimens. Section 2 is dedicated to the description of the fabrication and testing procedures, and Section 3 is dedicated to the analysis of experimental results and identification of improvements in the fabrication and testing procedures. In Section 4, we describe the model specimens that were fabricated and tested using lessons learned from the prototype specimens. In Section 5, we summarize significant results of this study and identify directions for future work.

2. Prototype specimens: fabrication and testing procedures

In this section, we are concerned with prototype specimens. These specimens allowed us to identify proper procedures for manufacturing, imaging, and mechanical testing of the model specimen.

2.1 Overview

At the beginning of the study, it was decided to work with specimens whose microstructure can be characterized with both CT and OM imaging. The primary restriction imposed by CT imaging is that the microscopic dimensions must be sufficiently large. Those distances include the particle size, the inter-particle distances, and the particle-specimen-surface distances. Since the resolution of current CT devices is 10 μm at best, the particles must be several hundred micron in size, and they must be positioned not too close to each other and not too close to the specimen surface. The secondary restriction imposed by CT imaging is that the macroscopic dimensions must be sufficiently small. This restriction is necessary in order to be able to store and process the CT data. By assuming that each elementary cube of material requires 10 bytes and the resolution is 10 μm, the data density is estimated as 10^{16} bytes/m^3. Accordingly, a specimen whose volume is 1 cm^3 would require 10^9 bytes worth of memory. For current workstations, this number is comparable to the hard drive memory rather than RAM memory. OM imaging imposes two restrictions. First, the specimen must be optically transparent. Second, the microstructure must be essentially two-dimensional.

In this study, we satisfied the restrictions imposed by CT and OM by working with rectangular specimens made of epoxy matrix filled with glass beads. The glass beads were close to 880 μm in diameter and they were arranged as a monolayer. For the prototype specimens, additional specifications for the macroscopic and microscopic dimensions were obtained by trial-and-error. For the model specimen, additional specifications for the macroscopic and microscopic dimensions were obtained based on the experience gained with the prototype specimens.

We worked with four types of prototype specimens, to which we refer as Epoxy A, Epoxy B, Composite A, and Composite B. The principal difference between the Epoxy A and Epoxy B specimens was the exposure period to room temperature and humidity prior to mechanical testing. For the Epoxy A specimens the exposure period was about one month and for the Epoxy B specimens the exposure period was about one week. The exposure period was reduced due to time constraints rather than scientific reasons. The matrix of the specimens Composite A (Composite B) was produced following the fabrication procedures for the specimens Epoxy A (Epoxy B).

2.2 Fabrication Procedures

2.2.1 Epoxy Specimens

The Epoxy A and Epoxy B specimens were made of a resin (Araldite GY502) mixed with an amido amine hardener (HY955) at a weight ratio 5:2 as proposed in [12]. The resin and the hardener were mixed completely and then the mixture was placed in a vacuum chamber at room temperature for 30 minutes to eliminate the air bubbles inside the mixture. The epoxy was then poured into a mold and cured at 160 $^{\circ}$F for 60 minutes. Finally, both the top and bottom surfaces of the pure epoxy specimen were machined flat and then they were cut to the various-length portions. The specimen dimensions are summarized in Table 2.1.

Note that the Epoxy B specimens are much longer than the Epoxy A specimens. This is because the Epoxy A specimens failed to provide an adequate constraint against the displacements along the length direction.

Prototype Specimens	l mm	w mm	t mm
Epoxy A	3.18	12.90	1.88
Epoxy A	6.35	12.90	1.88
Epoxy A	12.90	12.90	1.88
Epoxy A	22.90	12.90	1.88
Epoxy B	7.72	12.90	1.88
Epoxy B	38.10	12.90	1.88
Epoxy B	75.90	12.90	1.88
Composite A	3.18	12.93	1.88
Composite A	6.35	12.93	1.88
Composite A	12.90	12.93	1.88
Composite A	22.86	12.93	1.88
Composite B	75.92	12.90	1.87

Table 2.1   Geometry of the prototype specimens

2.2.2 Glass Beads

A batch of glass beads was supplied to us by a local manufacturer. In order to select glass beads suitable for CT imaging, we sieved the entire batch such using an 838 μm sieve (Buckbee-Meats St. Paul MN-62). As a result we obtained a batch of glass beads whose size distribution is characterized by the frequency diagram shown in Figure 2.1. This diagram was obtained by picking one hundred spheres at random and measuring their diameter using an optical microscope. The mean diameter of this sample was 880 μm, which is sufficient for CT imaging.
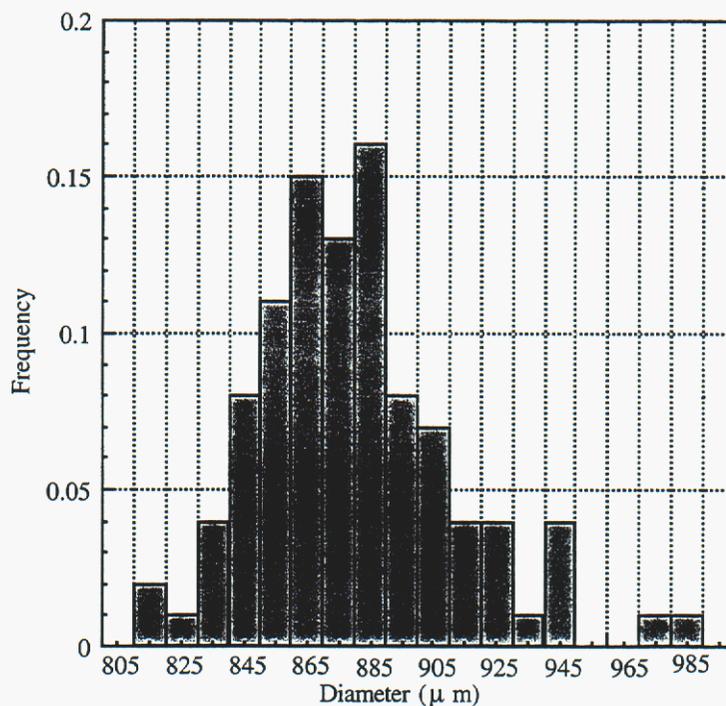
Figure 2.1   Histogram for one hundred glass beads after sieving

According to the literature [13], Young's modulus of glass does not vary significantly from one type of glass to another. Therefore, for glass beads, we adopted E=73 GPa (E-glass). Otherwise, we would have to determine the value of E using sub-millimeter glass beads, which is a non-trivial task.

2.2.3 Composite Specimens

Following the fabrication procedure for the Epoxy A and B specimens, the resin and the hardener were mixed completely and then the mixture was placed in a vacuum chamber at room temperature for 30 minutes to eliminate the air bubbles inside the mixture. The epoxy was then poured into a mold to make the first layer and cured at 160 °F for 60 minutes. After machining the surface of the specimen flat, tens of glass beads for Composite A and hundreds of glass beads for Composite B were placed at random on the first layer of epoxy. After that, more epoxy was poured over the glass-bead layer. The specimens were then cured for another 60 minutes at 160 °F. Finally, both the top and bottom surfaces of the specimens were machined to obtain the desired specimen thickness.

The Composite A specimens had the same dimensions as the Epoxy A specimens. These sets of specimens were used for evaluating the stiffening effect of the glass beads. Accordingly, these specimens were subjected to mechanical testing. In contrast, the Composite B specimen was used for evaluating the microstructural geometry. Accordingly, this specimen was subjected to CT, OM and weight measurements. The dimensions for the entire set of prototype specimens are summarized in Table 2.1.

2.3   Mechanical Testing

The microstructural geometry in the form of a monolayer requires one to work with thin specimens. Under simple tension or compression, such specimens are prone to instabilities even at low stress levels. Under simple tension, the instability is due to necking, and, under simple compression, the instability is due to buckling. To eliminate these instabilities, we decided to rely on constrained compression tests with loading in the thickness direction. Although this test does not involve instabilities, it requires careful evaluation of the in-plane constraints. To resolve this issue we conducted our tests using a plane strain compression device whose metal walls provide the constraint in one in-plane direction. In the other in-plane direction, the constraint was realized by choosing sufficiently long specimens.

The plane strain test device used in this study is shown in Figures 2.3. With this device, the specimens were loaded with a stiff metal compression bar at a rate of 1.27×10^{-2} mm/s. The force was measured with an 8.90 kN load cell. To ensure uniformity of the applied pressure on the specimen, a viscous compression platten was used. Additionally, two spring-loaded LVDTs were placed at both ends of the compression bar to measure the relative movement as shown in Figure 2.3. The alignment of the compression bar was easily confirmed from the two LVDT measurements. The surface of the compression bar was smooth enough, so that no

14
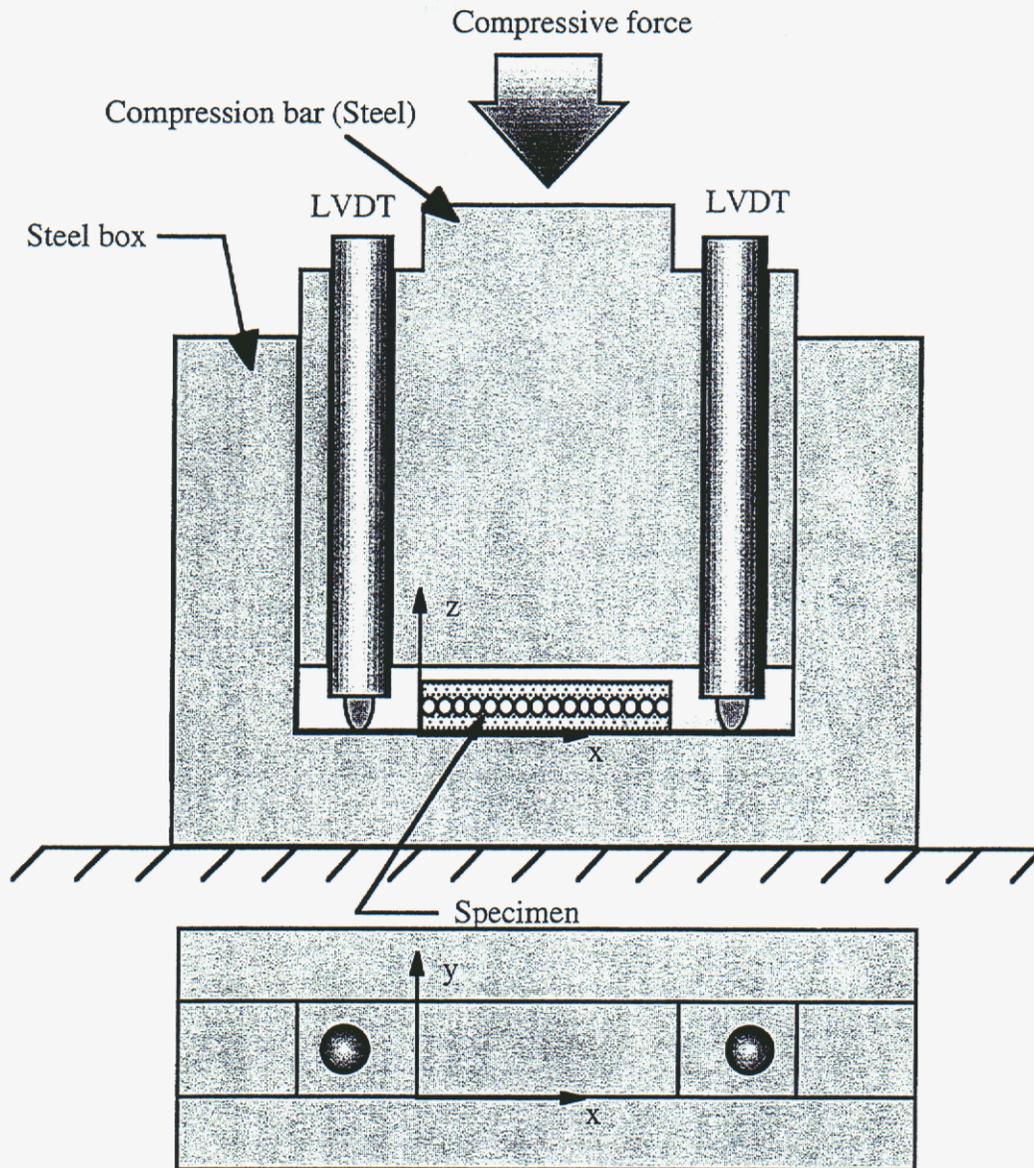
Figure 2.3   Schematic of the interior of the plane strain compression device

lubricant was required. Compression tests were conducted on the Epoxy A and B and

Composite A specimens. The objective of those tests was to evaluate the stiffening effect of the glass beads and the constraint strength along the length direction.

2.4 OM Imaging

Optical images were obtained with a Reichert-Jung MeF3 microscope and a camera KODAK MEGAPLUS Model 4.2. This set-up is shown in Figure 2.4. The magnification level was chosen such that the resolution was 4.41 μm. With this resolution, twenty images were required in order to obtain the image of the entire specimen. The images were processed with Adobe Photo Shop software.

2.5 CT Imaging

CT images were obtained using the device at the High Resolution X-ray CT Facility at The University of Texas at Austin (http://www.ctlab.geo.utexas.edu/). The data were obtained as a set of 27 slices. The distance between the slices was 112 μm and the in-plane resolution was 76.8 μm. Thus both dimensions were about one tenth of the size of the glass beads. These data were processed with a software package developed at Center for Computational Visualization at The University of Texas at Austin (http://www.ticam.utexas.edu/CCV). In terms of functionality, this package is similar to the public domain package VTK (Visualization Tool Kit), but it is faster and better suited for our purposes. The software creates both two- and three-dimensional images fully automatically, provided that the user prescribes the threshold value for the iso-contours that represent material interfaces.

2.6 Weight Measurement

Weight measurements provided us with a benchmark data point for the volume fraction of the glass beads. Weight measurements were done with a precision microbalance whose resolution was 0.1 mg. Each epoxy was weighed five times. We relied on a single weighing for glass in part because the density of glass is a fairly stable quantity, and in part because our measurement agreed very well with a handbook value. To determine the volume fraction of the glass beads we measured the specific weight for the glass, epoxy, and composite. Then the volume fraction of the glass beads was calculated using the rule of mixtures.
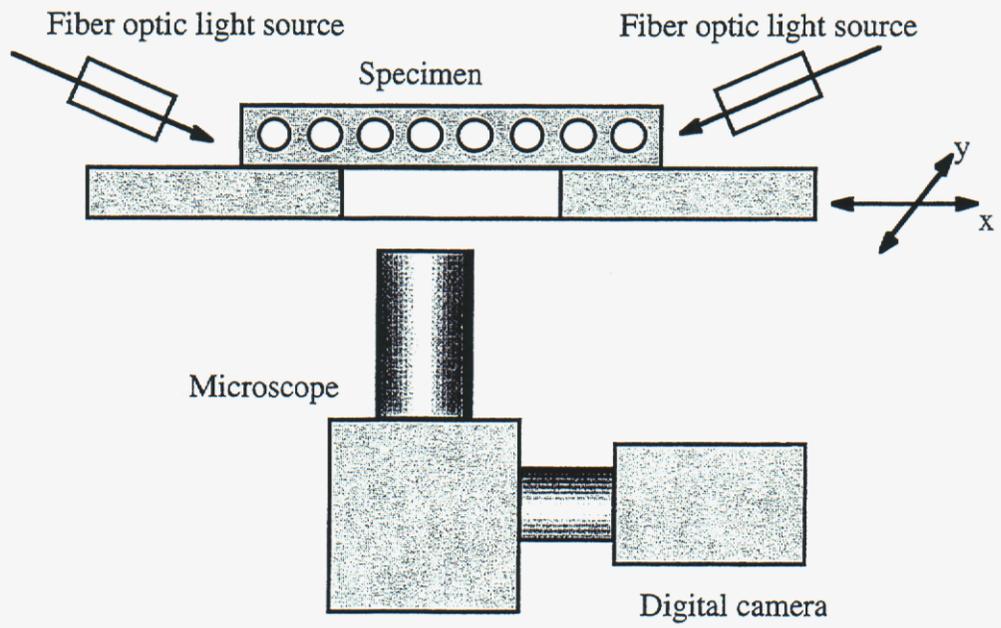
Figure 2.4   Schematic of the device for optical measurements

3. Prototype specimens: measurement results and analysis

3.1 Mechanical Testing
3.1.1 Epoxy A and Composite A Specimens

Results of the experiments on the Epoxy A and Composite A specimens are shown in Figure 3.1. There the apparent Young's modulus is plotted as a function of the specimen length; the apparent Young's modulus is defined as the stress-strain ratio along the thickness direction. The data were obtained using 4 Epoxy A and 4 Composite A specimens, and each specimen was tested 5 times.
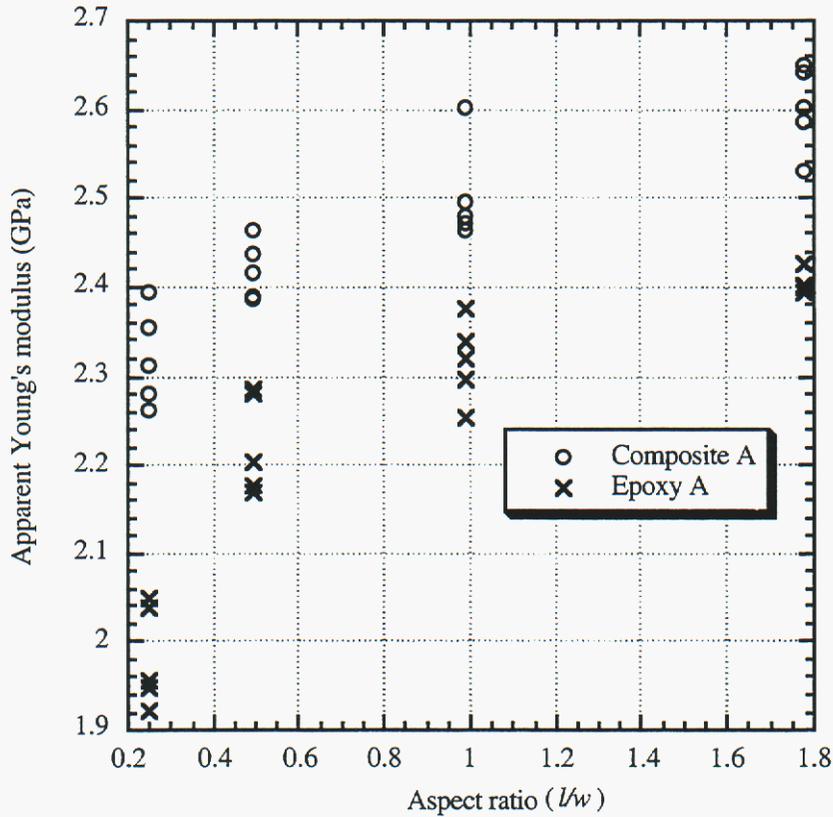
Figure 3.1   Apparent Young's modulus
of composite and pure epoxy specimens

For all specimens, the data exhibit scatter close to 5%, which cannot be correlated with any controlled experimental variables, and therefore the scatter should be attributed to experimental uncertainties. The apparent Young's modulus of the composite specimens is roughly 15% higher than that of the epoxy specimens, and it is independent of the specimen length. This behavior is expected, considering that the volume fraction of the glass beads is low, and therefore the glass beads cannot contribute significantly to

variations in the apparent Young's modulus from one specimen to another. The significant dependence of the apparent Young's modulus on the specimen length can be attributed to friction. For short specimens, those effects are insufficient to fully constrain the specimen along the length direction; in the other direction the specimen is constrained by the metal walls. Evidently, the data does not reach a plateau, and therefore additional experiments are required in order to determine the minimum specimen length that yields a fully constrained compression test.

3.1.2 Epoxy B Specimens

Based on the results for the Epoxy A and Composite A specimens, we concluded that the minimum specimen length can be determined using the Epoxy B specimens, rather than both Epoxy B and Composite B specimens. The apparent Young's modulus of the Epoxy A and Epoxy B specimens are plotted in Figure 3.2. The plot clearly shows the difference between the Epoxy A and Epoxy B specimens -- the former are significantly stiffer as a result of the longer exposure period under room temperature and humidity conditions. Also it is clear that the data for the Epoxy B specimens tends to reach a plateau. At least, the difference between the measurements for $l/w=3$ and $l/w=6$ is within 5%, which we adopted as the experimental uncertainty error. Further verification, using longer specimens, could not be carried out with our testing device.
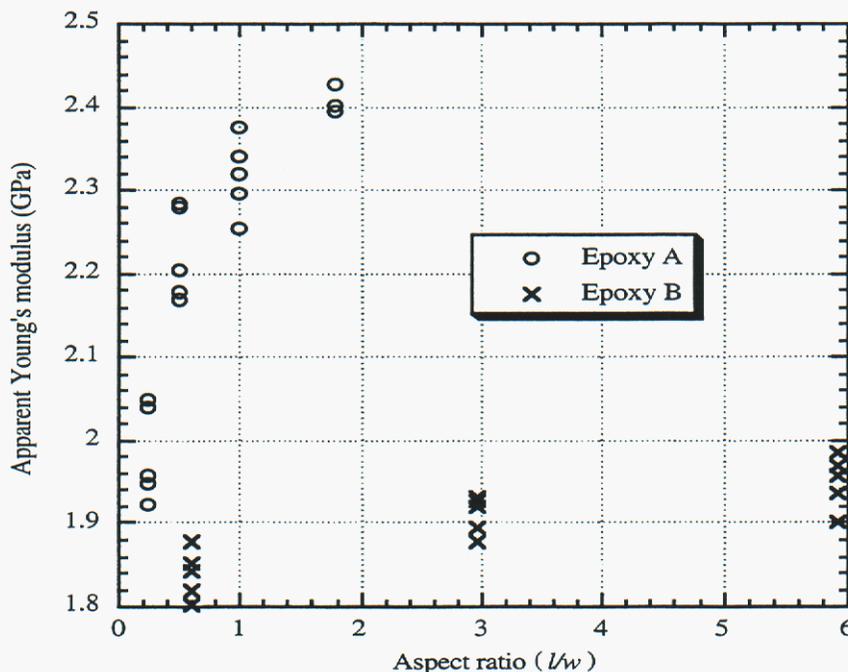
Figure 3.2   Apparent Young's modulus of pure epoxy specimens

19

3.2 OM Imaging

The optical image of the Composite B specimen is shown in Figure 3.3. This image is regarded as the benchmark for the CT image of the Composite B specimen. In addition we used this image for quantitative characterization of the microstructural geometry of the specimen. To this end, we assumed that the beads were spheres, and used the optical image to determine the center and radius of every sphere. These data were collected manually and recorded into a file. It was determined that the total number of spheres was 464, their mean diameter was 880 μm, and the volume fraction of glass was 7.6%. These data was also represented as a frequency plot for the sphere diameter (Fig. 3.4).
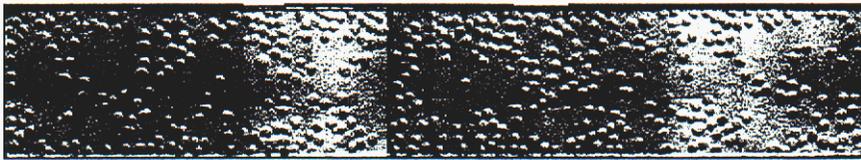
Fig. 3.3. Optical image of Composite B

To our surprise we discovered a relatively large percentage of small spheres. OM did not allow us to determine whether those spheres were glass beads or air bubbles. This issue was eventually resolved using the CT image, which can differentiate between the low-density air and high-density glass. Based on the CT image, we were able to conclude that the small inclusions were glass beads. Note that part of the fabrication process for the composite specimens included sieving of glass beads, in order to achieve a quasi-uniform population of the glass beads with respect to their size. Obviously the sieving procedure failed to exclude a relatively large number of small glass beads. We believe that those beads were not excluded because they were attached to large beads during the sieving procedure; a possible mechanism responsible for the attachment is static electricity.

We observed that many glass beads touched each other as well as the edge of the specimen. The random placing of hundreds of glass beads during its manufacturing process allowed them to touch to each other easily. This situation is typical for the majority of composite materials, since in most cases the volume fractions of particles or fibers significantly exceeds 7.6% considered in this study. The microstructures that involve many touching or almost-touching particles or fibers are particularly difficult for both experimental characterization and computational modeling because they require one to resolve the small gaps between the particles.
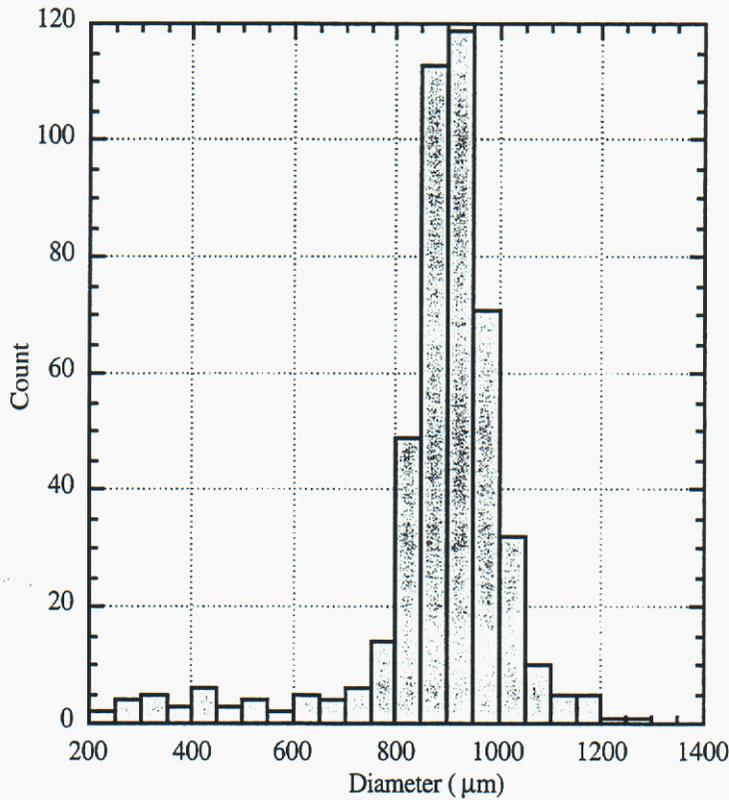
Figure 3.4   Histogram of the glass beads in Composite B

3.3 CT Imaging

Figure 3.5 shows a CT image of the middle slice of the specimen. Upon comparison of this image with the optical image in Figure 3.3, one can easily see major differences between the two images. In particular, within 1 cm from the left edge, the optical image shows about 20 spheres whereas the CT image does not show any spheres. Most likely, this discrepancy is due to tilting of the specimen during CT scanning. Apparently, the specimen was inadequately supported against the rotations about the x or y axes, which led to tilting and eventually to the poor image. Also, on the CT image, some particles do not at all look like spheres. This is expected for small particles whose

image consists of a small number of pixels and nearly touching particles whose image resembles a dumbbell.
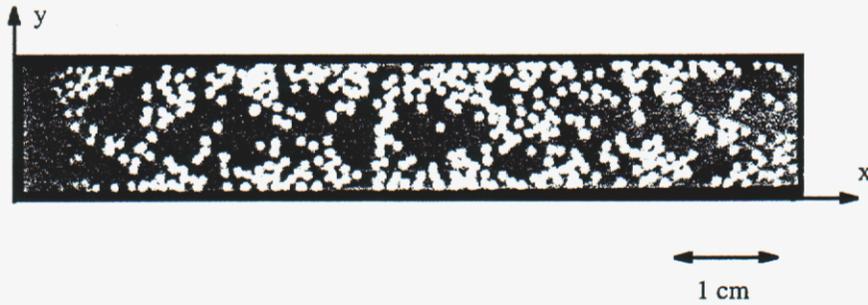
Figure 3.5   CT image of Composite B

Besides the errors associated with finite spatial resolution, the CT image is also prone to errors associated with data segmentation or the definition of a value for the data iso-contours that represent the glass-epoxy interfaces. It is clear that the quality of three-dimensional CT images cannot be better than that of two-dimensional CT images. From this perspective, a three-dimensional CT image of the Composite B specimen is worthless. Nevertheless, producing a three-dimensional was a useful exercise because it revealed to us two additional difficulties. First, the CT device assigned density values that varied from one slice to another. Some CT devices post-process the data so that this problem is eliminated, but not the device used in this study. Second, the data was polluted by the so-called ghost effect [14]. This effect can be eliminated if the data for each slice are collected after a certain, machine-dependent, time interval. The ghost effect cannot be eliminated with "creative" data post-processing.

3.4 Weight Measurement

Weight measurements for the largest Epoxy B specimen (l=75.9 mm) and Composite B specimens are summarized in Tables 3.1 and 3.2, respectively. In addition to the weight, these tables also contain the measurements of the specimen dimensions. Based on these data and the value for the specific weight of glass (see Table 3.3), we concluded that the volume fraction of glass beads was between 5.9% and 7.9%. This

range is consistent with the value of the volume fraction obtained from the optical image, which was 7.6%.

Measurement number	l mm	w mm	t mm	W g
1	75.89	12.90	1.88	2.072
2	75.90	12.92	1.88	2.072
3	75.89	12.90	1.88	2.072
4	75.90	12.90	1.88	2.072
5	75.90	12.88	1.88	2.072

Table 3.1   Density measurements for the Epoxy B specimen (l=75.9 mm)

Measurement number	l mm	w mm	t mm	W g
1	75.91	12.94	1.84	2.242
2	75.88	12.90	1.88	2.242
3	75.95	12.92	1.88	2.242
4	75.92	12.87	1.88	2.242
5	75.92	12.88	1.88	2.242

Table 3.2   Density measurements for the Composite B specimen

Radius μm	W mg	Density mg/mm^3
629	2.7	2.60

Table 3.3   Density measurement for a glass bead

23

3.5 Recommendations

The study of the prototype specimens fully confirmed the notion that detailed microstructural characterization is a challenging task, even for specimens with simple microstructure. Based on the study of the prototype specimens, we concluded that the fabrication and testing procedures for the model specimens must be improved as follows:

- In order to constrain the specimen along the length dimension, and thus realize a bi-axially constrained compression test, the specimen length l should exceed its width w by at least a factor of six.
- Fabrication procedures must be refined in order to eliminate small glass beads and control the distances between the glass beads and between the glass beads and the specimen surface. At this stage, the second requirement is critical for developing reliable material characterization procedures based on CT data.
- OM provides adequate images of the specimen, but it can benefit from CT imaging which can differentiate between small glass beads and air bubbles.
- CT data acquisition must be improved as follows:

    The model specimen must be smaller than the Composite B specimen in order to be scanned with better resolution.

    The specimen fixture inside the CT device must be improved in order to eliminate tilting.

    During scanning, the model specimen must be surrounded by a significant volume of glass, so that one can determine the peak corresponding to the density of glass.

    The scanning procedure must be improved in order to eliminate the ghost effect.

4. Model specimen

This section is concerned with the model specimen. This specimen is an improvement over the prototype specimens, both in terms of fabrication and testing. In order to eliminate repetition, our presentation emphasizes the differences between the model and prototype specimens.

4.1 Geometry

The model specimen was significantly smaller in size than the Composite B specimen involved in OM and CT imaging. The model specimen dimensions were $l = 11.08$ mm, $w = 1.93$ mm, and $t = 1.42$ mm. For the specimen microstructure, we selected eight glass beads and positioned them in a row along the length direction (Fig. 4.1). The beads had almost perfect spherical shape and their diameters were close to 900 μm. The beads were placed such that the gaps between them and the gaps between the beads and the specimen surface were on average 250 μm; the smallest distance was 137 μm.

(a)

1 mm

(b)

1 mm
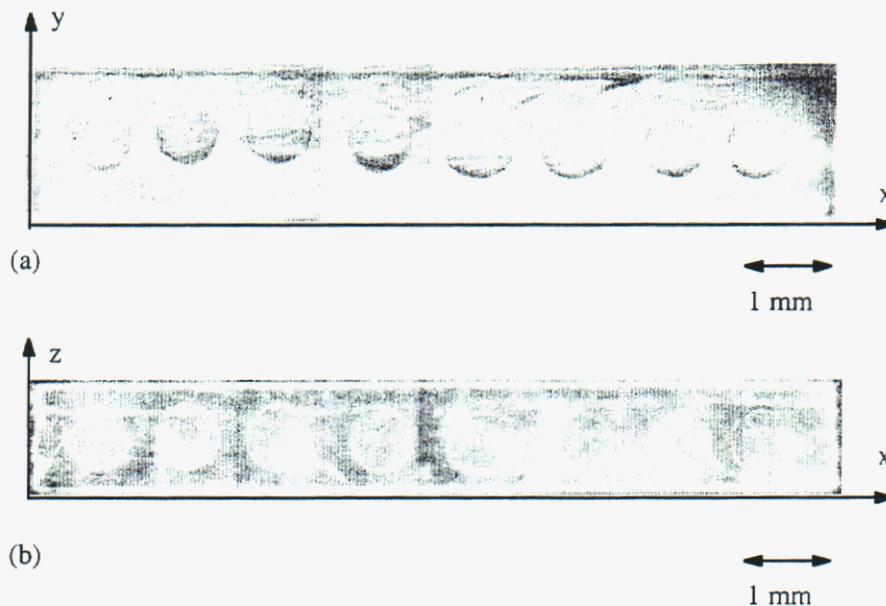
Figure 4.1   Optical images of the model composite specimen: (a) top view and (b) side view

4.2 Fabrication

The fabrication procedures for the model specimen had to be modified in order to realize controlled positioning of the glass beads. The modifications were as follows (see Fig. 4.2):

1. The positions of the glass beads were controlled manually. The optical microscope was used for monitoring the procedure.

2. Before placing the spheres, the top surface of the first layer was covered with a thin layer of uncured epoxy, which allowed us to maintain the beads in proper positions.

3. After placing the beads, the specimen was cured again at 160 °F for 60 minutes. This allowed the thin layer of epoxy to harden, and as a result the beads could be maintained in proper positions during the vacuuming stage.

4. To reduce the residual stress after curing inside the mold, the model specimen was annealed outside the mold at 160 °F for 60 minutes.

In addition to the model composite specimen, we fabricated an epoxy specimen with the same dimensions and curing history, except that the epoxy specimen was subjected to two rather than three curings. Both specimens were tested two weeks after fabrication.

Step 1: Pour and cure the first layer

Step 2: Apply thin layer of epoxy, position the glass beads, vacuum and cure

Step 3: Pour the 2nd layer of epoxy, vacuum and cure

Step 4: Remove model specimen from the mold and anneal
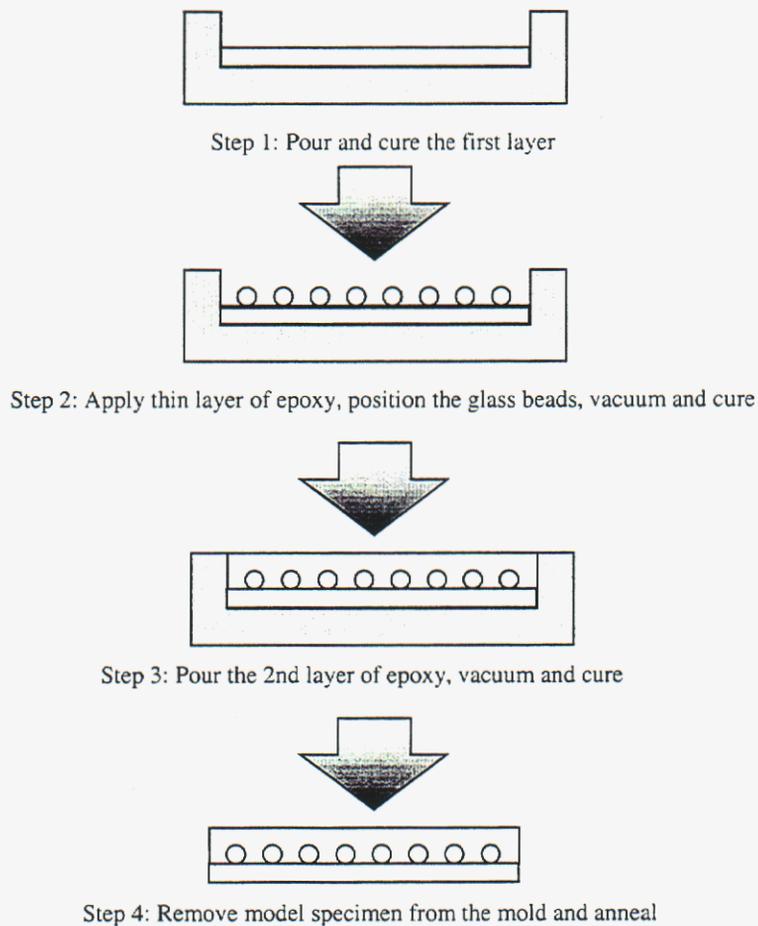
Figure 4.2　　Fabrication of the model composite specimen

4.3 Imaging Procedures

For the model specimen, we obtained optical images of the top and side views, using the resolution of 2.21 μm.

The CT data for the model composite specimen consisted of 36 slices along the thickness direction. The slices were spaced uniformly, and the inter-slice thickness was 46.7 μm. Within each slice, the pixel resolution was 25.5 μm, so that the CT image was obtained from the data whose resolution was about 3% of the mean diameter of the glass beads. These pixel resolution and inter-slice thickness were about one half of those of the Composite B specimen, so that the CT image was derived from the data whose resolution was about one twentieth of the mean diameter of the glass bead. These choices were made on the basis of experience with the Composite B specimen. In addition, care was taken to eliminate the ghost effect.

To obtain the histogram peak for the glass phase, the specimen was scanned together with a piece of industrial pure quartz, such that the total volume of glass was roughly equal to that of epoxy.

Also, as part of CT data processing, we included the median filter technique [15] to assess the CT data quality.

4.4 Mechanical Testing Procedures

Unfortunately, the dimensions of the model specimen were too small for constraining the deformation in the width direction; for the prototype specimen, the deformation was constrained by the loading device. This significantly complicates interpretation of the test results since we no longer can assess the macroscopic stress in the model specimen. Also, the model specimen was tested using a smaller load cell.

After the specimen had been tested in the elastic regime, it was tested in the inelastic regime until failure. The displacement was controlled such that the specimen was subjected to a series of loading-unloading tests. The peak load for each test was 450 N more than the peak load of the previous test. After each test, the specimen was examined using an optical comparator. The specimen images were recorded with a video camera, so that the specimen could be viewed at 10× magnification in an optical comparator connected to the DVD recorder (Panasonic WJDR 200). The schematic of this apparatus is shown in Figure 4.3.

Optical comparator

Digital video camera

OOO OOOOO

Specimen

DVD recorder & monitor

OOOOOOOO

Light sorce

Personal computer

OOOOOOOO

DVD-RAM
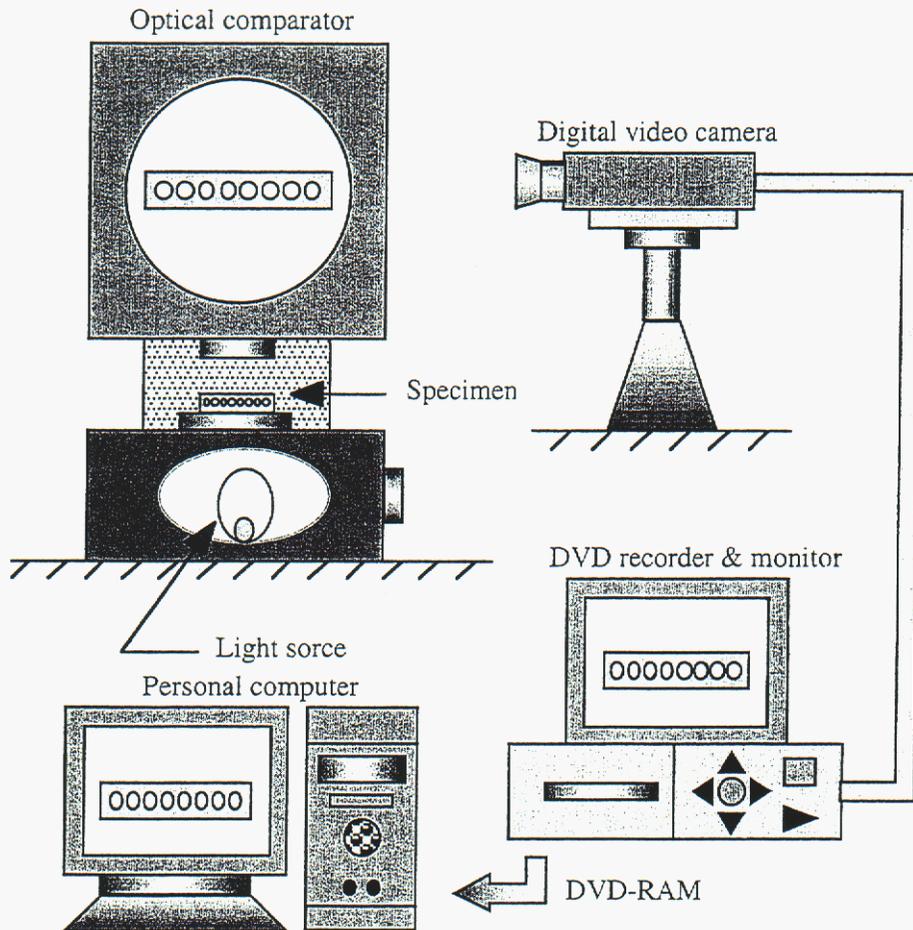
Figure 4.3   Schematic of the set-up for observation
               of the compression failure

4.5 Results

4.5.1 OM Imaging

Figure 4.1 shows the top and the side view optical images of the model specimen.
It is clear that we were successful in positioning the glass beads such that they were not
too close to each other and not too close to the specimen surface. OM revealed the
presence of a tiny glass bead next to the second main bead, if the main beads are counted
from left to right. The tiny bead cannot be seen in Figure 4.1 because it is not perfectly
focused, but it is clearly seen in Figure 4.7. We have no good explanation how the tiny

bead ended up inside the specimen, but the fact that the bead was there indicates that the fabrication procedures need further refinement.

By assuming that the beads are spherical, we estimated their mean diameter as 897 μm and the volume fraction of glass as 9.67%. In this measurement, the tiny glass bead was neglected.

4.5.2 CT Imaging

Figure 4.4 contains three histograms corresponding to the CT data for the middle slice. The black line represents the raw data and the red line represents the filtered data. The blue line was obtained by removing the data representing quartz followed by filtering the data representing the specimen. It is clear that all three peak, representing air, epoxy, and glass, could be easily identified with all three data processing procedures. On the other hand, both quartz and filtering enhance the glass peak.
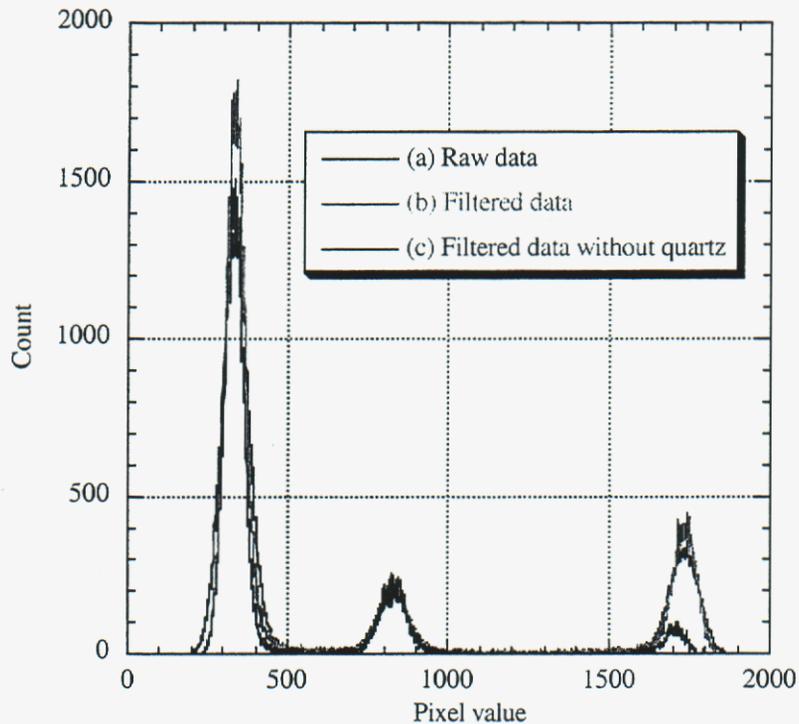
Figure 4.4   Histograms based on:
    (a) raw data
    (b) filtered data
    (c) filtered data without quartz

Figure 4.5 shows two images of the middle slice of the specimen; one image was obtained from the raw data and the other from the filtered data without quartz. The threshold values for both images were the same, and they were chosen as the midpoints

between the peaks. Both images are in good agreement with the optical image shown in Figure 1a. Filtering makes the second image less noisy than the first one, but the difference is very minor, at least visually.
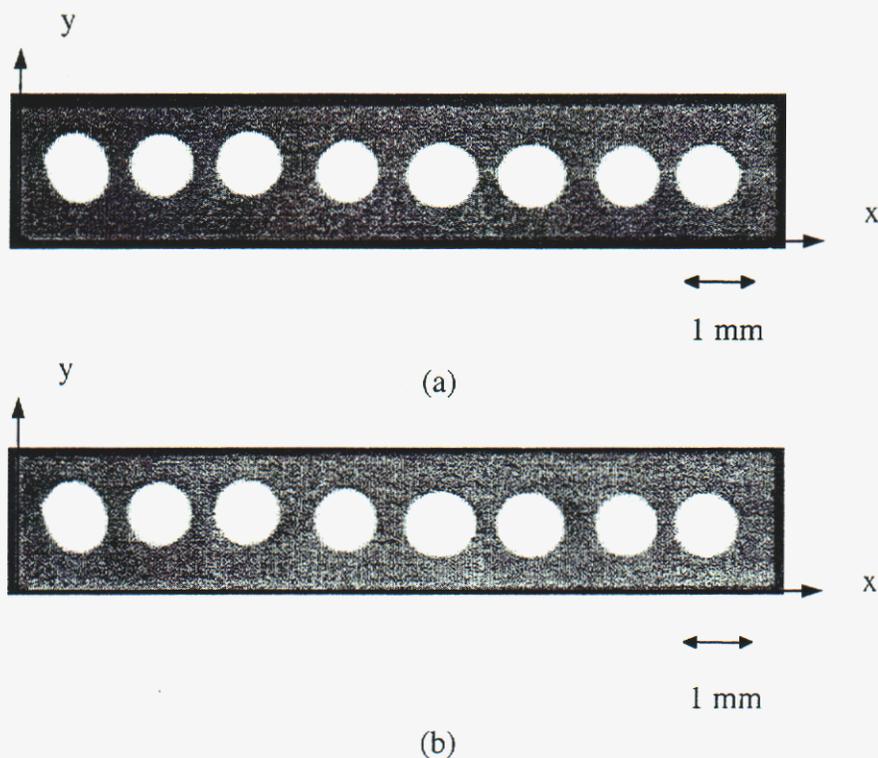
Figure 4.5   CT Images:
(a) raw data and (b) filtered data without quartz

Figure 4.6 contains thirteen histograms corresponding to various slices using filtered data without quartz. These histograms show that the density shifts from one slice to another were rather minor, and therefore the peaks did not have to be adjusted before making the three-dimensional image.
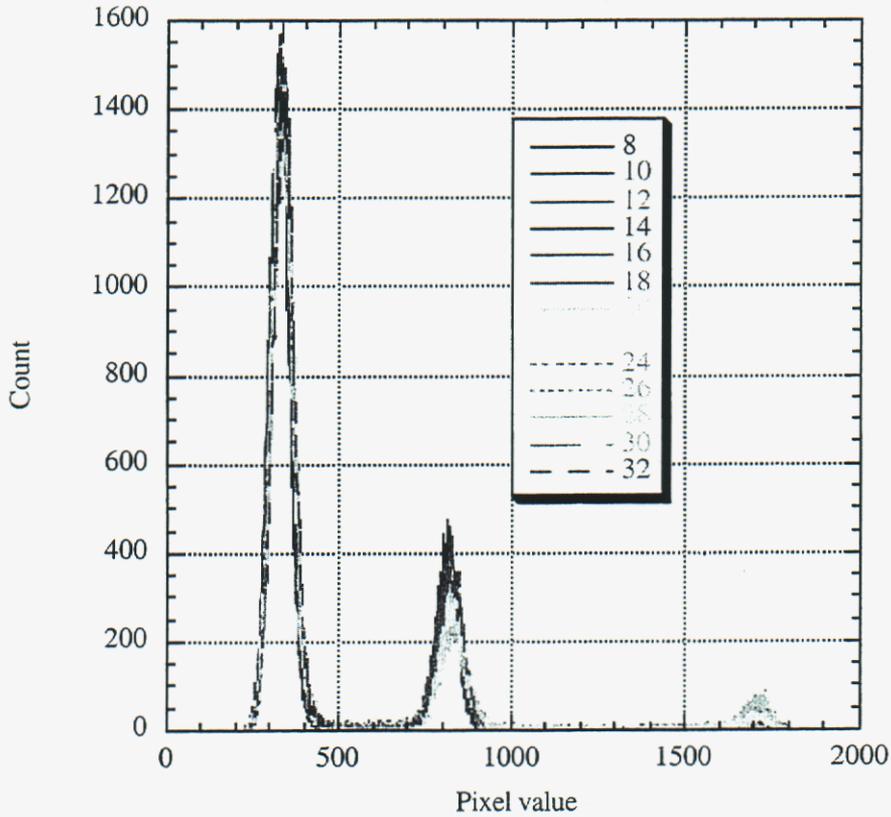
Figure 4.6 Histograms for various slices
for filtered data without quartz

Figure 4.7 contains a three-dimensional CT image of the specimen derived from the raw data. The threshold values were chosen equal to those for the middle slice. The image is consistent with the images shown in Figures 4.1 and 4.5. The image clearly shows the tiny glass bead. Also it is clear that this is a glass bead but not an air bubble. The tiny glass bead was also observed in the optical images but only after it had been detected by the CT device.

Figure 4.7   CT image of the model composite specimen for raw data

31

4.5.3 Weight Measurement

Based on weight measurements, it was estimated that the volume fraction of glass had to be between 9.4% and 9.6%. The OM estimate of 9.67% is outside this tight range, but this discrepancy can be easily attributed to deviations of the glass beads from the spherical shape.

4.5.4 Elastic Behavior

Figure 4.8 presents the stress-strain curves obtained by repeating the same test five times. According to Figure 4.8, the results are reproducible; in addition, the quality of these data was confirmed by monitoring alignment of LVDTs. From these data we calculated the apparent Young's modulus of the model specimen as 2.05 ± 0.03 GPa. Similarly, we determined the apparent Young's modulus of the corresponding epoxy specimen as 1.76 ± 0.06 GPa.
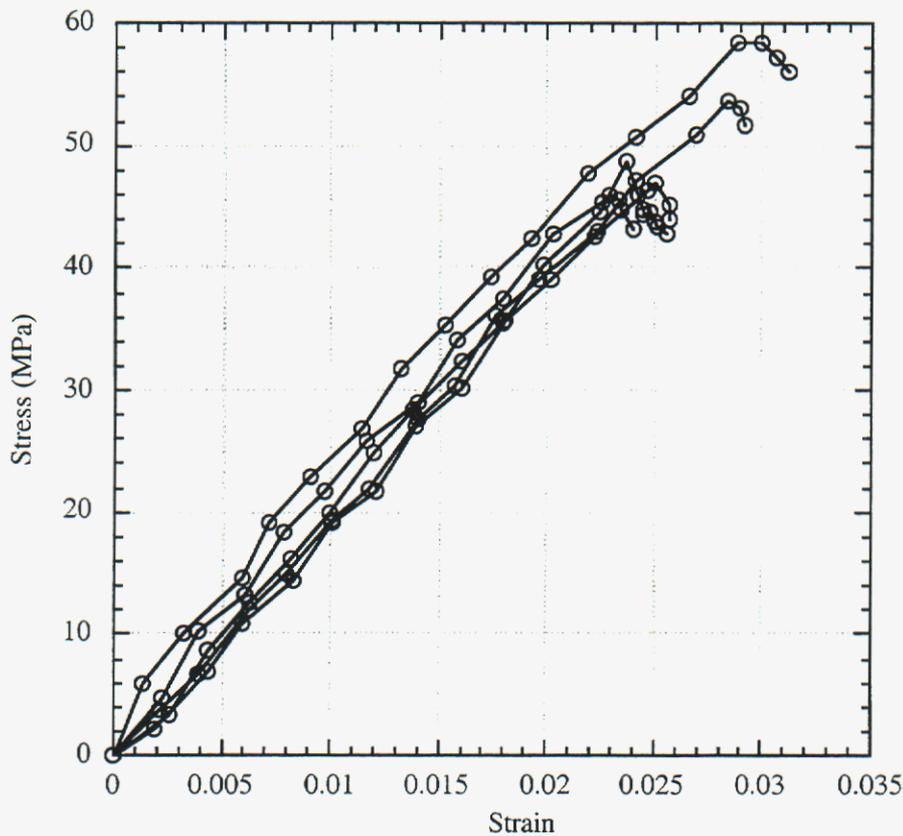
Figure 4.8   Stress-strain curves in the elastic regime

32

4.5.5 Fracture Behavior

Figure 4.9 contains a set of optical images that describe the evolution of fracture in the specimen. For the peak loads in Figures 4.9a-e, there are two images -- one presents the top view and the other the side view. For peak loads in Figures 4.9f-i, there is only image, for the top view. Apparently, fracture was initiated when the force P was between 1.74 and 1.93 kN. It appears that the cracks were initiated between the beads and the upper surface of the specimen. At $P=1.93$ kN, the cracks are particularly well developed near the middle beads. Those cracks overlap at the upper surface of the specimen but not at the mid-surface. As the load increases, the cracks in the middle continue to spread and at $P=2.48$ kN two of them overlap at the mid-surface. Also at this load, the side view of the specimen appears symmetric, which indicates that there was significant cracking between the beads and the lower surface. The cracks appear to stay in the mid-plane perpendicular to the y-axis for $P \leq 3.08$ kN. For $P \geq 3.57$ kN, the cracks start propagating out of this plane until a major crack develops at $P=4.73$ kN. Also, for $P \geq 3.57$ kN, the cracks interfered with optical imaging to the extend that the side view images were impossible to resolve.

(a)

(b)

(c)

1 mm
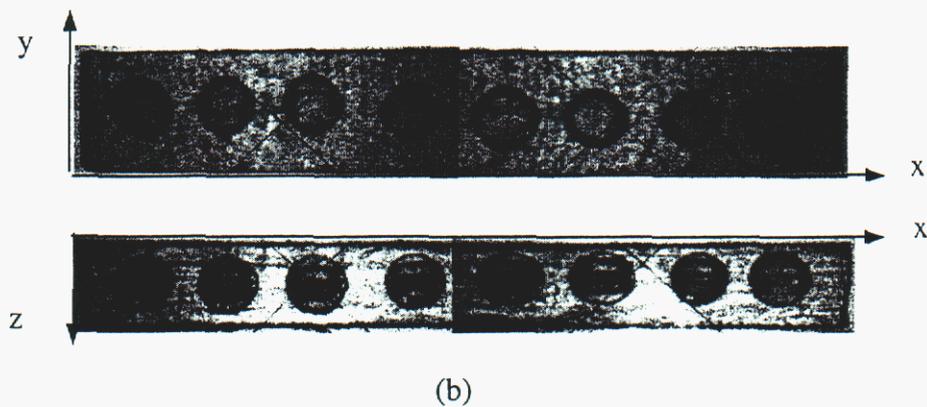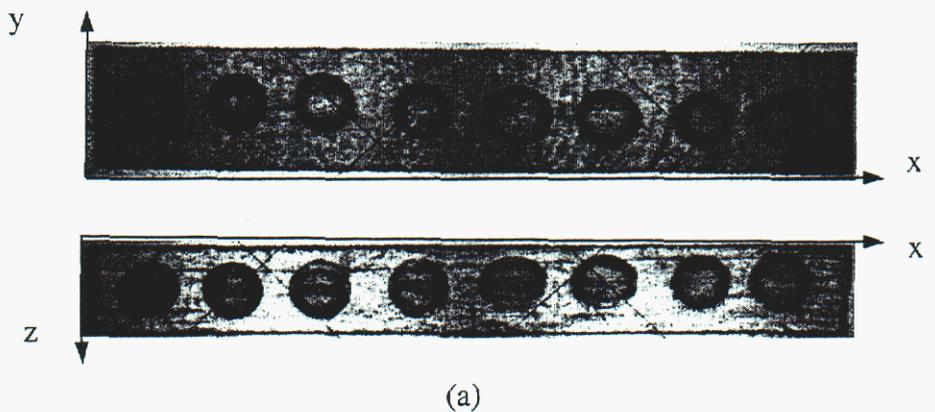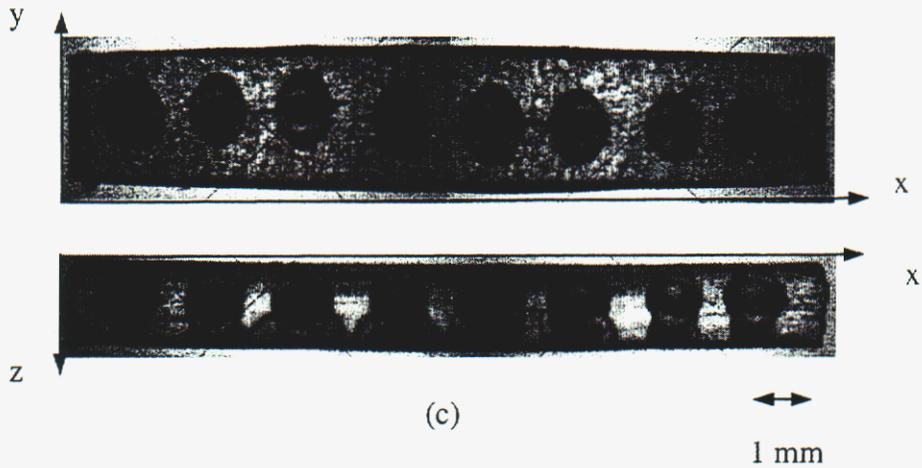
Figure 4.9A   Fracture patterns: (a) P=1.56 kN, (b) P=1.74 kl.., ⌐, ..93 kN

It is impossible to provide a detailed explanation of fracture evolution without detailed numerical analysis. Nevertheless, we believe that the tendency for the cracks to initiate at the upper rather than lower surface of the specimen may be related to the curing procedure -- the upper surface was cured once whereas the lower surface was cured three times. Also the tendency for the cracks to initiate near the beads in the middle may be related to friction effects, since we expect no slip conditions in the middle and easy slip at the ends of the specimen.

(d)

(e)
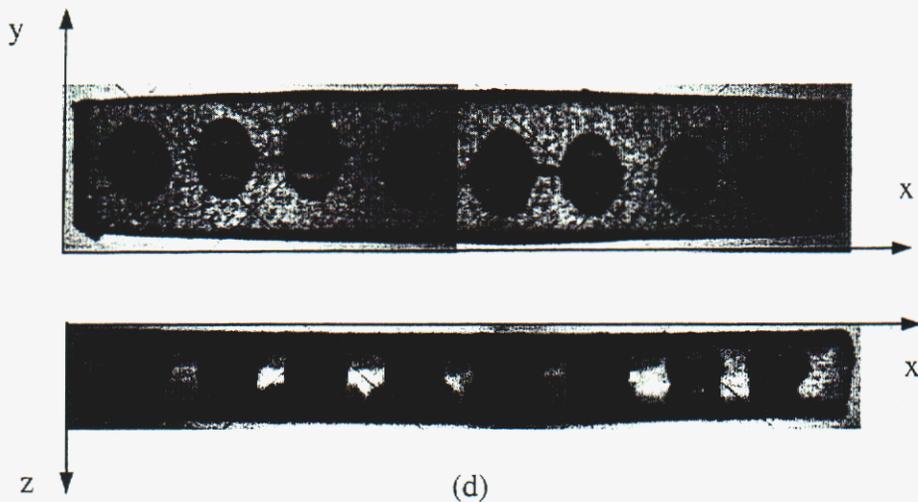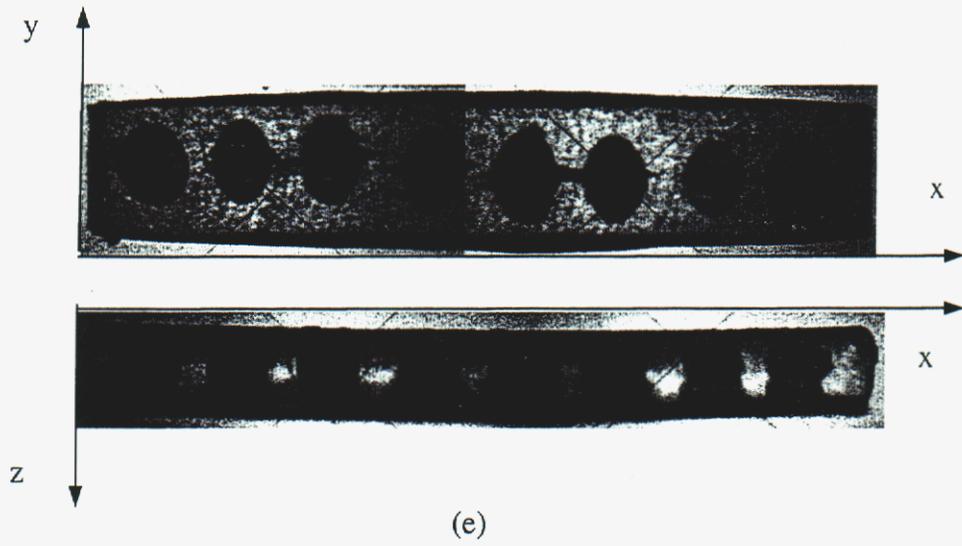
Figure 4.9B   Fracture patterns: (d) P=2.48 kN, (e) P=3.08 kN

(f)

(g)

y

x

(h)

y

x
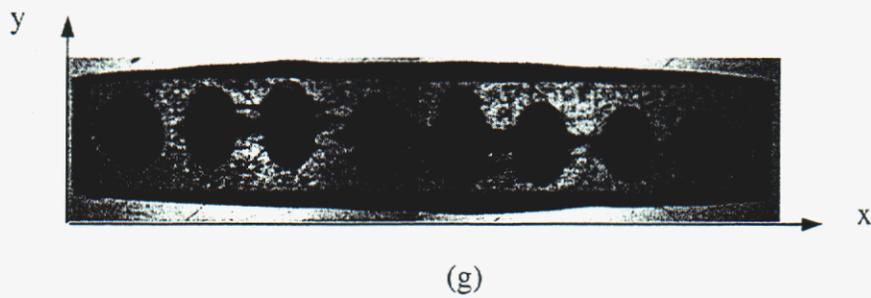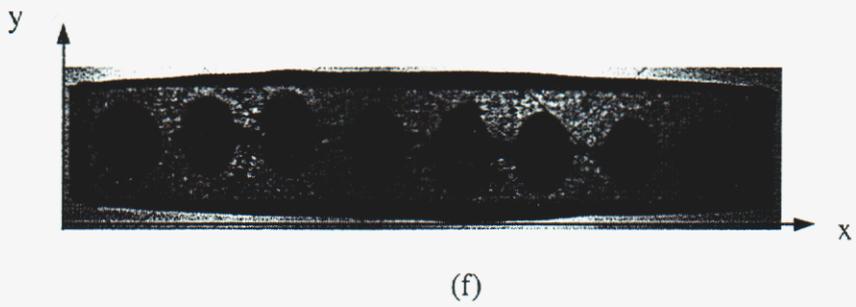
(i)

Figure 4.9C   Fracture patterns: (f) P=3.57 kN, (g) P=4.01 kN,
(h) P=4.55 kN, (i) P=4.73 kN

5. Conclusion

In this study, we considered several glass-epoxy specimens with two objectives in mind. First, we wanted to develop an understanding of how CT imaging could be utilized in microstructural characterization of mechanical behavior of composite materials. Second, we wanted to obtain experimental results that would allow us to benchmark computational modeling of composite materials.

We regard results of this study as mixed. On the one hand, we were able to develop a good practical understanding of what it takes to obtain a (visually) good CT image. On the other hand, such images can be obtained only for carefully crafted specimens, so that the practical use of CT imaging appears to be limited. This is certainly true for most fiber-reinforced materials with fibers close to 10 μm in diameter. Also, results of mechanical testing on the model specimen are of limited value because we failed to constrain the specimen so that we could unambiguously determine the induced macroscopic stress.

The improvements in the imaging procedures for the model specimen mostly due to a better understanding of CT imaging techniques that are well known to experienced CT technicians. Nevertheless each specimen has its own specifics and an understanding of those specifics is critical for producing good images, especially for features close in size to the resolution threshold.

Surprisingly, even for quasi-two-dimensional microstructures, CT imaging was able to identify certain features better than OM. CT was particularly useful in distinguishing between air bubbles and small glass beads. However, it would be counterproductive to consider OM and CT imaging methods as competing techniques. To the contrary, one should search for ways of integrating these methods to the maximum extent possible, so that it would be possible to exploit their advantages.

Based on results of this study, the following recommendations are made for future investigations:

(1) Consider modifications in the model specimen geometry and/or mechanical testing procedures that would lead to experimental data suitable for benchmarking of computational modeling methods.

(2) Compare experimental and computational data for a new model specimen, and establish possible ways of combining those data for integrated microstructural characterization of mechanical behavior of composite materials.

(3) Explore the applicability of CT imaging.

6. References

[1]   Z. Hashin, "Analysis of composite materials—a survey", *Journal of Applied Mechanics*, 1983, Vol. 105, pp. 481-505.

[2]   R.N. Yancey and S.J. Klima, "High-resolution computed tomography of advanced composite and ceramics materials", *Proceedings of the 15th Annual Conference on Composites and Advanced Ceramics*, AcerS, Cocoa Beach, FL, 1993, 13-16 January.

[3]   R.N. Yancey, " A Micromechanical constitutive damage model for use with computed tomography data," *NSF SBIR Phase I Final Report*, Grant III-9260829, 1993.

[4]   R.N. Yancey and G.Y. Baaklini, "Structural characterization of metal-matrix composites using computed tomography data", *SPIE*, 1996, Vol. 2944, pp. 2-12.

[5]   R.N. Yancey and R. Dzugan, "CT-assisted metrology for manufacturing applications", *SPIE*, 1996, Vol. 2948, pp. 222-230.

[6]   J.H. Stanley and R.N. Yancey, "CT-assisted agile manufacturing", *SPIE*, 1996, Vol. 2948, pp. 232-240.

[7]   R.H. Bossi, D.A. Cross and R.A. Mickelsen, "Radioscopy and computed tomography for failure analysis using microfocus X-ray", *SPIE*, 1996, Vol. 2944, pp. 13-20.

[8]   S.J. Hollister and N. Kikuchi, "Homogenization theory and digital imaging: a basis for studying the mechanics and design principles of bone tissue", *Biotechnol. Bioeng*, 1994, Vol. 43, No. 7, pp. 586-596.

[9]   K. Terada and N. Kikuchi, "Microstrucutral design of composites using the homogenization method and digital images", *Materials Science Research International*, 1996, Vol. 2, No. 2, pp. 65-72.

[10]   K. Terada, T. Miura and N. Kikuchi, "Digital image-based modeling applied to the homogenization analysis of composite materials", *Computational Mechanics*, 1997, Vol. 20, pp. 331-346.

[11]   C. Huet, "An integrated micromechanics and statistical continuum thermodynamics approach for studying the fracture behaviour of microcracked heterogeneous materials with delayed response", *Engineering Fracture Mechanics*, 1997, Vol. 58, No. 5-6, pp. 459-556.

[12]   J.G. Swadener and K.M. Liechti, "Asymmetric shielding mechanisms in mixed-mode fracture of a Glass/Epoxy interface", *Journal of Applied Mechanics*, 1998, Vol. 65, pp. 25-29.

[13]   C.C. Chamis, *Simple composite micromechanics equations for mechanical, thermal and moisture – related properties, Guide to composite materials*, ASE International, Materials, OH, 1987.

[14]   G.T. Herman, *Image reconstruction from projections: the fundamentals of computerized tomography*, New York: Academic Press, 1980.

[15]   C. Lindley, *Practical image processing in C*, 1991, pp. 378.

CHAPTER 2

Adaptive Modeling in Computational Solid Mechanics

J. Tinsley Oden, Kumar Vemaganti, Serge Prudhomme,

Daniel C. Hammerand, and Samuel W. Key

1. Introduction

Over the last few decades, the great advances in computational mechanics have been heralded as major milestones in science and engineering, providing the knowledge, tools, and techniques that, with the use of modern computers, may permanently change the way science and engineering is done. Furthermore, these advances may have a profound impact on virtually every aspect of human existence and well-being. At the heart of this promise is the predictive power of computer simulation, enabling the modeling of complex physical events and extending our ability to understand, control, and predict the behavior of physical and engineering systems.

There is, however, a growing concern as to the real extent to which this great promise will ultimately be fulfilled. Will the new developments lead to predictive tools of a fidelity, accuracy, and reliability sufficient to be used with confidence in applications of increasing complexity and importance? The predictive power of computer models depends upon two fundamental factors: 1) The accuracy with which the mathematical systems governing the model (the partial-differential equations, integral equations, constraints, etc.) are solved and 2) the suitability of the mathematical model selected to abstract a particular set of physical events of interest. The first factor leads to questions of *verification*; as Roache [26] put it "verification asks the question, have the equations been solved right?" The second factor leads to questions of *validation*; or, from [26], "have the right equations been solved?" Issues of verification call upon methods of *a priori* and *a posteriori* error estimation, benchmarking, software engineering. Issues of validation traditionally call upon physical experiments and testing. Our interest here is in the latter category, model selection. In particular, can computational and mathematical procedures be developed that aid or complement

the validation process? We provide what may be an affirmative answer to this question embodied in the notion of hierarchical modeling and model adaptivity.

The selection of the mathematical model, the fundamental step in performing a computer simulation of a physical event, is a step left largely to heuristic arguments, judgment and experience of the analyst, or is based on incomplete empirical data. Indeed, different analysts frequently choose different models to study the same physical phenomenon. Nevertheless, the selection of the model and the parameters that characterize it is arguably the most important step in computer-based science and engineering and is often the dominant source of error in computer simulations. Since all models of physical events are approximations to nature, all models are in a sense wrong. According to Box [5], "All models are wrong, but some are useful;" Easterling [9] observes that "Useful models are those for which the prediction error, the difference between nature and computation, is tolerable in the context in which the model is to be used. The problem is to establish 'usefulness.'"

The notion of hierarchical modeling, while itself based on a selection process and also based on mathematical (and, therefore, "wrong") models, provides a potentially more systematic approach for selecting models from a well-defined class of models. In this approach, we seek to define a class $\mathcal{M}$ of mathematical models of a certain class of physical phenomena, which includes models that are presumably candidates for modeling all events of interest. Within this class $\mathcal{M}$ is a so-called fine model, which is also a mathematical model, but one of such detail, sophistication, and complexity that all phenomena of interest in a set of simulations are predictable by it with sufficient accuracy. Many of the properties of the fine model may presumably be determined through experiments, testing, imaging, and other means. All other models within $\mathcal{M}$ are coarser or simplified models. While general and inclusive, the fine model is often too complex to be used to obtain quantitative results; thus solutions of the fine model are never actually computed, except possibly for very special cases. The fine model is used only as a datum against which modeling error in coarser models is measured. The fine model, for example, may characterize phenomena occurring at many spatial and temporal scales and may embody many interacting physical effects, while various coarser models may be characterized by averaged mechanical properties and simplified laws. Ultimately, the suitability of the fine model itself must be estimated by determining its predictive limits within the context of a still larger class of models or, unavoidably, through more traditional validation procedures such as physical experiments. Once a set of hierarchical models is defined, the suitability of a given (tractable) coarse model is determined by *a posteriori* modeling error estimates, bounds in various norms of the error in a solution of a coarse model compared to the unknown fine model solution.

Once the *a posteriori* estimates of modeling errors are in hand, adaptive procedures can be designed to make possible the automatic selection of appropriate models within the class $\mathcal{M}$.

2. Adaptive Modeling for Heterogeneous Materials

The study of heterogeneous materials offers an example of a class of models $\mathcal{M}$, wherein the finest model leads to a problem that is virtually impossible to solve, while a coarse model invariably leads to an erroneous solution. Direct numerical simulation of the deformation of highly heterogeneous materials – even under the assumption that the constituents are linearly elastic – is well beyond the capacity of modern computers, due mainly to the vast number of degrees of freedom required to capture and resolve the heterogeneities. Traditionally, the body is treated as a homogeneous medium characterized by averaged properties. While this approach leads to manageable problem sizes, it also results in the loss of crucial fine-scale information.

The methodology of hierarchical modeling, applied to the analysis of linearly elastic heterogeneous materials, was first described in [30, 20]. In these papers, a posteriori bounds on the error in solutions to elastostatics problems induced by replacing fine-scale micromechanical properties by coarser scale or effective properties were derived in global energy norms. These error estimates were then used as a basis for an adaptive modeling process in which only enough fine-scale information sufficient to deliver results of a preset accuracy, measured in energy norms, is used to characterize the model. The resulting adaptive process can lead to significant computational savings, making possible the analysis of micromechanical effects in some cases that are intractable by traditional approaches. Preliminary results on extensions of these adaptive approaches to a class of models depicting material damage were discussed in [21].

More recently, the theory of a posteriori modeling error estimation for heterogeneous materials was extended to "quantities of interest" that represent local features of the response [18]. These quantities of interest could represent, for example, average stresses on material interfaces, boundary displacements, or mollified pointwise displacements, strains, or stresses. Mathematically, a quantity of interest is any feature of the fine-scale solution that can be characterized as a continuous linear functional on the space of functions to which the fine-scale solution belongs. Computable upper and lower bounds and sharp estimates of the errors in such quantities are established in [18]. Based on these estimates, an adaptive modeling algorithm, referred to as the GOALS (Goal-Oriented Adaptive Local Solution) algorithm, is developed.

The theory assumes that the microstructure, and hence the elasticity tensor, is known almost everywhere in the body. However, in most engineering materials, the number of microstructural components in enormous, often exceeding millions of constituents, and the geometry, orientation, and mechanical properties are unknown or not known with significant precision. Thus, it is usually necessary to use specialized technologies to determine the microstructure. In [29], Computerized Tomography (CT) images generated by X-ray devices are used to determine the microstructure locally. A detailed description of a computational environment that integrates CT imaging and h-p adaptive finite element methods with the GOALS adaptive modeling algorithm is available in [29].

In the remainder of this section, we describe various aspects of this integrated adaptive modeling methodology. First, the overall adaptive modeling strategy is described. Next, each module of this strategy is individually discussed.

2.1 Overview of the adaptive modeling strategy

The general ideas of the adaptive-modeling algorithm depicted in Fig. 2.1, are

1. A structural component is given or fabricated and a list of analysis goals, so-called *quantities of interest*, is identified; the accurate calculation of these quantities is the goal of the analysis;

2. X-ray tomography is used to scan the specimen or various portions of it to approximately define the internal microstructure;

3. A homogenization module accepts imaging data taken from sampled sections of the body and computes effective mechanical properties using an adaptive finite element method;

4. The effective properties are used as input data in the adaptive hp finite element program which computes a highly accurate "homogenized" solution to the equations of elastostatics for given loading and applied force data (a step in which the response of the body is modeled mathematically as a homogeneous, linearly elastic body);

5. The homogenized solution is input to a module which implements the GOALS algorithm which estimates the errors in the quantities of interest (errors due to modeling the heterogeneous material as a homogenized medium) and adaptively adjusts the calculated quantities (by using additional microscale data supplied by the imaging process) until preset levels of accuracy are attained;

42
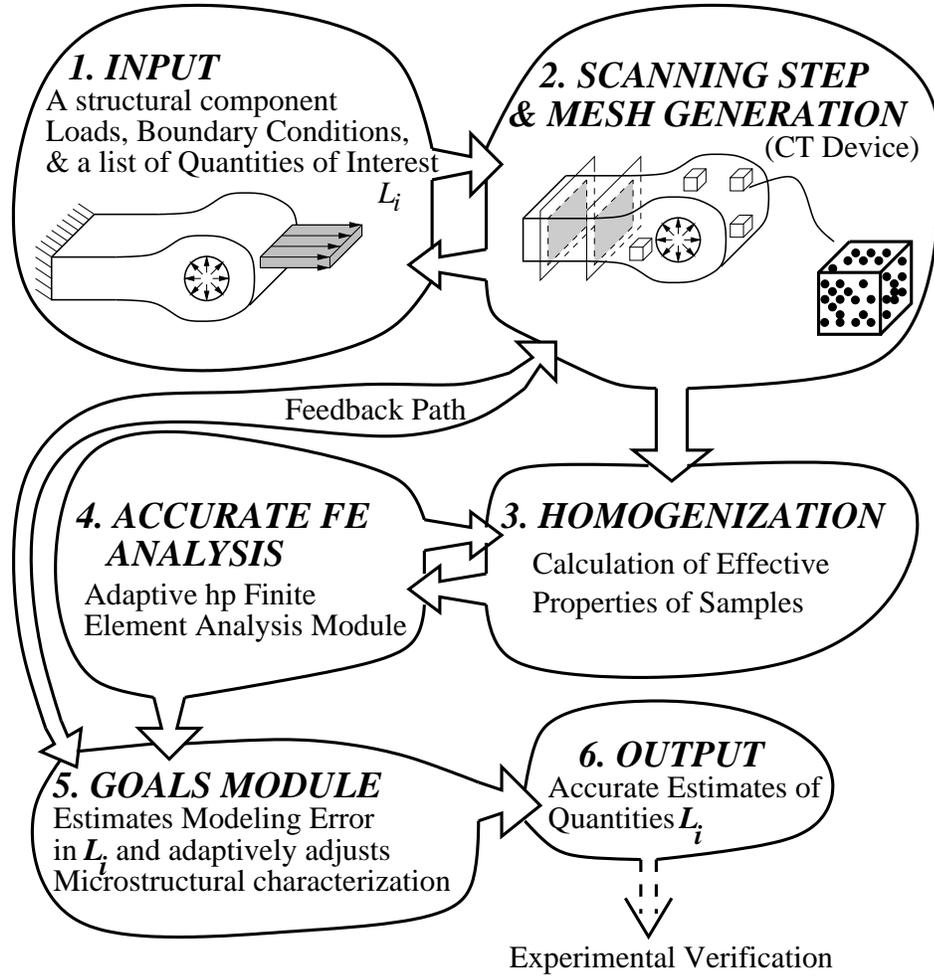
Figure 2.1: Schematic of the adaptive modeling strategy.

6. The output is the set of quantities of interest; when possible, these results are verified by independent physical experiments.

   In general, there is feedback between steps 1 and 2, and step 5 and 2: from 2 to 1, as imaging data may be needed to identify quantities of interest, and from 5 to 2, as additional imaging data on microstructure may be needed to adaptively improve the accuracy of the model.

2.2 Computerized tomography and mesh generation

Tomography refers to the cross-sectional imaging of an object from data collected by subjecting the object to electro-magnetic radiation from different directions. X-ray CT refers to the use of X-rays to analyze a given cross-section. Since a 3-D description of an object can, in principle, be assembled from a series of 2-D (planar) descriptions, we focus our attention here on the analysis of 2-D CT

images.

The output from the CT analysis of a material is a planar map of the attenuation coefficient[1] of the material, given as a discrete function over a cartesian grid in the form of a gray-scale. Each element of this grid is termed a *pixel*.

For a detailed discussion of CT imaging, see [16]. In the current study, we use a X-ray CT device with a resolution of about 10 μm.

The Segmentation Problem

Segmentation refers to the process of identifying individual constituents and the interfaces between constituents in a material from its gray-scale image. For a two-phase material, this amounts to classifying a pixel as belonging to one of two materials and grouping like pixels. The range of the gray-scale depends on the number of "bits" of information stored per pixel; an 8-bit gray-scale, for instance, ranges from 0 to 255 in value. Thus, to segment an image of a two-phase material, it is necessary to select a threshold value, above which a pixel is considered to be of one material, and below which the pixel belongs to the other material. For multi-phase materials, several threshold values must be identified to delineate various micromechanical constituents. The choice of this threshold value, evidently, has a significant impact on the outcome of the segmentation process. Once the threshold parameter, also known as an isovalue, is selected, the isocontour – the surface on which the value of the gray-scale function equals the selected isovalue – has to be determined. Here, we comment only about the selection of the isovalue; a discussion of isocontouring is beyond the scope of this paper.

Our approach to the segmentation problem consists of augmenting the Contour Spectrum approach proposed by Bajaj et al [2] with known information about the specimen. In particular, we use information about the volume fraction of the specimen to arrive at a segmented image. In the current approach, the threshold value is varied from one end of the gray-scale to the other, and for each value of the threshold parameter, the volume fraction of the resulting segmented image is computed. Then, the threshold value that predicts the known volume fraction of the specimen is selected as the true threshold value. For more details on the Contour Spectrum approach and the isocontouring algorithms used in this work, see [2, 3].

---

[1]The attenuation coefficient of a material, as the name suggests, measures the attenuation or the loss of intensity of an X-ray while passing through the material as a result of the photoelectric absorption effect and the Compton effect. In general, the attenuation coefficient of a material can be related accurately to material densities.
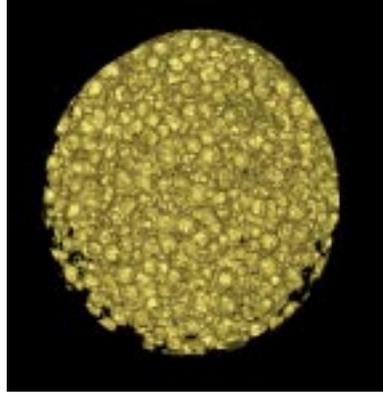
Figure 2.2: CT image of sample of rock.

As an example, we consider a rock sample whose CT image is shown in Fig. 2.2. Figure 2.3 (a) shows the Contour Spectrum for this sample, with the blue line representing the variation of the volume fraction with the threshold parameter. Finally, in Figs. 2.3 (b) and (c), we show the resulting segmented images for two different choices of the threshold parameter. As is seen from these figures, an arbitrary selection of the threshold parameter can result in a highly inaccurate picture of the microstructure.

Mesh Generation

The final step in using the original CT image is to generate a mesh on which computations can be performed. We now discuss the generation of 2-D meshes of quadrilateral elements for the case of two-phase composites with cylindrical (circular) inclusions embedded in a matrix material. Mesh generation using triangles for more general inclusions has been performed but will not be considered here. For the purpose of the present discussion, we assume that the centers and the radii of the inclusions have been extracted from the segmented image of the specimen.

The main steps in the mesh generation algorithm are as follows:

1. Construct the weighted Voronoi diagram of centers of the circles, with the weight proportional to radius.

2. Merge short edges of the Voronoi diagram based on user-specified threshold. If this leads to an intersection between an edge and a circle, the operation is not performed.

3. Divide faces of Voronoi cells into n pieces (n is user-specified) by introducing vertices (nodes) on the faces. If a face is shared, this division must be unique.
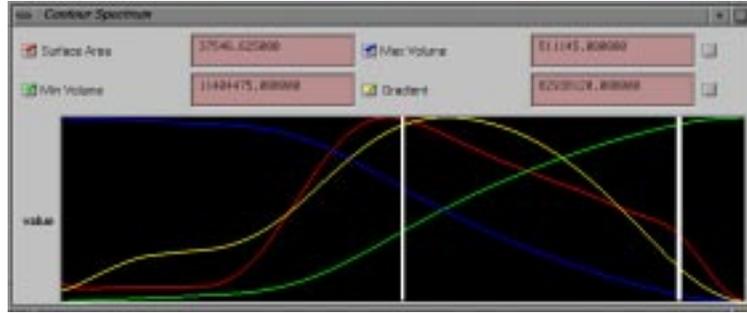
(a)
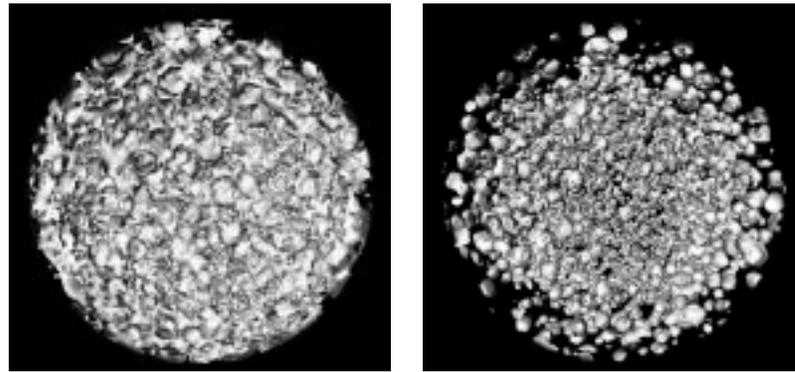
(b)   (c)

Figure 2.3: (a) Contour Spectrum for the rock sample of Fig. 2.2. (b) and (c) Segmented images for two different choices of the threshold parameter.

4. Connect centers of circles to the vertices on the faces, leading to a collection of super-triangles. Each such super-triangle is divided into a quadrilateral and a triangle by the presence of the circle. Optionally, additional division can be performed in the radial direction.

5. Subdivide each triangle into three quadrilaterals.

6. Smooth the mesh using, say, centroid smoothing, wherein an internal node is relocated to the geometric center, or centroid, of the polygon comprised of the elements containing the internal node.

In Fig. 2.4, some of the steps in the algorithm are illustrated for a simple case. More involved example problems employing the above algorithm will be presented later. It should be noted that the approach described here is not related to the VCFEM of Ghosh and Moorthy (e.g. [11]). Here, the Voronoi cells are not finite elements; they are partitions that encapsulate inclusions and provide a
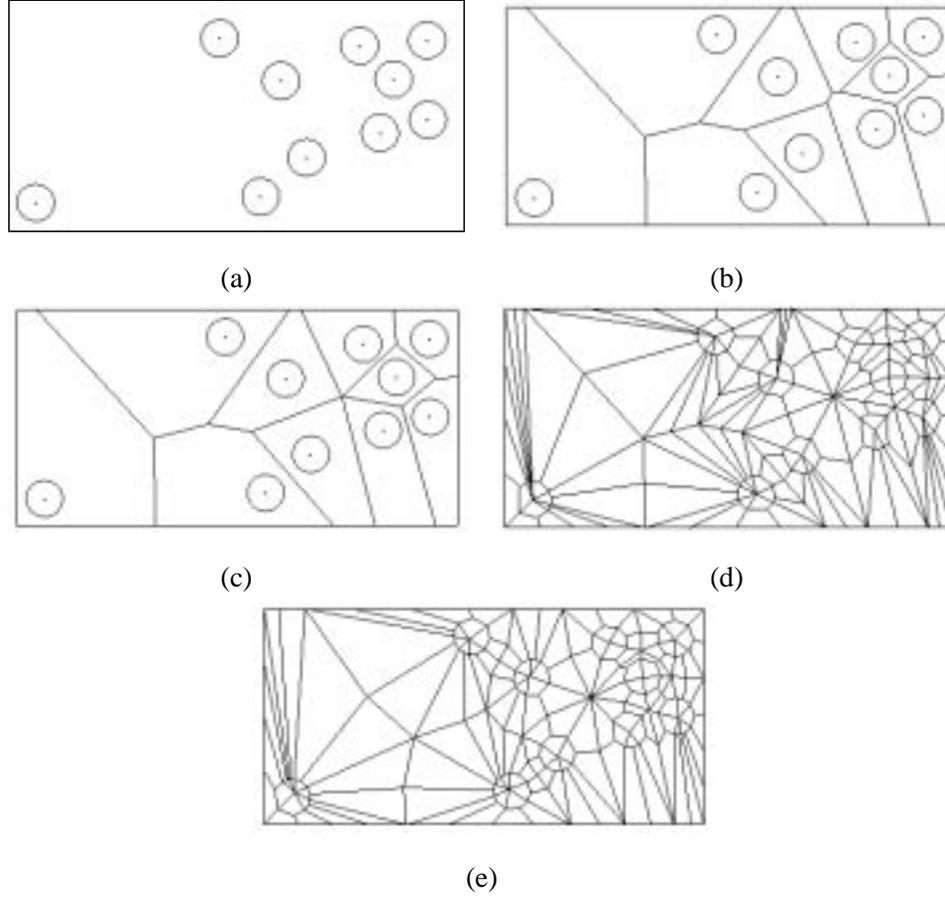
Figure 2.4: Meshing Algorithm: (a) Original configuration of the inclusions, (b) Weighted Voronoi diagram for this configuration, (c) Voronoi diagram after removal of short edges, (d) Initial mesh, and (e) Mesh after 5 iterations of centroid smoothing.

geometric description of needed for automatic meshing. These meshes are adapted using a parallel hp adaptive finite element system discussed later.

2.3 The mathematical model

The actual response of the material body under study can be thought of as being depicted by an abstract mathematical model, provided by the equations of linear elasticity for heterogeneous bodies:

$$
\begin{aligned}
-\mathrm{div}\Big(\mathbf{E}(\mathbf{x})\boldsymbol{\nabla}\mathbf{u}(\mathbf{x})\Big) &= \mathbf{f}(\mathbf{x}), \ \mathbf{x} \in \Omega \\
\mathbf{n}(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x})\boldsymbol{\nabla}\mathbf{u}(\mathbf{x}) &= \mathbf{t}(\mathbf{x}), \ \mathbf{x} \in \Gamma_t \\
\mathbf{u}(\mathbf{x}) &= \mathcal{U}(\mathbf{x}), \ \mathbf{x} \in \Gamma_u
\end{aligned}
\tag{2.1}
$$

47

Here, $\Omega \subset \mathbb{R}^N (N = 1, 2, 3)$ is the open interior of the material body under consideration, Γ_t and Γ_u are portions of the boundary $\partial\Omega$ of Ω on which tractions $\mathbf{t}$ and displacements $\mathcal{U}$, respectively, are prescribed, $\mathbf{f}$ is the body force per unit volume, and $\mathbf{n}$ is the unit exterior normal to $\partial\Omega$. In (2.1):

$\mathbf{E}(\mathbf{x})$ = the value of the elasticity tensor $\mathbf{E}$ at a point $\mathbf{x} \in \Omega$

$\mathbf{u}(\mathbf{x})$ = the value of the fine-scale displacement field $\mathbf{u}$ at a point $\mathbf{x} \in \Omega$

It will always be assumed that $\mathbf{E}$ has the standard symmetries and ellipticity properties of elasticity tensors: $E_{ijkl}(\mathbf{x}) = E_{jikl}(\mathbf{x}) = E_{ijlk}(\mathbf{x}) = E_{klij}(\mathbf{x})$, for almost every $\mathbf{x} \in \Omega, 1 \leq i, j, k, l \leq N$; there exist constants $\alpha_0, \alpha_1 > 0$ such that for almost all $\mathbf{x} \in \Omega$,

$$\alpha_0 \varepsilon_{ij}\varepsilon_{ij} \leq E_{ijkl}(\mathbf{x})\varepsilon_{ij}\varepsilon_{kl} \leq \alpha_1 \varepsilon_{ij}\varepsilon_{ij} \tag{2.2}$$

for any $\varepsilon_{ij} \in \mathbb{R}^N \times \mathbb{R}^N, \varepsilon_{ij} = \varepsilon_{ji}$, and repeated indices are summed throughout their range, $1 \leq i, j, k, l \leq N$. Owing to the possibly very irregular distribution of multiphase microstructural features within the body, $\mathbf{E}$ will, in general, be a highly oscillatory, rapidly varying function of $\mathbf{x}$.

It is well known that for general domains Ω and general loading and boundary conditions, a classical solution to problem (2.1) does not exist. One is then led to consider a weak or variational form. In the case of (2.1), we have the corresponding weak problem:

$$\text{Find } \mathbf{u} \in \{\hat{\mathbf{u}}\} + \mathbf{V}(\Omega) \text{ such that}$$
$$\mathcal{B}_\Omega(\mathbf{u}, \mathbf{v}) = \mathcal{F}_\Omega(\mathbf{v}) \ \forall \ \mathbf{v} \in \mathbf{V}(\Omega). \tag{2.3}$$

Here, $\mathbf{V}(\Omega)$ is the space of admissible displacements,

$$\mathbf{V}(\Omega) \stackrel{\text{def}}{=} \left\{ \mathbf{v} \in \left(H^1(\Omega)\right)^N : \mathbf{v} = \mathbf{0} \text{ on } \Gamma_u \right\}. \tag{2.4}$$

Also, $\hat{\mathbf{u}}$ is an $(H^1(\Omega))^N$ function whose trace on Γ_u is the Dirichlet data $\mathcal{U}$, and

$$\mathcal{B}_\Omega(\mathbf{u}, \mathbf{v}) \stackrel{\text{def}}{=} \int_\Omega \boldsymbol{\nabla}\mathbf{v} : \mathbf{E}\,\boldsymbol{\nabla}\mathbf{u}\,\mathrm{dx}, \tag{2.5}$$

$$\mathcal{F}_\Omega(\mathbf{v}) \stackrel{\text{def}}{=} \int_\Omega \mathbf{f} \cdot \mathbf{v}\,\mathrm{dx} + \int_{\Gamma_t} \mathbf{t} \cdot \mathbf{v}\,\mathrm{ds}. \tag{2.6}$$

In (2.5), the integrand is an $L^1(\Omega)$ function, $\boldsymbol{\nabla}\mathbf{v} : \mathbf{E}\,\boldsymbol{\nabla}\mathbf{u} = \frac{\partial v_i}{\partial x_j}(\mathbf{x})E_{ijkl}(\mathbf{x})\frac{\partial u_k}{\partial x_l}(\mathbf{x})$ (repeated indices summed; $1 \leq i, j, k, l \leq N, N = 1, 2, 3$). In (2.6), it is implicitly assumed that $\mathbf{f} \in (L^2(\Omega))^N$ and $\mathbf{t} \cdot \mathbf{v}$ is integrable on Γ_t for $\mathbf{v} \in \mathbf{V}(\Omega)$.

The function $\mathbf{E} = \mathbf{E}(\mathbf{x})$ will rarely ever be completely known, and therefore, the fine-scale displacement field $\mathbf{u} = \mathbf{u}(\mathbf{x})$ will also never (or rarely ever) be known. The best that one can usually hope for is that a CT imaging device can be used to *sample* the specimen and determine an approximate restriction of $\mathbf{E}$ to the sampled subdomains.

2.4 Effective properties: the homogenized problem

Since our knowledge of $\mathbf{E}$ is incomplete and since the fine-scale displacement field is unknown and generally impossible to determine, we resort to a classical approach to analyze heterogeneous media: the elasticity tensor $\mathbf{E}$ is replaced by a smeared, or homogenized tensor $\mathbf{E}^0$ of *effective properties*. A large body of work exists on this subject and the underlying mathematical theory, and we refer to standard references for full details: see, for example, Bensoussan, Lions, and Papanicolaou [4], Sanchez-Palencia [23] and Jikov, Kozlov, and Oleinik [15], or, for an engineering-oriented treatment, Christensen [8].

In our adaptive modeling process, imaging data from the samples is used to compute effective properties. This is accomplished by generating a finite element mesh of the microstructure and calling an hp finite element module (discussed below) to calculate effective moduli. Note that the determination of effective properties is not the goal of this analysis; homogenization is only an artifact in a broader computational strategy. Thus, in the present approach,

- $\mathbf{E}^0$ is computed from data supplied by the CT image of samples of the body;

- $\mathbf{E}^0$ may be a constant tensor, or it can vary over a large group of samples, being constant, however, over each sample.

With $\mathbf{E}^0$ known, we replace (2.1) by the *homogenized problem*:

$$
\begin{aligned}
-\mathrm{div}\mathbf{E}^0 \boldsymbol{\nabla} \mathbf{u}^0(\mathbf{x}) &= \mathbf{f}(\mathbf{x}), \ \ \mathbf{x} \in \Omega \\
\mathbf{n}(\mathbf{x}) \cdot \mathbf{E}^0(\mathbf{x}) \boldsymbol{\nabla} \mathbf{u}^0(\mathbf{x}) &= \mathbf{t}(\mathbf{x}), \ \ \mathbf{x} \in \Gamma_t \\
\mathbf{u}^0(\mathbf{x}) &= \mathcal{U}(\mathbf{x}), \ \ \mathbf{x} \in \Gamma_u.
\end{aligned}
\tag{2.7}
$$

The weak form of the homogenized problem (2.7) is:

$$
\begin{aligned}
&\text{Find } \mathbf{u}^0 \in \{\hat{\mathbf{u}}\} + \mathbf{V}(\Omega) \text{ such that} \\
&\mathcal{B}^0_\Omega(\mathbf{u}^0, \mathbf{v}) = \mathcal{F}_\Omega(\mathbf{v}) \ \ \forall \ \mathbf{v} \in \mathbf{V}(\Omega),
\end{aligned}
\tag{2.8}
$$

where

$$
\mathcal{B}^0_\Omega(\mathbf{u}^0, \mathbf{v}) \overset{\text{def}}{=} \int_\Omega \boldsymbol{\nabla} \mathbf{v} : \mathbf{E}^0 \, \boldsymbol{\nabla} \mathbf{u}^0 \, \mathrm{dx},
\tag{2.9}
$$

and $\mathbf{E}^0$ denotes the elasticity tensor for the homogenized problem. The displacement field $\mathbf{u}^0 = \mathbf{u}^0(\mathbf{x})$ is called the *homogenized displacement field*. While we do not know the fine-scale field $\mathbf{u}$,

we can nevertheless estimate quite accurately, the *homogenization error*

$$\mathbf{e}^0 = \mathbf{u} - \mathbf{u}^0 \tag{2.10}$$

using methods described in [18, 30], and summarized below.

2.5 Error estimation and adaptive modeling: the GOALS algorithm

We now review the major results regarding modeling error estimation and adaptive modeling presented in [18]. First, we review global energy bounds on the modeling error $\mathbf{e}^0$. Next, bounds on the modeling error in local quantities of interest are briefly discussed. This is followed by a review of the Goal Oriented Adaptive Local Solution (GOALS) algorithm [18], a procedure for delivering accurate values of quantities of interest.

Energy Error Estimates

Let

$$\mathcal{I}_0 = (\mathbf{I} - \mathbf{E}^{-1}\,\mathbf{E}^0), \tag{2.11}$$

where $\mathbf{I}$ is the identity tensor. Next, for $\mathbf{g} \in \mathbf{V}(\Omega)$, define the associated linear *residual functional* $\mathcal{R}_{\mathbf{g}} : \mathbf{V}(\Omega) \to \mathbb{R}$,

$$\mathcal{R}_{\mathbf{g}}(\mathbf{v}) = -\int_\Omega \boldsymbol{\nabla}\mathbf{v} : \mathbf{E}\,\mathcal{I}_0\boldsymbol{\nabla}\mathbf{g}\,d\mathbf{x}, \quad \mathbf{v} \in \mathbf{V}(\Omega). \tag{2.12}$$

Finally, define the energy norm of an admissible function $\mathbf{v} \in \mathbf{V}(\Omega)$,

$$\|\mathbf{v}\|_{E(\Omega)} \stackrel{\text{def}}{=} \sqrt{\mathcal{B}_\Omega(\mathbf{v},\mathbf{v})}, \tag{2.13}$$

where $\mathcal{B}_\Omega(\cdot,\cdot)$ is the bilinear form defined in (2.5).

Theorem 2.1 *Let $\mathbf{u}$ and $\mathbf{u}^0$ be the solutions to problems (2.3) and (2.8) respectively. Then the following holds:*

$$\zeta_{\text{low}} \leq \|\mathbf{e}^0\|_{E(\Omega)} = \|\mathbf{u} - \mathbf{u}^0\|_{E(\Omega)} \leq \zeta_{\text{upp}}, \tag{2.14}$$

where

$$\zeta_{\text{low}} \stackrel{\text{def}}{=} \frac{|\mathcal{R}_{\mathbf{u}^0}(\mathbf{u}^0)|}{\|\mathbf{u}^0\|_{E(\Omega)}}, \quad \zeta_{\text{upp}} \stackrel{\text{def}}{=} \left\{ \int_\Omega \left(\mathcal{I}_0\boldsymbol{\nabla}\mathbf{u}^0 \right) : \mathbf{E}\,\mathcal{I}_0\boldsymbol{\nabla}\mathbf{u}^0\,d\mathbf{x} \right\}^{\frac{1}{2}}. \tag{2.15}$$

$\square$

For proofs, see [30] and [22].

Local Error Estimates

Let L be a quantity of interest that is characterized by a continuous linear functional $\mathbf{V}(\Omega)$, $L \in \mathbf{V}'(\Omega)$. The problem

$$
\boxed{
\begin{array}{c}
\text{Find } \mathbf{w} \in \mathbf{V}(\Omega) \text{ such that} \\
\mathcal{B}_\Omega(\mathbf{v}, \mathbf{w}) = L(\mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{V}(\Omega)
\end{array}
}
\tag{2.16}
$$

is referred to as the *adjoint fine-scale problem*. The solution $\mathbf{w}$ to the adjoint fine-scale problem is termed the *fine-scale influence function*. The homogenized version of this problem is referred to as the *adjoint homogenized problem* and reads

$$
\boxed{
\begin{array}{c}
\text{Find } \mathbf{w}^0 \in \mathbf{V}(\Omega) \text{ such that} \\
\mathcal{B}_\Omega^0(\mathbf{v}, \mathbf{w}^0) = L(\mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{V}(\Omega).
\end{array}
}
\tag{2.17}
$$

The solution to this problem is the *homogenized influence function*. In what follows, we sometimes refer to the problems (2.3) and (2.8) as the *primal fine-scale problem* and *primal homogenized problem*, respectively. Functions $\mathbf{w}$ and $\mathbf{w}^0$ exist and are uniquely defined. The modeling error in the influence function is given by

$$
\bar{\mathbf{e}}^0 \overset{\text{def}}{=} \mathbf{w} - \mathbf{w}^0.
\tag{2.18}
$$

Also, $\bar{\mathbf{e}}^0$ satisfies the following relationship:

$$
\bar{\zeta}_{\text{low}} \leq \|\bar{\mathbf{e}}^0\|_{E(\Omega)} = \|\mathbf{w} - \mathbf{w}^0\|_{E(\Omega)} \leq \bar{\zeta}_{\text{upp}}
\tag{2.19}
$$

where

$$
\bar{\zeta}_{\text{low}} \overset{\text{def}}{=} \frac{|\mathcal{R}_{\mathbf{w}^0}(\mathbf{w}^0)|}{\|\mathbf{w}^0\|_{E(\Omega)}}; \quad \bar{\zeta}_{\text{upp}} \overset{\text{def}}{=} \left\{ \int_\Omega \left(\mathcal{I}_0 \boldsymbol{\nabla} \mathbf{w}^0 \right) : \mathbf{E}\, \mathcal{I}_0 \boldsymbol{\nabla} \mathbf{w}^0 \, \mathrm{d}\mathbf{x} \right\}^{1/2}.
\tag{2.20}
$$

We now state the main result on the estimation of modeling error in quantities of interest.

Theorem 2.2 *Let $\mathbf{u}^0$ and $\mathbf{w}^0$ be the solutions to problems (2.8) and (2.17), respectively. Then,*

$$
\eta_{\text{low}} \leq L(\mathbf{e}^0) \leq \eta_{\text{upp}}
\tag{2.21}
$$

where

$$
\eta_{\text{low}} \overset{\text{def}}{=} \frac{1}{4}(\eta_{\text{low}}^+)^2 - \frac{1}{4}(\eta_{\text{upp}}^-)^2 + \mathcal{R}_{\mathbf{u}^0}(\mathbf{w}^0),
\tag{2.22}
$$

51

$$\eta_{\text{upp}} \stackrel{\text{def}}{=} \frac{1}{4}(\eta_{\text{upp}}^+)^2 - \frac{1}{4}(\eta_{\text{low}}^-)^2 + \mathcal{R}_{\mathbf{u}^0}(\mathbf{w}^0), \tag{2.23}$$

with arbitrary $s \in \mathbb{R}^+$,

$$\eta_{\text{upp}}^\pm \stackrel{\text{def}}{=} \sqrt{s^2\zeta_{\text{upp}}^2 \pm 2\int_\Omega \mathcal{I}_0\boldsymbol{\nabla}\mathbf{u}^0 : \mathbf{E}\,\mathcal{I}_0\boldsymbol{\nabla}\mathbf{w}^0\,d\mathbf{x} + s^{-2}\bar{\zeta}_{\text{upp}}^2}, \tag{2.24}$$

and

$$\eta_{\text{low}}^\pm \stackrel{\text{def}}{=} \frac{|\mathcal{R}_{s\mathbf{u}^0 \pm s^{-1}\mathbf{w}^0}(\mathbf{u}^0 + \theta^\pm\mathbf{w}^0)|}{\|\mathbf{u}^0 + \theta^\pm\mathbf{w}^0\|_{E(\Omega)}}, \tag{2.25}$$

where ζ_{upp} and $\bar{\zeta}_{\text{upp}}$ are defined by (2.15) and (2.20), respectively, and $\theta^\pm$ is given by

$$\theta^\pm = \frac{\mathcal{B}_\Omega(\mathbf{u}^0,\mathbf{w}^0)\mathcal{R}_{\mathbf{u}^0}(s\mathbf{u}^0 \pm s^{-1}\mathbf{w}^0) - \mathcal{B}_\Omega(\mathbf{u}^0,\mathbf{u}^0)\mathcal{R}_{\mathbf{w}^0}(s\mathbf{u}^0 \pm s^{-1}\mathbf{w}^0)}{\mathcal{B}_\Omega(\mathbf{u}^0,\mathbf{w}^0)\mathcal{R}_{\mathbf{w}^0}(s\mathbf{u}^0 \pm s^{-1}\mathbf{w}^0) - \mathcal{B}_\Omega(\mathbf{w}^0,\mathbf{w}^0)\mathcal{R}_{\mathbf{u}^0}(s\mathbf{u}^0 \pm s^{-1}\mathbf{w}^0)}. \tag{2.26}$$

$\square$

See [22] for the proof. The scalar parameter s is a scaling factor and its optimal value is $s^* = \sqrt{\bar{\zeta}_{\text{upp}}/\zeta_{\text{upp}}}$. Also, in our numerical experiments, we employ the following *estimate* of the modeling error in the quantity of interest:

$$L(\mathbf{e}^0) \approx \eta_{\text{est}} \stackrel{\text{def}}{=} \frac{1}{4}(\eta_{\text{upp}}^+)^2 - \frac{1}{4}(\eta_{\text{upp}}^-)^2 + \mathcal{R}_{\mathbf{u}^0}(\mathbf{w}^0). \tag{2.27}$$

Definition of the Local Fine-scale Problem

For the purpose of simplicity, we assume that the quantity of interest L is a functional of the form

$$L(\mathbf{v}) = \int_\omega l(\mathbf{v})\,d\mathbf{x}, \tag{2.28}$$

where l is a linear map $l : \mathbf{V}(\Omega) \to L^1_{\text{loc}}(\Omega)$. Here, ω is some subset of the domain Ω. Functionals of other types can be accommodated very easily in our approach. Let Ω_L be a subset of the domain Ω that contains ω: $\omega \subset \Omega_L$. We shall refer to Ω_L as the functional's "domain of influence", and it's determination will be discussed shortly.

In order to define the local fine-scale problem on Ω_L, we introduce some notation. Let

$$\Gamma_{L_t} \stackrel{\text{def}}{=} \partial\Omega_L \cap \Gamma_t, \quad \Gamma_{L_u} \stackrel{\text{def}}{=} \partial\Omega_L \setminus \Gamma_{L_t}. \tag{2.29}$$

Define the local function space on Ω_L as

$$\mathbf{V}(\Omega_L) = \left\{ \mathbf{v} \in \mathbf{V}(\Omega), \mathbf{v} = \mathbf{0} \text{ on } \Omega \setminus \overline{\Omega_L}, \mathbf{v}|_{\Gamma_{L_u}} = \mathbf{0} \right\}. \tag{2.30}$$

Next, an extension operator $\mathcal{E}_L : \mathbf{V}(\Omega_L) \to \mathbf{V}(\Omega)$ is introduced, defined by:

$$\mathbf{v}_L \in \mathbf{V}(\Omega_L), \ \mathcal{E}_L(\mathbf{v}_L) = \mathbf{v} \text{ such that } \mathbf{v}|_{\Omega_L} = \mathbf{v}_L, \mathbf{v}|_{\Omega \setminus \Omega_L} = \mathbf{0}. \tag{2.31}$$

The restriction of the homogenized solution $\mathbf{u}^0$ to the domain of influence Ω_L is defined as $\mathbf{u}_L^0 \colon \mathbf{u}_L^0 \overset{\text{def}}{=} \mathbf{u}^0|_{\Omega_L}$. Then, the following weak boundary value problem is referred to as the *local fine-scale problem*:

$$\boxed{\begin{aligned} &\text{Find } \tilde{\mathbf{u}}_L \in \{\mathbf{u}_L^0\} + \mathbf{V}(\Omega_L) \text{ such that} \\ &\mathcal{B}_L(\tilde{\mathbf{u}}_L, \mathbf{v}_L) = \mathcal{F}_L(\mathbf{v}_L) \ \ \forall \mathbf{v}_L \in \mathbf{V}(\Omega_L), \end{aligned}} \tag{2.32}$$

where the bilinear and linear forms are defined as

$$\mathcal{B}_L(\tilde{\mathbf{u}}_L, \mathbf{v}_L) \overset{\text{def}}{=} \int_{\Omega_L} \boldsymbol{\nabla}\mathbf{v}_L : \mathbf{E}\,\boldsymbol{\nabla}\tilde{\mathbf{u}}_L \,\mathrm{d}\mathbf{x}, \tag{2.33}$$

and

$$\mathcal{F}_L(\mathbf{v}_L) \overset{\text{def}}{=} \int_{\Omega_L} \mathbf{f} \cdot \mathbf{v}_L \,\mathrm{d}\mathbf{x} + \int_{\Gamma_{L_t}} \mathbf{t} \cdot \mathbf{v}_L \,\mathrm{d}\mathbf{s}, \tag{2.34}$$

respectively. Thus, $\tilde{\mathbf{u}}_L$ is a perturbation to the homogenized solution $\mathbf{u}^0$ on Ω_L that takes into account the fine-scale microstructure. It equals the primal homogenized solution $\mathbf{u}^0$ on the Γ_{L_u} portion of its boundary. Using the extension operator $\mathcal{E}_k$, we define the *locally enhanced* function $\tilde{\mathbf{u}} \in \mathbf{V}(\Omega)$ as:

$$\tilde{\mathbf{u}} \overset{\text{def}}{=} \mathbf{u}^0 + \mathcal{E}_L(\tilde{\mathbf{u}}_L - \mathbf{u}_L^0). \tag{2.35}$$

Finally, it is noted that the modeling error $\mathbf{u} - \tilde{\mathbf{u}}$ can also be estimated – both globally in the energy norm, and locally in the quantity of interest – using the results presented in [18]. For extensions of the above results to perforated domains, see [28].

The GOALS Algorithm

The GOALS algorithm provides an adaptive procedure for accurately computing a quantity of interest $L(\mathbf{u})$ by determining the size of its domain of influence. For this purpose, we introduce a partition $\mathcal{P}$ of the domain Ω into cells Θ_k, $1 \le k \le N(\mathcal{P})$, where $N(\mathcal{P})$ is the total number of cells in the partition. The following modeling *error indicators* are used in the GOALS algorithm:

$$\begin{aligned} \zeta_{k,\text{upp}} &\overset{\text{def}}{=} \left\{ \int_{\Theta_k} \mathcal{I}_0 \boldsymbol{\nabla}\mathbf{u}^0 : \mathbf{E}\,\mathcal{I}_0 \boldsymbol{\nabla}\mathbf{u}^0 \,\mathrm{d}\mathbf{x} \right\}^{\frac{1}{2}} \\ \bar{\zeta}_{k,\text{upp}} &\overset{\text{def}}{=} \left\{ \int_{\Theta_k} \mathcal{I}_0 \boldsymbol{\nabla}\mathbf{w}^0 : \mathbf{E}\,\mathcal{I}_0 \boldsymbol{\nabla}\mathbf{w}^0 \,\mathrm{d}\mathbf{x} \right\}^{\frac{1}{2}}, \end{aligned} \tag{2.36}$$

and

$$\beta_k \stackrel{\text{def}}{=} \zeta_{k,\text{upp}}\bar{\zeta}_{k,\text{upp}} + \zeta_{k,\text{upp}}\|\mathbf{w}^0\|_{E(\Theta_k)}. \tag{2.37}$$

The indicators $\zeta_{k,\text{upp}}$ and $\bar{\zeta}_{k,\text{upp}}$ are contributions of a cell to the modeling errors estimates ζ_{upp} and $\bar{\zeta}_{\text{upp}}$ (recall (2.15) and (2.20)), respectively. The measure β_k is an indicator of the local contribution of a cell to the modeling error in the quantity of interest.

The outline of the GOALS algorithm is as follows:

Step 1. Initialization. Given the initial data Ω, Γ_u, Γ_t, $\mathbf{E}$, $\mathbf{f}$ and $\mathbf{t}$, construct a non-overlapping partition of the domain $\mathcal{P} = \{\Theta_k\}, k = 1, 2 \ldots N(\mathcal{P})$. Specify error tolerance parameters α_{TOL} and δ_{TOL}, $0 < \delta_{\text{TOL}} < 1$.

Step 2. Homogenization. Compute the homogenized elasticity tensor $\mathbf{E}^0$. Solve the primal homogenized problem (2.8) for $\mathbf{u}^0$ and the adjoint homogenized problem (2.17) for $\mathbf{w}^0$.

Step 3. Modeling Error Estimation. Compute error indicators ζ_k, $\bar{\zeta}_k$ and β_k for $1 \leq k \leq N(\mathcal{P})$, using (2.36) and (2.37). Estimate the modeling error in the quantity of interest using Theorem 3.1. Denote this estimate by η_{est}.

Step 4. Tolerance Test. If $\eta_{\text{est}} \leq \alpha_{\text{TOL}} \times L(\mathbf{u}^0)$, STOP.

Step 5. Domain of Influence. Determine initial guess for "domain of influence" Ω_L as all the cells that intersect ω, the region over which the quantity of interest is defined:

$$\overline{\Omega_L} = \overline{\cup_{j \in \mathcal{J}}\Theta_j} \quad \mathcal{J} \stackrel{\text{def}}{=} \{j : \Theta_j \cap \omega \neq \emptyset\}. \tag{2.38}$$

Compute the quantities ζ_L, $\bar{\zeta}_L$, and β_L:

$$\zeta_L \stackrel{\text{def}}{=} \left\{\sum_{k \in \mathcal{J}} \zeta_{k,\text{upp}}^2\right\}^{\frac{1}{2}}, \ \bar{\zeta}_L \stackrel{\text{def}}{=} \left\{\sum_{k \in \mathcal{J}} \bar{\zeta}_{k,\text{upp}}^2\right\}^{\frac{1}{2}}, \ \beta_L \stackrel{\text{def}}{=} \zeta_L \bar{\zeta}_L + \zeta_L\|\mathbf{w}^0\|_{E(\Omega_L)} \tag{2.39}$$

Step 6. Update Domain of Influence. Determine the "bad neighbors" of Ω_L, i.e., if $\beta_i > \delta_{\text{TOL}} \times \frac{|\Theta_i|}{|\Omega_L|} \times \beta_L$, mark Θ_i as bad and update Ω_L:

$$\Omega_L \leftarrow \Omega_L \cup \{\text{ bad neighbors }\}. \tag{2.40}$$

Update the quantities ζ_L, $\bar{\zeta}_L$, and β_L.

Step 7. Solution of Local Problem. Solve local problem (2.32) on Ω_L for $\tilde{\mathbf{u}}_L$. Construct the locally enhanced solution $\tilde{\mathbf{u}} \in \mathbf{V}(\Omega)$ using (2.35).

Step 8. Estimate Modeling Error. Estimate the modeling error $L(\mathbf{u} - \tilde{\mathbf{u}})$ and denote the estimate by η_{est}. If $\eta_{\text{est}} \leq \alpha_{\text{TOL}} \times L(\tilde{\mathbf{u}})$, STOP. ELSE, GOTO Step 6.

2.6 Parallel hp adaptive FEM for material modeling

The adaptive modeling process described above requires the solution of three different boundary value problems (BVPs): (1) the BVP on the unit cell Y for the computation of the homogenized elasticity tensor $\mathbf{E}^0$, (2) the homogenized primal and adjoint problems for $\mathbf{u}^0$ and $\mathbf{w}^0$, and (3) the local fine-scale problems on domains of influence Ω_L for $\tilde{\mathbf{u}}_L$. Additionally, the primal and adjoint fine-scale problems are sometimes solved for $\mathbf{u}$ and $\mathbf{w}$, respectively, to obtain reference solutions so that the accuracy of the modeling error estimates can be verified and the adaptive modeling procedure validated.

In order to isolate the modeling error, it is necessary to solve these boundary value problems with very high accuracy. This in turn requires the solution of very large sparse systems of equations. Also, the modeling algorithm imposes certain requirements on the implementation. To list a few,

- Both the global and local error estimates involve the computation of integrals over the domain ω. These integrals have to be computed with high accuracy.

- Boundary conditions for the local fine-scale problems need to be extracted from the homogenized solution.

- In order to handle the output from the meshing code, the representation of element geometry has to be well-separated from the approximation shape functions.

With these requirements in mind, a two-dimensional *parallel* and *adaptive* finite element solver was developed for the implementation of the adaptive modeling procedure. The main features of the code are:

1. *Language*: The code is written in C++ and uses the MPICH version of the Message Passing Interface (MPI) [12, 17].

2. *Adaptivity*: Mesh refinement is done using 1-irregular divisions of elements. Hierarchical shape functions are used for p adaptivity. For a discussion of hp adaptive finite elements, see [25].

3. *Parallelism*: The code is designed to run on distributed memory machines. The notion of Space Filling Curves (SFCs) is used to perform partitioning of the domain and achieve load-balancing [24, 10].

4. *Solution strategy*: In the current version, the degrees of freedom corresponding to the bubble functions are first eliminated, and the resulting system of equations is solved using either (a) a preconditioned Conjugate Gradient algorithm or (b) the SParse Object-Oriented Linear Equations Solver (SPOOLES) package, a public domain software [27].

5. *Datastructure*: A Hashtable based datastructure is used for storing element, node and DOF classes. Very general classes have been implemented for storing material and boundary condition data.

6. *Organization, structure and others*: The code is organized as a library-style collection of routines, with the user supplying a few routines that specify the problem parameters. Post-processing is user defined. User-specified integration rules can be used instead of the default integration routines. Additionally, the code is capable of running in batch and interactive modes.

The code has been tested on (a) cluster of PCs running Linux, (b) cluster of SGI workstations running IRIX 6.5 and (c) an IBM SP running AIX 4.1.4.

2.7 Numerical experiments and examples

We now reproduce some representative numerical results from [29]. The first example demonstrates the integration of two important technologies discussed in this paper: imaging and meshing techniques, and adaptive material modeling. We study the deformation of a fabricated composite material that has a single layer of glass beads (E=69GPa, ν=0.22) distributed in an epoxy matrix(E=4.6GPa, ν=0.36). The dimensions of the specimen are shown in Fig. 2.5. The average diameter of the glass beads is 800 μm. Images of the specimen were obtained both with a CT device (resolution=10μ) and an optical microscope (resolution=1μ).
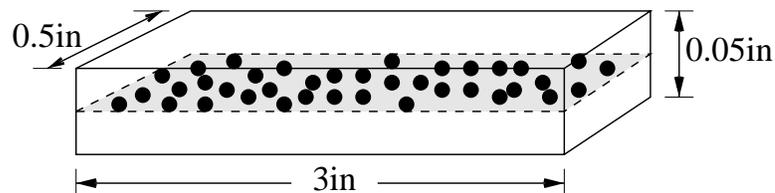
Figure 2.5: Dimensions of the epoxy-glass specimen.

We consider the two-dimensional problem of finding the response of the above specimen to

compressive loads, as shown in Fig. 2.6. We pick, as a quantity of interest, the σ_{xx} component of the stress tensor averaged over the inclusion denoted by ω in Fig. 2.6.
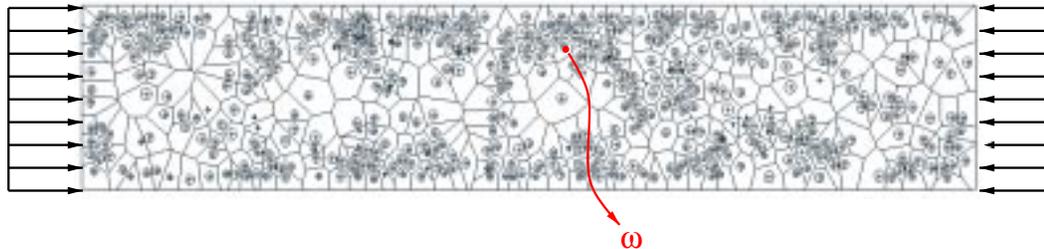
Figure 2.6: Schematic for the 2-D problem of manufactured specimen under compressive loads.

To compute a reference solution $\mathbf{u}$, the meshing algorithm presented in Section 2.2 is used to generate a mesh. The resulting initial mesh is shown in Fig. 2.7 (a). The mesh after three iterations of centroid smoothing is shown in Fig. 2.7 (b).
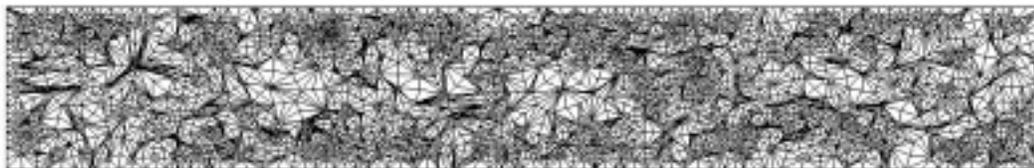
(a)
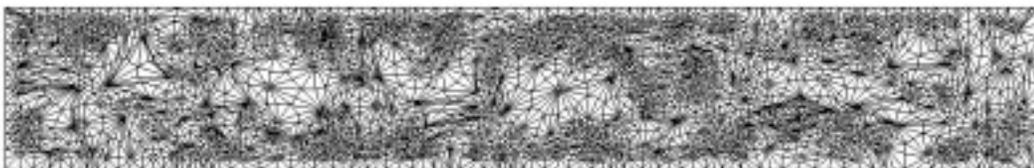
(b)

Figure 2.7: (a) Initial mesh, and (b) Mesh after smoothing.

Next, the homogenized primal and adjoint solutions $\mathbf{u}^0$ and $\mathbf{w}^0$ are obtained using the hp finite element code. The modeling error indicators are then computed and the adaptive procedure is carried out. The sequence of domains of influence and the resulting modeling errors are shown in Fig. 2.8. As can be seen, the quantity of interest can be predicted accurately using only local microstructural information.

The next numerical experiment deals with the analysis of a wrench made of a particulate heterogeneous material, modeled as a two-dimensional object. In this example, the internal microstructure of the body is not known completely. However, as pointed out in [29], it is possible to predict local quantities of interest without knowing the microstructure throughout the domain.

(a) % Error = 17.0   (b) % Error = 9.05   (c) % Error = 2.1
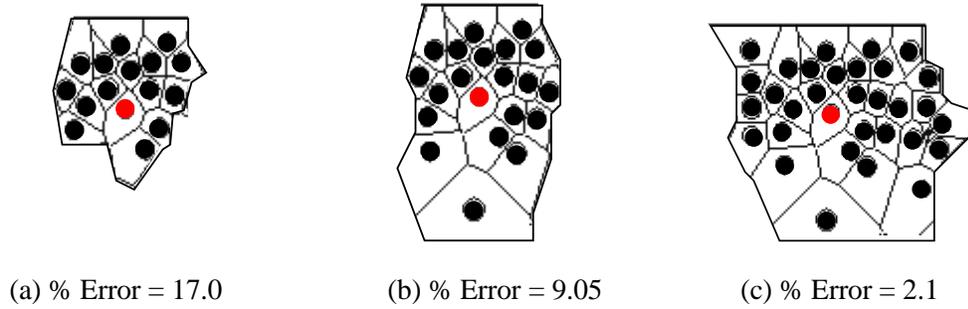
Figure 2.8: Results of the adaptive modeling procedure: domains of influence and resulting modeling errors.

The objective is to study the microscale solution at one corner of the hexagonal part of the boundary of the wrench shown in Fig. 2.9. The hexagonal part of the boundary is constrained, whereas the circular part of the interior boundary is loaded as shown. The internal microstructure of the wrench is assumed to be known only in the vicinity of the corner of interest, in a circular area of radius 0.04 in. The microstructure in this region consists of randomly distributed circular inclusions, with a local volume fraction of 0.4. Based on the number of inclusions in this region and based on the area of the wrench, it is estimated that the body has about 128,000 inclusions. Judging from results obtained in previous calculations, an adapted hp-FEM mesh of the entire structure sufficient to produce global solutions within 1% error in an energy norm would require a computational model with on-the-order-of one billion degrees of freedom.

The quantity of interest is taken to be the average σ_{yy} stress on the inclusion ω, shown in red. The material properties are taken to be $E = 100\text{GPa}$, $\nu = 0.2$ for the matrix material, and $E = 1000\text{GPa}$, $\nu = 0.2$ for the inclusions. The body is homogenized using the Hashin-Shtrikman lower bound. Also, for the homogenized problem, the domain is slightly modified by ignoring the fillets. This simplification is shown in Fig. 2.10 (a), marked by arrows. This, of course, results in artificial corners in the domain, and hence leads to singularities in the homogenized solution. This is reflected in the plot of σ_{yy} component of the stress tensor, shown in Fig. 2.10 (b).

Next, the homogenized influence function $\mathbf{w}^0$ is computed for the specified quantity of interest, and the modeling error in the quantity of error is estimated. The relative modeling error $L(\mathbf{e}^0)/L(\mathbf{u}^0)$ is found to be 0.49. The error indicators are then computed and two steps of the adaptive modeling algorithm are carried out. The domains of influence and the resulting estimated modeling errors are shown in Fig. 2.11. Note that for the local fine-scale problem, the fillet at the vertex of the hexagon is not ignored (Fig. 2.11 (b)). Here, the GOALS adaptive modeling algorithm

58
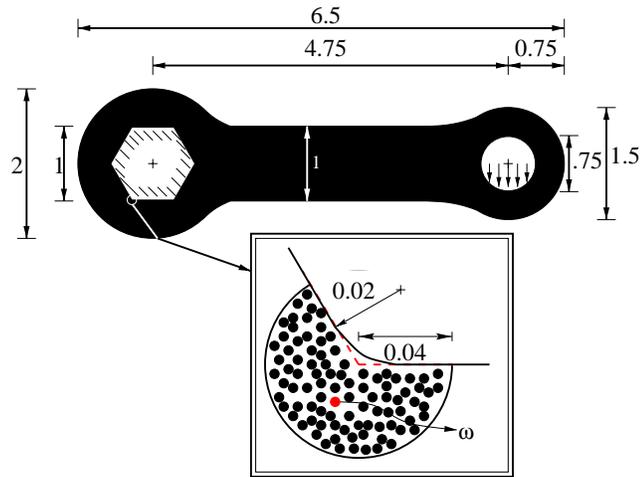
Figure 2.9: Schematic of the wrench. All dimensions are in inches.

(a)

(b)
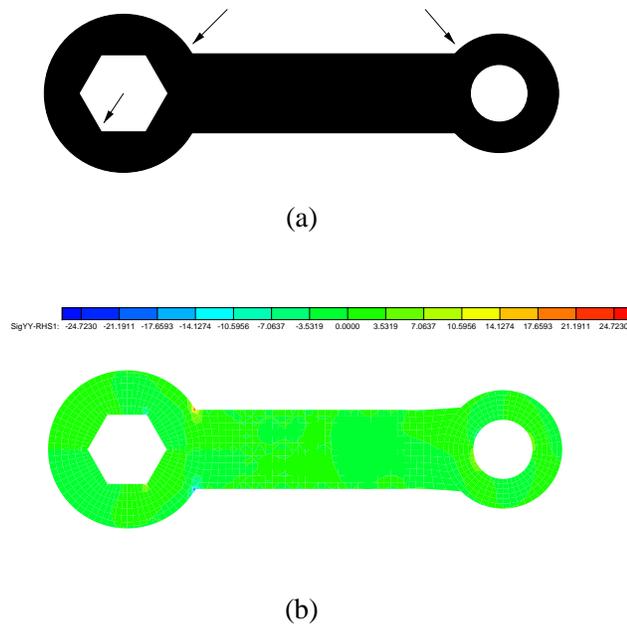
Figure 2.10: (a) Simplified domain for the primal homogenized problem; Ignoring fillets (shown by arrows) results in artificial corners, (b) σ_{yy} component of the homogenized stress field (GPa).

(a) Estimated final error = 12.2%.   (b) Estimated final error = 4.2%

Figure 2.11: Domains of influence and the resulting (estimated) modeling errors in the local solutions.

allows for the prediction of the quantity of interest to within an estimated 4%.

3. Global Estimates of Modeling Error in Nonlinear Continuum Mechanics

As a general example of a basis for hierarchical modeling, we consider a large class of physical phenomena that we assume can be captured by the general equations of nonlinear continuum mechanics. If this assumption proves to be invalid in actual predictions, then a larger class of models must be defined which includes models of non-continuum effects, such as, for example, those of molecular dynamics or kinetic theory. As an example of the nonlinear continuum class of models, consider those characterized by the (class of) weak initial-value problems given in Fig. 3.1.

Here Ω_0 is the reference configuration of a material body (an open, bounded region in $\mathbb{R}^n$, $n = 1, 2$, or 3) with boundary $\partial\Omega_0$ consisting of portions Γ_D^0, Γ_N^0 $\left(\partial\Omega_0 = \overline{\Gamma_D \cup \Gamma_N}\right)$, ρ_0 is the mass density in the reference configuration, $\boldsymbol{b}$ the body force per unit mass, $\boldsymbol{g}$ the traction on Γ_N, $\overline{\boldsymbol{U}}$ is a function in V that produces the prescribed displacement on Γ_D^0, and $\boldsymbol{U}_0, \boldsymbol{V}_0$ are initial displacements and velocities, respectively. The displacement gradient history $\boldsymbol{\nabla}\boldsymbol{u}^t(s) = \{\boldsymbol{\nabla}\boldsymbol{u}(\boldsymbol{X}, t-s)$, $\boldsymbol{X} \in \Omega_0$, $t \geq s \geq 0\}$ depends upon the material point $\boldsymbol{X} \in \Omega_0$ (a.e.), and times in the interval $[0, t]$; $dX = dX_1 dX_2 dX_3$ (for $n = 3$); $\boldsymbol{S}$ is the second Piola-Kirchhoff stress, determined by a constitutive functional $\boldsymbol{\mathcal{S}}(\cdot)$ on the histories of $\boldsymbol{\nabla}\boldsymbol{u}$ and of a set of additional variables $\boldsymbol{A}$, which may include the temperature and various state/internal variables. In (3.1c), the dependence of var-

<div style="border:1px solid black;">

Model $\mathcal{M}$

Given the set of data,

$$\mathcal{D} = \left\{ \Omega_0, \Gamma_D^0, \Gamma_N^0, \rho_0, \boldsymbol{b}, \boldsymbol{g}, \boldsymbol{U}_0, \boldsymbol{V}_0 \right\}, \tag{3.1a}$$

and the experimentally determined parameters and functions determining the constitutive equation

$$\boldsymbol{S} = \boldsymbol{\mathcal{S}} \left(\boldsymbol{\nabla} \boldsymbol{u}^t(s), \ \boldsymbol{A}^t(s); \ \boldsymbol{X}, t \right), \tag{3.1b}$$

find, for each $t \geq 0$, the displacement field $\boldsymbol{u}(\cdot, t) \in V + \{\overline{\boldsymbol{U}}\}$ such that

$$\int_{\Omega_0} \rho_0 \ddot{\boldsymbol{u}}(t) \cdot \boldsymbol{v} \, dX + \int_{\Omega_0} \boldsymbol{F}(\boldsymbol{u}; t) \, \boldsymbol{\mathcal{S}} \left(\boldsymbol{\nabla} \boldsymbol{u}^t(s), \boldsymbol{A}^t(s); t \right) : \boldsymbol{\nabla} \boldsymbol{v} \, dX$$
$$= \mathcal{F} \left(\boldsymbol{u}(t); \ \boldsymbol{v}, t \right) \quad \forall \boldsymbol{v} \in V \tag{3.1c}$$

</div>
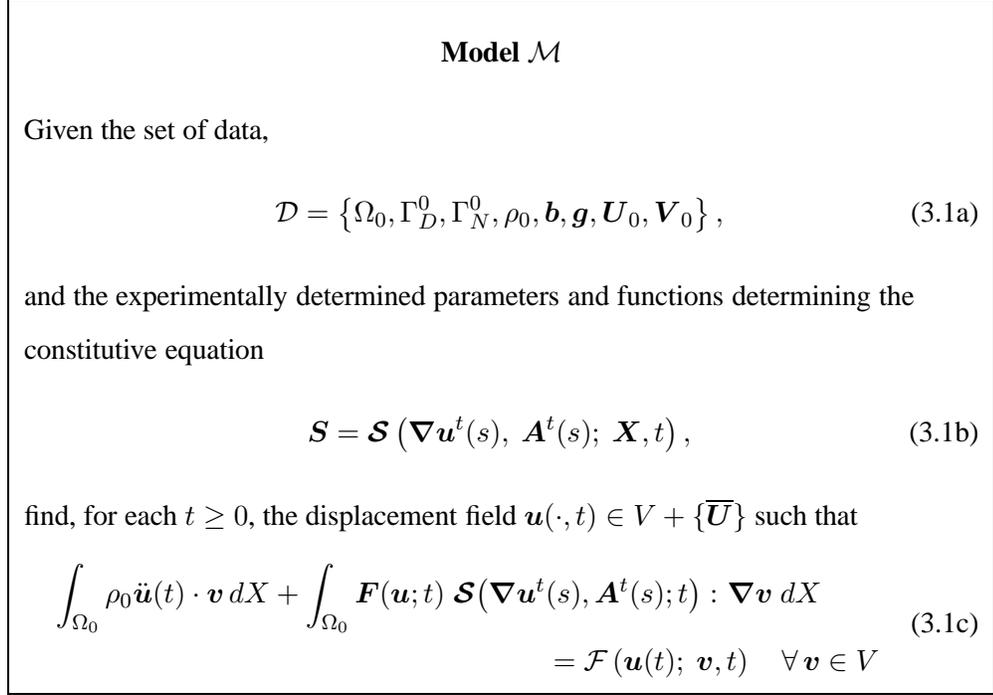
Figure 3.1: Model $\mathcal{M}$

ious quantities on $\boldsymbol{X}$ is suppressed for simplicity, $\ddot{\boldsymbol{u}} = \partial^2 \boldsymbol{u} / \partial t^2$, and $\boldsymbol{F}(\boldsymbol{u}; t)$ is the deformation gradient, $\boldsymbol{F} = \boldsymbol{I} + \boldsymbol{\nabla} \boldsymbol{u}$. The symbol V denotes the space of admissible displacements (e.g. $V = \{\boldsymbol{v} \in (W^{1,p}(\Omega_0))^n \colon \boldsymbol{v}|_{\Gamma_D} = \boldsymbol{0}\}$), and $\mathcal{F}(\cdot)$ is the loading functional, which depends upon $\boldsymbol{b}, \boldsymbol{g}$, and possibly $\boldsymbol{u}$; e.g.

$$\mathcal{F} \left(\boldsymbol{u}(t); \boldsymbol{v}, t \right) = \int_{\Omega_0} \rho_0 \boldsymbol{b}(t) \cdot \boldsymbol{v} \, dX + \int_{\Gamma_N^0} \boldsymbol{g}(\boldsymbol{u}, t) \cdot \boldsymbol{v} \, ds_0.$$

Remarks:

1. The set $\mathcal{D}$, as well as the coefficients, kernels, and various parameters appearing in the constitutive functional (3.1b), as well as the boundary and initial data $(\overline{\boldsymbol{U}}, \boldsymbol{U}_0, \boldsymbol{V}_0)$, can only be determined approximately, through, for instance, imaging data, testing, and through experimentally characterizing functionals the forms of which are themselves postulated. Thus, they can only be represented in some statistical sense, so that the solutions to the general fine-scale model are, in this sense, random variables. We ignore this fact here and consider only deterministic models.

2. Even when the data is known precisely, the fine model may be, in general, intractable; the governing system may be too complex to be solved by the largest and most sophisticated computers available. We do not, in general, expect to solve the fine-scale model. As noted earlier, it only serves as a datum with respect to which coarse, simplified models are measured.

Model $\mathcal{M}_0 \in \mathcal{M}$

Given the data set $\mathcal{D}$ and the constitutive equation

$$\boldsymbol{S}_0 = \boldsymbol{\mathcal{S}}_0 \left(\boldsymbol{\nabla u}_0^t(s), \boldsymbol{A}_0^t(s); \; \boldsymbol{X}, t \right) \qquad (3.2a)$$

find, for each $t \geq 0$, $\boldsymbol{u}_0(t) \in V_0 + \{\overline{\boldsymbol{U}}\}$ such that

$$\int_{\Omega_0} \rho_{00} \ddot{\boldsymbol{u}}_0(t) \cdot \boldsymbol{v} \, dX + \int_{\Omega_0} \boldsymbol{F}(\boldsymbol{u}_0; t) \, \boldsymbol{\mathcal{S}}_0 \left(\boldsymbol{\nabla u}_0^t(s), \boldsymbol{A}_0^t(s); t \right) : \boldsymbol{\nabla v} \, dX$$

$$= \mathcal{F}\left(\boldsymbol{u}_0(t); \boldsymbol{v}, t \right) \quad \forall \, \boldsymbol{v} \in V_0$$

$$(3.2b)$$

Figure 3.2: Model $\mathcal{M}_0$

3. We are not concerned here with verification of approximations of the models in class $\mathcal{M}$. Thus, we assume that the simplified or coarse model can be solved exactly. In practical applications of our modeling strategies, we use highly accurate finite element approximations of solutions of the simplified model. For instance, the accuracy can be determined and controlled by *a posteriori* estimates of *approximation* error and associated adaptive methods (see [1]). ∎

In actual applications, we use a coarser or simplified model $\mathcal{M}_0$ within the class $\mathcal{M}$. Let us suppose that the simplified model is characterized as shown in Fig. 3.2. Here it is understood that the coarse (simplified) solution $\boldsymbol{u}_0(\cdot, t) \in V_0 + \{\overline{\boldsymbol{U}}\}$, $V_0 \subset V$, and, importantly, $\boldsymbol{u}_0$ satisfies the same boundary and initial conditions as the fine solution $\boldsymbol{u}$. The dependence on $\boldsymbol{X}$ of the integrands is suppressed. In the simplified problem, the mass density ρ_{00} and the constitutive equation for stress may be different than that of the fine model problem. As remarked earlier, the coarse model is assumed to be solvable, and we assume that we can calculate an exact or extremely accurate approximation solution $\boldsymbol{u}_0$ at each $(\boldsymbol{X}, t) \in \overline{\Omega}_0 \times [0, T]$. In general, no additional data or testing are needed to define the model $\mathcal{M}_0$ as it is a special subclass of $\mathcal{M}$. However, some processing, such as averaging or homogenization, may be used to reduce $\mathcal{M}_0$ from the fine scale model. In fact, for many cases, more than one coarse model approximation to the fine scale model can be defined.

62

Modeling Error Bounds

Let us assume that the motion (particularly the displacement field) $\boldsymbol{u}_0 \in V$ of the coarse model is known for all $t \in [0, T]$. Then the modeling error in various terms at time t are defined by

$$\boldsymbol{e_u}(t) = \boldsymbol{u}(t) - \boldsymbol{u}_0(t) \tag{3.3}$$

$$\boldsymbol{E_S}(t) = \boldsymbol{S}\left(\boldsymbol{\nabla u}^t(s), \boldsymbol{A}^t(s); t\right) - \boldsymbol{S}\left(\boldsymbol{\nabla u}_0^t(s), \boldsymbol{A}^t(s); t\right) \tag{3.4}$$

$$e_{\mathcal{F}}(\boldsymbol{v}, t) = \mathcal{F}(\boldsymbol{u}(t); \boldsymbol{v}, t) - \mathcal{F}(\boldsymbol{u}_0(t); \boldsymbol{v}, t) \tag{3.5}$$

$$e_\rho = \rho_0 - \rho_{00} \tag{3.6}$$

The variables $\boldsymbol{A}^t(s)$ in (3.4) are computed using the fine model evolution equations with the displacement gradient history indicated in the same argument of $\boldsymbol{S}$. For each $t \geq 0$, the modeling error is defined by the linear functional,

$$\mathcal{E}(t) \in V':$$
$$\langle \mathcal{E}(t), \boldsymbol{v} \rangle : = \int_{\Omega_0} \left\{ \left(\rho_0 \ddot{\boldsymbol{e}}_{\boldsymbol{u}}(t) - e_\rho \ddot{\boldsymbol{u}}_0(t)\right) \cdot \boldsymbol{v} + \boldsymbol{\nabla e_u}(t) \boldsymbol{S}\left(\boldsymbol{\nabla u}_0^t(s), \boldsymbol{A}^t(s); t\right) : \boldsymbol{\nabla v} \right.$$
$$\left. + \boldsymbol{F}\left(\boldsymbol{u}_0 + \boldsymbol{e_u}; t\right) \boldsymbol{E}_S(t) : \boldsymbol{\nabla v} \right\} dX - e_{\mathcal{F}}(\boldsymbol{v}, t) \tag{3.7}$$

where $\langle \cdot, \cdot \rangle$ denotes duality pairing on $V' \times V$, V' being the dual of V.

The *global a posteriori modeling error indicator* is defined by the one-parameter scalar-valued function,

$$\zeta(t): = \left\{ \int_{\Omega_0} \left| \boldsymbol{F}\left(\boldsymbol{u}_0(t); t\right) \left(\boldsymbol{S}_0\left(\boldsymbol{\nabla u}_0^t(s), \boldsymbol{A}_0^t(s); t\right) - \boldsymbol{S}\left(\boldsymbol{\nabla u}_0^t(s), \boldsymbol{A}^t(s); t\right) \right) \right|^2 dX \right\}^{1/2}. \tag{3.8}$$

It is straightforward to show that the following global bound on modeling error holds (see [13, 19]):

$$\|\mathcal{E}(t)\|_{V'} \leq \zeta(t) \tag{3.9}$$

where $\| \cdot \|_{V'}$ is the norm in the dual V' of V. An interpretation of the error measure on the left-hand side of (3.8) as a global smoothing of total error is given in [13].

3.1 Generalization of error indicator

The global error indicator ζ can be used to define adaptive strategies which can be used to select various models over given spatial and temporal regions. Some of the many strategies that can be

defined will be detailed in the following subsection. Before preceding to detail any such strategies, the global error indicator ζ will be generalized to account for the time history of what particular models are used for a given subdomain. That is the governing equations are discretized in the time domain and the models used over each time step are selected adaptively.

We have two types of models which are of interest. One is the equilibrium model which is used to integrate the constitutive equations over the current time step in enforcing quasi-static or dynamic equilibrium at the end of the current time step. The other is a comparison model used in determining alternate stresses at the end of the current time step. Consider the case where the equilibrium model is a coarse model. Then, the fine model may be used as the comparison model in order to judge the accuracy of the equilibrium model. Another possibility is that the comparison model is an even coarser model which is a candidate to be the next equilibrium model. In fact, multiple comparisons between various models may be made at the end of a time step. The error indicator associated with comparing the equilibrium (E) and comparison (C) models is then written as follows:

$$
\begin{aligned}
\zeta_{E/C}(t) \;=\; & \left\{ \int_{\Omega_0} \left| \boldsymbol{F}(\boldsymbol{u}_E(t); t) \left[\boldsymbol{S}_E \left(\boldsymbol{\nabla} u_H^t(s), \boldsymbol{A}_{H,E}^t(s); t \right) \right. \right. \right. \\
& \left. \left. \left. - \boldsymbol{S}_C \left(\boldsymbol{\nabla} u_H^t(s), \boldsymbol{A}_{H,C}^t(s); t \right) \right] \right|^2 \, dX \right\}^{1/2}
\end{aligned}
\tag{3.10}
$$

where $\boldsymbol{\nabla} u_H^t(s)$ is the displacement gradient history from the various equilibrium models actually used and the state variables are computed using the equilibrium model displacements as follows:

$$
\boldsymbol{A}_{H,E}^t(s) \;=\; \begin{cases} \boldsymbol{A}_H(t-s) & t \geq s \geq \Delta t \quad (0 \leq t - s \leq t - \Delta t) \\ \boldsymbol{A}_E(t-s) & \Delta t \geq s \geq 0 \quad (t - \Delta t \leq t - s \leq t) \end{cases}
\tag{3.11}
$$

$$
\boldsymbol{A}_{H,C}^t(s) \;=\; \begin{cases} \boldsymbol{A}_H(t-s) & t \geq s \geq \Delta t \quad (0 \leq t - s \leq t - \Delta t) \\ \boldsymbol{A}_C(t-s) & \Delta t \geq s \geq 0 \quad (t - \Delta t \leq t - s \leq t) \end{cases}
\tag{3.12}
$$

Here $\boldsymbol{A}_H(t-s)$ denotes the state variable history from the actual equilibrium models used.

Note that the error indicator defined by (3.10), (3.11), and (3.12) is local in time, i.e., it characterizes only the differences between the equilibrium and comparison models over the current time step. That is, in integrating the constitutive equations, the state variables and stresses at the start of the time step are taken to be the same in both models. The accumulation of errors from previous time steps is not considered.

In order to make adaptivity decisions, $\zeta_{E/C}$ will be normalized by $\hat{\zeta}_E$ which is given as follows:

$$
\hat{\zeta}_E(t) = \left\{ \int_{\Omega_0} \left| \boldsymbol{F}(\boldsymbol{u}_E(t); t) \left[\boldsymbol{S}_E \left(\boldsymbol{\nabla} u_H^t(s), \boldsymbol{A}_{H,E}^t(s); t \right) \right] \right|^2 \, dX \right\}^{1/2}
\tag{3.13}
$$

Let us suppose that the domain Ω_0 is partitioned into a collection of non-overlapping subdomains Ω_{0K}:

$$\mathcal{P}\colon \overline{\Omega}_0 = \bigcup_{K=1}^{N(\mathcal{P})} \overline{\Omega}_{0K}; \qquad \Omega_{0K} \cap \Omega_{0J} = \varnothing,\ K \neq J. \tag{3.14}$$

The global model error indicator $\zeta_{E/C}(t)$ of (3.10) can be written as the sum,

$$\zeta_{E/C}(t) = \left\{ \sum_{K=1}^{N(\mathcal{P})} \left(\zeta_{E/C}^K(t) \right)^2 \right\}^{1/2} \tag{3.15}$$

where

$$
\begin{aligned}
\left(\zeta_{E/C}^K(t) \right)^2 &= \int_{\Omega_{0K}} \left| \boldsymbol{F}(\boldsymbol{u}_E(t); t) \left[\boldsymbol{S}_E \left(\boldsymbol{\nabla} u_H^t(s), \boldsymbol{A}_{H,E}^t(s); t \right) \right.\right.\\
&\qquad\qquad \left.\left. - \boldsymbol{S}_C \left(\boldsymbol{\nabla} u_H^t(s), \boldsymbol{A}_{H,C}^t(s); t \right) \right] \right|^2 \, dX
\end{aligned} \tag{3.16}
$$

Similar equations hold for $\hat{\zeta}_E(t)$ and $\hat{\zeta}_E^K(t)$. For finite element analysis, a natural choice is to have each element correspond to a subdomain.

The $\zeta_{E/C}^K(t)$ are the contributions of the error over Ω_{0K} to the global modeling error indicator $\zeta_{E/C}(t)$. They do not represent local modeling errors, as the actual local error is generally polluted by errors in remote subdomains. Nevertheless, the $\zeta_{E/C}^K(t)$ are used as an indication of the relative error in various subdomains/elements and to identify subdomains/elements where model refinement is required or where model coarsening is allowed.

3.2 Adaptivity schemes

In order to perform model adaptivity, three questions must be answered which are when?, where?, and how? A variety of methods that answer these questions can be developed to adaptively select from among the members of a defined model family. These schemes come from selecting among explicit and implicit implementations, global and local use of the developed error indicators, smart and dumb model selection, and equilibrium and fine model state variable updates. The schemes have been implemented in the Sandia code JAS3D[6] which can solve for quasi-static equilibrium states of solid structures using either conjugate gradient or dynamic relaxation iterative solvers. In all of the implemented schemes, adaptivity calculations are only performed when equilibrium has been achieved within a specified tolerance using the chosen equilibrium models. Also, each element has been chosen to be a separate subdomain for the adaptivity calculations in each adaptive algorithm.

Explicit/Implicit Implementations

When computing an adaptive solution, a choice must be made between bounding the values of the error indicator for a minimum computational cost and achieving minimum error indicator values for a given computational cost. While not doing so exactly, the chosen implicit and explicit algorithms attempt to offer solutions for each choice.

An implicit scheme allowing multiple passes through a given solution period is suited to bounding the error indicator values. However, such a method may require more than two passes and, thus, may expend sizable computational effort to do so. If the chosen solution period is composed of multiple time steps, model switching could be delayed until a critical number of elements required refinement or allow for coarsening. When such a point is reached, switching would be performed and the calculations would restart from the start of the solution period. However, implementing an implicit scheme in this manner would require either additional input/output or memory to be used to store all of the information associated with the solution period start point. For implementation purposes, only the case where each time step is considered as a separate solution period is chosen. In fact, three types of implicit methods have been implemented. In each case the implicit calculations over a single time step proceed until the error indicator(s) fall below a predefined limit.

The first method is a pure implicit implementation where refinement and coarsening are allowed for any element during any pass through the current time step. The difficulty with this approach is that model selection may cycle through set of choices over and over again without ever achieving the required limits on the error indicator(s). Another approach referred to as restricted implicit allows refinement whenever required, but coarsening only in elements which have not required refinement previously during one of the previous passes through the current time step. The final approach allows both coarsening and refinement at the end of the first pass through the current time step. Then for additional passes through the current time step only refinement is allowed. This final approach is termed "one time" implicit as coarsening of any element is allowed only one time during each time step.

Explicit schemes where a single pass is performed through a solution period will not limit the error indicator values, but, of course, will not compute multiple solutions for any solution period. The drawback, of course, is that it is not possible to try correct any large errors identified by large error indicator values. Rather, the error indicators are just used to select the model that will be used in each element for the next time step. In order to limit error growth, these explicit techniques

can be mated to selection processes where the fine model is chosen immediately when refinement is indicated as opposed to just refining to a slightly better coarse model. Such model selection schemes will be discussed later. Similar to the implicit implementations, the explicit codes have solution periods that each correspond to a separate time step.

Local/Global Use of Error Indicator

Although the error indicator $\zeta_{E/C}(t)$ that has been derived is global in nature, the element contributions $\zeta_{E/C}^K(t)$ can be used in determining where refinement is required or coarsening is allowed. The difference between the methods which are here characterized as either local or global in nature come about from exactly how the adaptivity questions of when? and where? are answered. From this point onward these types of methods will be termed as either local or global schemes.

In the local type of schemes, all choices about adaptivity are answered using only information from the subdomain/element under consideration. The local error indicator $\bar{\zeta}_{E/C}^K$ is defined as

$$\bar{\zeta}_{E/C}^K := \sqrt{\left(\zeta_{E/C}^K\right)^2 \Big/ \left(\hat{\zeta}_{E/C}^K\right)^2} \tag{3.17}$$

For each equilibrium configuration, refinement and coarsening are indicated as follows:

- Refinement required:

$$\bar{\zeta}_{E/F}^K \geq TOL \tag{3.18}$$

- Coarsening allowed:

$$\bar{\zeta}_{E/F}^K < TOL \quad \& \quad \bar{\zeta}_{E/C}^K < TOL \tag{3.19}$$

where F denotes the fine model, TOL is a predefined tolerance and the comparison model C in (3.19) is, of course, coarser than the equilibrium model.

Because a structure may consist of many different components and materials, JAS3D performs many calculations on a material block by block basis where a material/element block is simply a group of elements. In keeping with this architecture of JAS3D, the global schemes as implemented consider adaptivity for each material block separately (sums and maximum are taken over all the elements in a given material block). For a problem involving only a single material block, the global schemes are indeed truly global. For the global methods, the following model selection criteria are used:

- Adaptivity (refinement or coarsening) allowed:

$$\frac{\zeta_{E/F}(t)}{\hat{\zeta}_E(t)} \geq \gamma \qquad (3.20)$$

- Refinement required:

$$\frac{\zeta_{E/F}^K(t)}{\max_K \zeta_{E/F}^K(t)} \geq \alpha \qquad (3.21)$$

- Coarsening allowed:

$$\frac{\zeta_{E/F}^K(t)}{\max_K \zeta_{E/F}^K(t)} < \beta_1 \qquad (3.22)$$

and

$$\frac{\zeta_{E/C}^K(t)}{\max_K \zeta_{E/C}^K(t)} < \beta_2 \qquad (3.23)$$

Here γ, α, β_1, and β_2 are predefined tolerances. In (3.21) and (3.22), all elements in a block are considered in determining the maximum $\zeta_{E/F}^K(t)$. Elements that require refinement are not considered in determining the maximum $\zeta_{E/C}^K(t)$ in (3.23). Note that in determining the maximum $\zeta_{E/F}^K(t)$ and $\zeta_{E/C}^K(t)$, the equilibrium and coarse comparison models are likely to differ from element to element. The decision of whether to allow adaptivity or not as given by (3.20) means that adaptivity is likely to be performed only for select time steps.

Smart/Dumb Model Selection

After the needed error indicators have been computed and signify either refinement or coarsening based on the predefined tolerances, the question of which model will be selected to become the new equilibrium model must be answered. In the present research, two alternatives have been investigated to select from among the members in a nested family of models.

The smart model selection procedure allows models to be selected non-sequentially. When refinement is required, the fine model is chosen to be the next equilibrium model. On the other hand, when coarsening is allowed, the coarsest model with an acceptable error indicator is selected when error measure is being used locally. For the codes using the error measure globally, the coarse model selected for computation of (3.23) is the one giving the lowest $\zeta_{E/C}^K(t)$. Obviously, the refinement procedure requires that the fine model be computationally tractable. This model

selection scheme has been designed especially for explicit implementations to minimize the error growth when refinement is indicated and to select the cheapest model possible when coarsening is allowed.

The dumb model selection approach selects models sequentially. That is, the new equilibrium model chosen when adaptivity is indicated is either a single level above or below the current equilibrium model. This approach is particularly well-suited to implicit schemes which allow multiple passes through a time step to select the cheapest models possible for a given maximum error indicator.

Equilibrium/Fine Model State Variable Updates

In all of the adaptivity algorithms implemented, the first step after an equilibrium state has been converged is to assess whether the equilibrium models used in each element have generated acceptable results. This, of course, involves comparing the actual equilibrium and approximate fine model stresses. The question is whether any of the computed fine model information can be used to enhance the equilibrium model results. Note that the stresses from the comparison fine model are only approximate and do not correspond to an equilibrium state, unless of course, the equilibrium model is in fact the fine model. Furthermore, the stresses from the equilibrium model for a particular element depend on what models are used in other elements. Nevertheless, it may be acceptable to save the approximate fine model state variable updates along with the equilibrium model displacements. Whether this is satisfactory or not depends on the details of the defined model family. For models in a family which differ only in how the state variable updates are calculated, saving the fine model state variable updates is akin to using the fine model with a coarser tolerance for halting the force equilibrium calculations. This may not be too dangerous, as equilibrium will be forced again in the next set of equilibrium calculations.

Implemented Algorithms

Fourteen schemes corresponding to various combinations of the different options detailed above have been implemented for the nonlinear viscoelastic model family to be described later. Acronyms for each algorithm have been generated as follows. The first letter is either E, I, R, or O corresponding to explicit, pure implicit, restricted implicit, or one time implicit, respectively. The second letter which is either L or G, respectively, denotes whether the error indicator is used locally or globally. The third letter can be either S or D for the cases of smart or dumb model selection, respectively.

69

Finally, E or F, respectively, is used in the final slot to refer to whether the equilibrium or fine model state variable updates are saved. For example, the explicit code using the error measure locally with dumb model selection and equilibrium model state variable updates is denoted as ELDE. The schemes that have been implemented in JAS3D are ELSE, ELSF, ELDE, ELDF, EGSE, EGSF, EGDE, EGDF, ILSE, RLSE, OLSE, OLDE, OGSE, and OGDE.

3.3 Application to nonlinear viscoelastic polymers

An application of the present theory to a family of material laws for polymers is now considered. First, a brief overview of the nonlinear viscoelastic (NLVE) family will be given and followed by details of each model in the NLVE family. Adaptive results will be given for an example involving tension of a plate with a hole.

Nonlinear Viscoelasticity (NLVE) Overview

The constitutive law developed by Chambers, Adolf and Caruthers [7] describes the nonlinear viscoelastic response of glassy polymers. This material model was developed using a thermodynamically consistent rational mechanics approach. The material constants in the model were fit to an epoxy so that a wide range of physical phenomena can be modeled both comprehensively and quantitatively. A key feature of the detailed nonlinear viscoelastic model used in this example is that the physical phenomena of interest, such as yielding, volume recovery, or enthalpy relaxation, are captured through the use of a material clock assuming rheological simplicity. The material/reduced time scale on which viscoelastic relaxations proceed is controlled by the horizontal shift factor a as given by Eq. (3.29) which is to follow.

The various models in the NLVE family are delineated based on what physics are incorporated into $\log_{10}[a]$. This leads to models which differ only in terms of how the state variables are updated over each time step. The fine scale model includes the effects of thermal, volumetric strain and stress histories in the horizontal shift factor which is given in terms of the configurational energy. The first level of approximation has a shift factor which depends upon configurational entropy and includes the effects of the temperature and volume histories on $\log_{10}[a]$. A still coarser model is where $\log_{10}[a]$ depends only on the current temperature through the well-known Williams-Landel-Ferry (WLF) equation. This model assumes thermodynamic equilibrium and has been found to be valid for polymers under moderate stresses at temperatures ranging from the glass transition temperature θ_g to $\theta_g + 100$ K. For polymers under moderate stresses at temperatures below θ_g, the configurational

Table 3.1: Models in the NLVE family ordered in terms of computational expense.

Model	Material Relaxations
(1B) Rubbery elastic	all relaxations completed by end of time step
(1D) Pseudo elastic	no relaxations over time step
(2B) WLF	$a = a(\theta)$
(3B) Configurational entropy	$a = a(\theta, \text{volumetric strain})$
(4B) Configurational energy	$a = a(\theta, \text{volumetric strain, stress})$

entropy model must be used. Finally, two elastic levels are defined which correspond to no material relaxations occurring over the current time step (pseudo elastic model) or all of the relaxations reaching completion by end of the current time step (rubbery elastic model). A summary of the NLVE models is presented in Table 3.1 where they have been numbered and ordered according to increasing computational cost.

Nonlinear Viscoelastic (NLVE) Family of Material Models

In rational mechanics, expressions for the stress, entropy and internal energy are determined from the expression for the specific Helmholtz free energy ψ (J/g) which is given in a generalized Lamé constant form for the nonlinear viscoelastic family presently being examined as follows:

$$
\begin{aligned}
\psi(t) \;=\; & \psi^\infty(t) + \Delta G_0^1 \sum_{i=1}^{N} g_i^1 (\boldsymbol{I}_{\mathcal{C}}^i(t) : \boldsymbol{I})^2 + \Delta G_0^2 \sum_{i=1}^{N} g_i^2 (\boldsymbol{I}_{\mathcal{C}}^i(t) : \boldsymbol{I}_{\mathcal{C}}^i(t)) \\
& + \Delta A_0 \sum_{i=1}^{N} a_i (\boldsymbol{I}_{\mathcal{C}}^i(t) : \boldsymbol{I}) I_\theta^i(t) + \Delta C \sum_{i=1}^{N} c_i (I_\theta^i(t))^2
\end{aligned}
\tag{3.24}
$$

where $\boldsymbol{I}_{\mathcal{C}}^i$ and I_θ^i are state variables that represent integral histories of strain and temperature, respectively. Furthermore in Eq. (3.24), $\Delta(\cdot)$ denotes $(\cdot)^g - (\cdot)^\infty$ (i.e., the difference between the glassy and rubbery value of a relaxation modulus) and these coefficients have been taken to be constant. Also, all four relaxation spectra have been expressed as Prony series with identical distributions of relaxation times. The quantity $\boldsymbol{I}_{\mathcal{C}}^i(t)$ is given by

$$
\boldsymbol{I}_{\mathcal{C}}^i(t) = \int_{-\infty}^{t} \exp\left[-\frac{(t^* - \xi^*)}{\tau_i} \right] \frac{d\mathcal{C}}{d\xi} \, d\xi
\tag{3.25}
$$

71

where $t^* - \xi^*$ represents a difference in the reduced or material time scale, τ_i is the ith relaxation time and $\mathcal{C}$ is the Hencky strain measure given by

$$\mathcal{C} = \boldsymbol{I} + \ln \boldsymbol{C} \tag{3.26}$$

and $\boldsymbol{C}$ is the right Cauchy-Green deformation tensor given by

$$\boldsymbol{C} = \boldsymbol{F}^T \boldsymbol{F} \tag{3.27}$$

Similarly, $I_\theta^i(t)$ is given by

$$I_\theta^i(t) = \int_{-\infty}^{t} \exp\left[-\frac{(t^* - \xi^*)}{\tau_i}\right] \frac{d\theta}{d\xi} \, d\xi \tag{3.28}$$

The reduced and physical time scales are related as follows:

$$t^* - \xi^* = \int_{\xi}^{t} \frac{1}{a(t')} \, dt' \tag{3.29}$$

where a is the shift factor which differs in each member of the NLVE family.

The stress, entropy and the rate of entropy production are determined from

$$\frac{d\psi}{dt} = \frac{1}{\rho} \boldsymbol{S}_H : \frac{d\mathcal{C}}{dt} - \eta \frac{d\theta}{dt} + \dot{\sigma} \tag{3.30}$$

where η is the specific entropy, $\boldsymbol{S}_H$ is the stress that is work conjugate with the Hencky strain measure and $\dot{\sigma}$ is the rate of entropy generation. The Hencky stress is found to be

$$\frac{1}{2\rho_g} \boldsymbol{S}_H = \frac{1}{2\rho_g} \boldsymbol{S}_H^\infty + 2\Delta G_0^1 \sum_{i=1}^{N} g_i^1 (\boldsymbol{I}_\mathcal{C}^i : \boldsymbol{I}) \boldsymbol{I} + 2\Delta G_0^2 \sum_{i=1}^{N} g_i^2 \, \boldsymbol{I}_\mathcal{C}^i + \Delta A_0 \sum_{i=1}^{N} a_i I_\theta^i \boldsymbol{I} \tag{3.31}$$

where ρ_g is the reference density in a stress free state at the glass transition temperature θ_g. The equilibrium contribution to the Hencky stress is as follows:

$$\boldsymbol{S}_H^\infty = \left[4\rho_g \left(G_0^1\right)^\infty (\mathcal{C} : \boldsymbol{I} - 3) + 2\rho_g A_0^\infty (\theta - \theta_g)\right] \boldsymbol{I} + 4\rho_g \left(G_0^2\right)^\infty (\mathcal{C} - \boldsymbol{I}) \tag{3.32}$$

The second Piola Kirchhoff stress tensor is given in terms of the Hencky stress tensor by the following relationship:

$$\boldsymbol{S} = \boldsymbol{S}_H : \frac{d\mathcal{C}}{d\boldsymbol{C}} \tag{3.33}$$

The $d\mathcal{C}/d\boldsymbol{C}$ tensor is computed numerically.

Configurational Energy Model

The most sophisticated and, hence, accurate member of the NLVE family of material models uses a shift factor based on a quantity termed the configurational internal energy. Briefly, the configurational internal energy E_c is the internal energy of the actual viscoelastic material minus that coming from its glassy response to the same volumetric and thermal history. Recall that the specific internal energy E is as follows:

$$E = \psi + \theta\eta \tag{3.34}$$

The current value of the configurational energy E_c depends on the current values of $\boldsymbol{I}_{\mathcal{C}}^i$ and I_θ^i. Then, the shift factor in terms of the configurational energy is

$$\log_{10}[a] = B\left(\frac{1}{E_c} - \frac{1}{\Delta_{ref}}\right) \tag{3.35}$$

where B is a constant in the present work. Using a shift factor which depends upon the configurational internal energy gives a material clock which depends upon the thermal, volumetric strain and stress histories through $\boldsymbol{I}_{\mathcal{C}}^i$ and I_θ^i.

The quasi-static equilibrium states are determined at discrete times in the finite element solution with numerical integration used for the constitutive equation. The history integrals in this model and all other models except for the two elastic models to be defined later are marched in time using a modified central difference scheme [31, 7] as follows:

$$\boldsymbol{I}_{\mathcal{C}}^i(t^n) = \left(\frac{2a_{avg}\tau_i - \Delta t_{min}}{2a_{avg}\tau_i + \Delta t_{min}}\right)\boldsymbol{I}_{\mathcal{C}}^i(t^{n-1}) + \left(\frac{2a_{avg}\tau_i\Delta t_{min}}{2a_{avg}\tau_i + \Delta t_{min}}\right)\left(\frac{\boldsymbol{C}(t^n) - \boldsymbol{C}(t^{n-1})}{\Delta t^n}\right) \tag{3.36}$$

and

$$I_\theta^i(t^n) = \left(\frac{2a_{avg}\tau_i - \Delta t_{min}}{2a_{avg}\tau_i + \Delta t_{min}}\right)I_\theta^i(t^{n-1}) + \left(\frac{2a_{avg}\tau_i\Delta t_{min}}{2a_{avg}\tau_i + \Delta t_{min}}\right)\left(\frac{\theta(t^n) - \theta(t^{n-1})}{\Delta t^n}\right) \tag{3.37}$$

where

$$\Delta t_{min} = \min\left\{\Delta t^n, 2a_{avg}\tau_i\right\} \tag{3.38}$$

For the configurational energy model (4B), a_{avg} is given by

$$\log_{10}[a_{avg}] = B\left(\frac{2}{E_c(t^n) + E_c(t^{n-1})} - \frac{1}{\Delta_{ref}}\right) \tag{3.39}$$

A simple fixed point iteration scheme is used to converge the nonlinear constitutive calculations given $\boldsymbol{C}(t^{n-1})$, $\boldsymbol{C}(t^n)$, $J(t^n)$, $\theta(t^{n-1})$, $\theta(t^n)$, $\boldsymbol{I}_{\mathcal{C}}^i(t^{n-1})$, $I_\theta^i(t^{n-1})$ and $E_c(t^{n-1})$ where $J = \det \boldsymbol{F}$.

Configurational Entropy Model

Similar to the configurational internal energy, the configurational internal entropy is defined as that coming from the actual specimen minus the internal entropy from the specimen's glassy response under the same volumetric strain and thermal histories. Noting Eq. (3.34), it should be apparent that the configurational internal energy contains terms corresponding to the configurational entropy. In order to be consistent with the configurational energy model, the shift factor in the configurational entropy model (3B) is actually based on $\theta_g \eta_c$ where η_c is the actual configurational entropy. For this coarse model, $\log_{10}[a]$ is written as

$$\log_{10}[a] = B \left(\frac{1}{\theta_g \eta_c} - \frac{1}{\Delta_{ref}} \right) \tag{3.40}$$

For the numerical integration, a_{avg} is determined from

$$\log_{10}[a_{avg}] = B \left(\frac{2}{\theta_g \eta_c(t^n) + \theta_g \eta_c(t^{n-1})} - \frac{1}{\Delta_{ref}} \right) \tag{3.41}$$

Similar to the configurational energy model (4B), a fixed point integration scheme is used to converge the nonlinear constitutive calculations that result from η_c being a function of $\boldsymbol{I}_C^i$ and I_θ^i. The configurational entropy model (3B) includes the effects of the volumetric strain and thermal histories on $\log_{10}[a]$ and is a good approximation to the configurational energy model (4B) for the case where the stresses are at or below moderate levels.

WLF Model

For the case of a polymer subjected to low to moderate stresses when the polymer is at temperatures ranging from θ_g to approximately $\theta_g + 100\,\text{K}$, wide experimental evidence indicates that $\log_{10}[a]$ can be expressed strictly in terms of the current temperature by the well-known Williams-Landel-Ferry [32] relationship which is given as follows:

$$\log_{10}[a] = \frac{-C_1(\theta - \theta_g)}{C_2 + (\theta - \theta_g)} \tag{3.42}$$

where C_1 and C_2 are material constants. Because the horizontal shift factor in this case depends only on the temperature and not on the stress or strain, it is properly termed a thermorheologically simple linear viscoelastic model. Of course, the material response is nonlinear in terms of the thermal history. It should also be noted that the proper kinematics for large deformations/strains are still used in this and all other models. For the numerical integration of the constitutive law, a_{avg} is determined using Eq. (3.42) with θ replaced by $\theta_{avg} = [\theta(t^n) + \theta(t^{n-1})]/2$. Because temperature is assumed to be specified, no iterations are required for the material law calculations in this model.

Elastic Models

The material response of a cross-linked polymer at very short or very long elapsed times since a load was applied can be characterized elastically using the appropriate constants that describe the polymer's glassy and rubbery moduli. That is, it is not necessary to consider any material relaxations in order to find the initial and final viscoelastic response of the material. For such a glassy response, essentially no material relaxations have had a chance to occur since the load was applied, whereas for the corresponding rubbery response, all material relaxations have been completed since the load was applied. Furthermore, these glassy or rubbery moduli could be used to determine the initial or final viscoelastic structural responses for all path-independent problems, respectively. The idea of not computing any material relaxations will be generalized into two elastic models which can be used in combination with the previously presented viscoelastic models in computing the nonlinear, possibly path-dependent, structural responses for structures composed of materials falling into the NLVE family. The pseudo elastic model (1D) that will be defined corresponds to the case where all material relaxations are proceeding extremely slowly over each time step in which the model is used, whereas the rubbery elastic model (1B) will give the result for the case where all relaxations have reached completion by the end of the time step in which it is used. Because neither model includes the explicit use of a shift factor a, no iterations are required to converge the numerical computation of either constitutive model.

Pseudo Elastic Model

For the pseudo elastic response over the current time step, consider the following exact equation for $I_\theta^i(t^n)$:

$$I_\theta^i(t^n) = \exp\left[-\frac{(\Delta t^*)^n}{\tau_i}\right] I_\theta^i(t^{n-1}) + \int_{t^{n-1}}^{t^n} \exp\left[-\frac{(t^*)^n - \xi^*}{\tau_i}\right] \frac{d\theta}{d\xi} \, d\xi \qquad (3.43)$$

For this model, the assumption is that t^* changes very little over the current time step so that the following approximation for $I_\theta^i(t^n)$ is acceptable:

$$I_\theta^i(t^n) \approx I_\theta^i(t^{n-1}) + \theta(t^n) - \theta(t^{n-1}) \qquad (3.44)$$

Likewise, $\boldsymbol{I}_C^i(t^n)$ is updated using

$$\boldsymbol{I}_C^i(t^n) = \boldsymbol{I}_C^i(t^{n-1}) + \boldsymbol{C}(t^n) - \boldsymbol{C}(t^{n-1}) \qquad (3.45)$$

It should be apparent from Eq. (3.29) that small Δt^* over a time step results from Δt being very small and/or $\log_{10}[a]$ tending to be positive and relatively large over the time step. Note that if other models in the NLVE family have been used before the pseudo elastic approximation (1D), the pseudo elastic (1D) results will not correspond simply to using $\boldsymbol{S}_H = \boldsymbol{S}_H^\infty$ with the rubbery moduli replaced by the corresponding glassy values in Eq. (3.32). However, if only this model has been used for all times over a given region, then the Hencky stress for that region would be $\boldsymbol{S}_H = \boldsymbol{S}_H^\infty$ with the rubbery moduli replaced by the corresponding glassy values. That is, using the pseudo elastic model for all time steps gives purely a glassy elastic response. Regardless of what other models may have been used in previous time steps, the response of this model to any load increments applied in the current time step is the traditional glassy elastic response of the polymer. Hence, this model can be thought of as an enhanced glassy elastic model that can also be used in time steps where Δt^* is small.

Rubbery Elastic Model

For the rubbery elastic model (1B), all material relaxations which started before and during the current time step are assumed to reach completion by the end of the current time step. For this to be true, the elapsed reduced time since any loading was applied needs to be relatively large. Exactly how large this is can be determined from the largest relaxation time appearing in the Prony series that are used to characterize the viscoelastic material. A large amount of elapsed reduced time since a load was applied may or may not correspond to a large amount of elapsed physical time. Recall that the reduced and physical time scales are related as given by Eq. (3.29). For instance, if $\log_{10}[a]$ is constant at -2, the rate at which the reduced time proceeds would be two orders of magnitude larger than the rate at which physical time elapses. Because all relaxations are assumed to be completed by the end of the time step, all integral histories $\boldsymbol{I}_C^i$ and I_θ^i are set to zero, while the Hencky stress at the end of the time step is simply $\boldsymbol{S}_H^\infty$ as given by Eq. (3.32).

Hierarchy of NLVE Material Models

In terms of least to largest amount of computational cost, the models are ordered as rubbery elastic (1B), pseudo elastic (1D), WLF (2B), configurational entropy (3B), and configurational energy (4B). In terms of the physics captured in $\log_{10}[a]$, the non-elastic models are ordered as WLF (2B), configurational entropy (3B), and configurational energy (4B). Discounting any possible pathological examples, if the WLF model (2B) is accurate, then so is the configurational entropy model (3B).

76

Although they do not compute what the actual viscoelastic relaxations are over a given time step, either of the two elastic models may be accurate over certain time steps for a given loading scenario. As the most detailed model in the NLVE family, the configurational energy model (4B) is assumed to capture all of the relevant physics. Hence, the assessment of accuracy of any of the coarse NLVE models must be made with respect to the configurational energy model (4B).

3.4 NLVE numerical results

The present example considers a square plate with a centrally-located hole under tension. Symmetry boundary conditions are used on two edges of the domain so that only one-fourth of the plate has to be modeled. The plate has a length of 50.8 mm, a thickness of 0.635 mm and hole with a radius of 6.35 mm. The plate geometry is shown in Fig. 3.3. The plate is assumed to be stress free and in thermodynamic equilibrium in the initial state. The temperature field is uniform and held constant at θ_g throughout the entire analysis. The tension boundary conditions are applied as a prescribed displacement on the exterior edge that is perpendicular to the global X axis. The prescribed displacement in the X direction increases linearly with time at a rate of 0.254 mm/sec.

A large set of numerical results have been generated corresponding to all of the implemented schemes. For each method except ILSE, at least two levels of model adaptivity tolerances are used. For the codes involving local use of the error measure, TOL is set to either 0.05 or 0.01. For all of the codes which use the error measure globally, the adaptivity parameters are either the combination ($\gamma = 0.1$, $\alpha = 0.8$, $\beta_1 = 0.3$, $\beta_2 = 0.2$) or ($\gamma = 0.01$, $\alpha = 0.7$, $\beta_1 = 0.2$, $\beta_2 = 0.1$). Additional results corresponding to other combinations of (γ, α, β_1, β_2) have been generated for some codes. Also, a baseline solution corresponding to using the configurational energy model (4B) for all elements over all time steps has been computed. Unless otherwise noted, all results correspond to using a relatively fine mesh of 7803 elements. This mesh and element groupings for post processing are shown in Fig. 3.3. A single element block is used for the schemes using the error measure in a global fashion. For some methods, additional results have been generated using a coarser mesh of 867 elements. In all cases, 100 equal time steps have been used to integrate the results over the first 10.0 sec. The fine mesh and time domain discretizations have been determined using a convergence study for the baseline solution. The assumption is made that the approximation errors resulting from these discretizations can be ignored at present. The interaction between modeling and approximation errors can be formally explored in a later work.

For brevity, only a few representative numerical results from the baseline and ELDE (explicit,
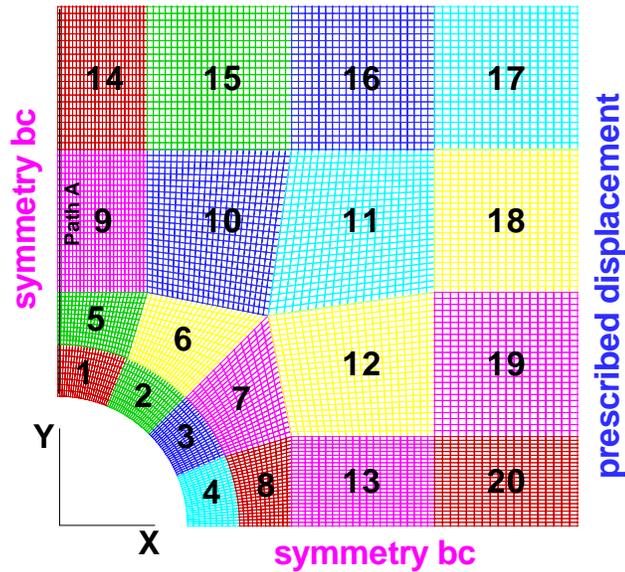
Figure 3.3: Quarter plate geometry and mesh of 7803 elements. In addition, the elements have been grouped into twenty numbered regions for reporting average quantities over a region. Also shown is path A which connects the centroids of the elements on the vertical symmetry edge, beginning at the top of the hole and proceeding toward the plate edge.

local use of error measure, dumb model switching, equilibrium model state variable updates) solutions will be given. Some results from using ELSE with the coarse mesh of 867 elements are given in Ref. [13]. However, some tabular data and general observations considering all of the generated results will be presented later on.

Reference Solution

Contours of the Cauchy stress component σ_{xx} for the reference solution are shown in Fig. 3.4. For these and all other contour plots presented for this example, the results are presented on deflected meshes. Contours of σ_{yy} are not shown for brevity. For relatively small times, the nonlinear viscoelastic results give stress concentrations that resemble what would occur for a similar elastic case. That is, the location of the highest tensile σ_{xx} occurs at the intersection of the hole and symmetry plane perpendicular to the X-axis, whereas the highest compressive stress σ_{yy} occurs at the intersection of the hole and the symmetry plane perpendicular to the Y-axis. However, as the time and

loading increase, the σ_{xx} stress concentration moves slightly along the X-symmetry plane away from the hole. These changes in the distribution of σ_{xx} result from the fact that the stresses affect the rate at which viscoelastic relaxations occur in the configurational energy model (4B). The compressive σ_{yy} stresses are not large enough to cause yielding at the intersection of the Y-symmetry plane and the hole. Hence, the location of the highest compressive σ_{yy} remains stationary.

ELDE Solutions with $\mathbf{TOL} = 0.05$ and $\mathbf{0.01}$

Results generated using the ELDE implementation with $TOL = 0.05$ and 0.01 will now be presented. In these and all other adaptive cases considered, all elements begin their constitutive calculations using the WLF model (2B). Shown in Fig. 3.5 are contour plots of σ_{xx} for the case where $TOL = 0.05$. It is clear that the stress σ_{xx} in this adaptive solution follows the same trends as in the reference baseline solution. However, it is also apparent that the deformations near the top of the hole are somewhat different. The time history of the element-wise distribution of models used are presented in Fig. 3.6 for the $TOL = 0.05$ case. The band where the configurational energy model (4B) is used corresponds to that where σ_{xx} is relatively high. Outside of this band, the WLF model (2B) and the pseudo elastic model (1D) are predominantly used.

For $t = 5.0$ and 10.0 sec, contours of σ_{xx} and element-wise distribution of model usage are shown in Fig. 3.7 for the ELDE scheme with $TOL = 0.01$. Examining Fig. 3.7, it is clear that using a finer tolerance leads to better stress results, but at the expense of greater usage of the configurational energy model (4B).

Figure 3.8 shows a plot the average σ_{xx} over region 1 as function of time, whereas Fig. 3.9 gives the spatial variation of σ_{xx} along path A at discrete results. Once again, the tighter tolerance is observed to give much better results, especially in terms of the smoothness of the stress variation along path A.
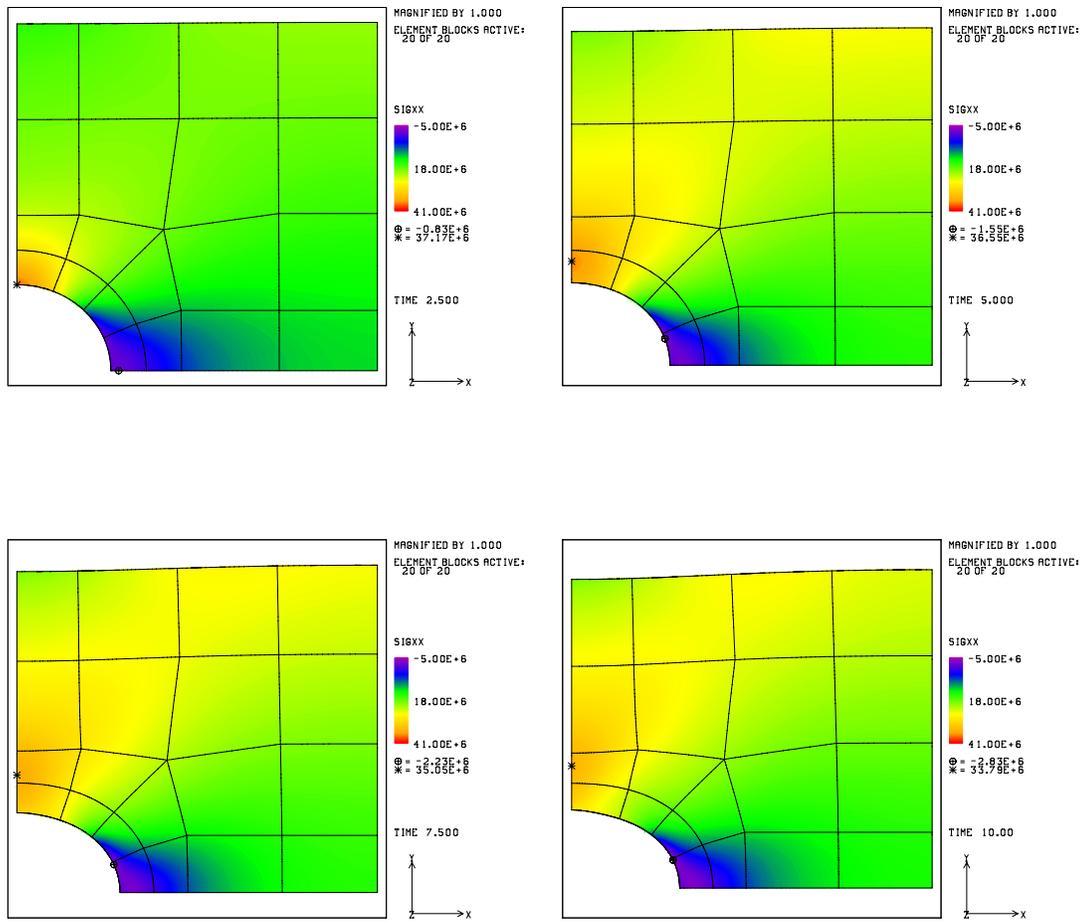
Figure 3.4: Reference solution: Cauchy stress σ_{xx} at $t = 2.5$, 5.0, 7.5 and 10 seconds (left to right, top to bottom) for the case where the configurational energy model (4B) is used for all calculations.
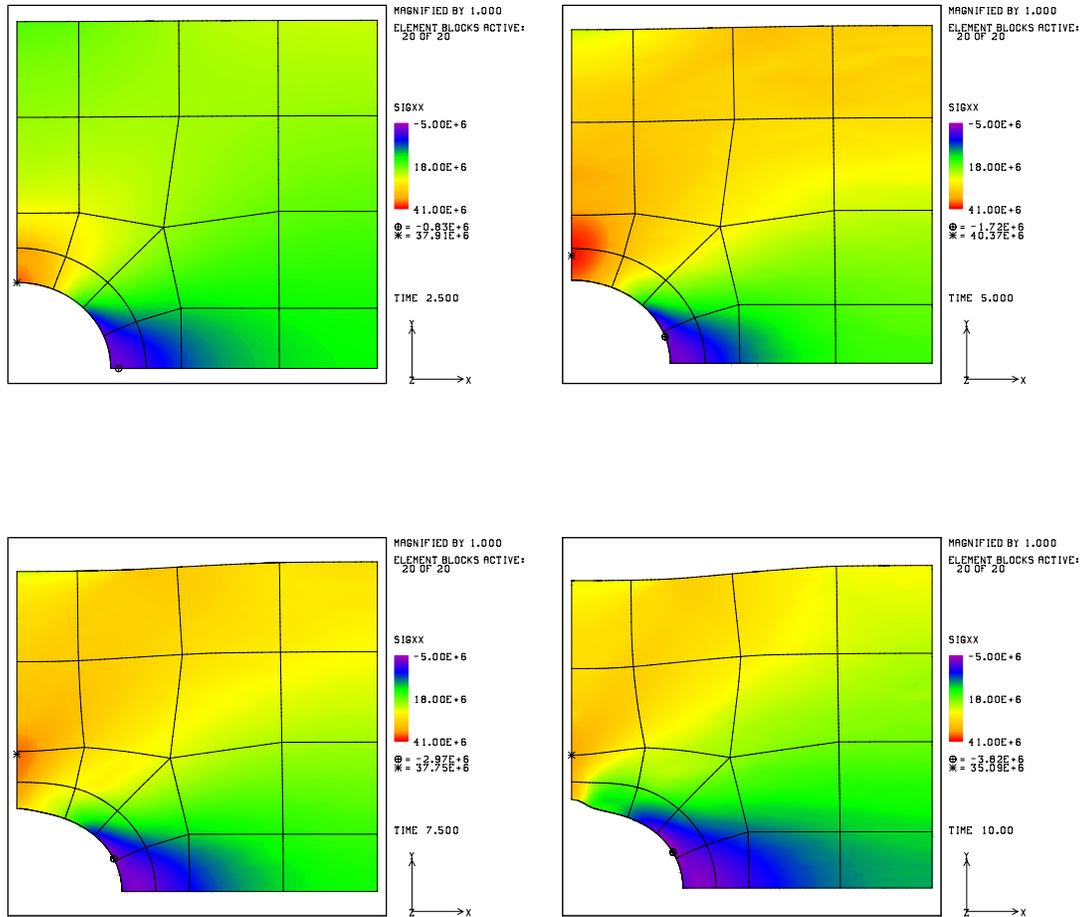
Figure 3.5: Cauchy stress σ_{xx} at $t = 2.5$, 5.0, 7.5 and 10 seconds (left to right, top to bottom) for the ELDE scheme with $TOL = 0.05$.
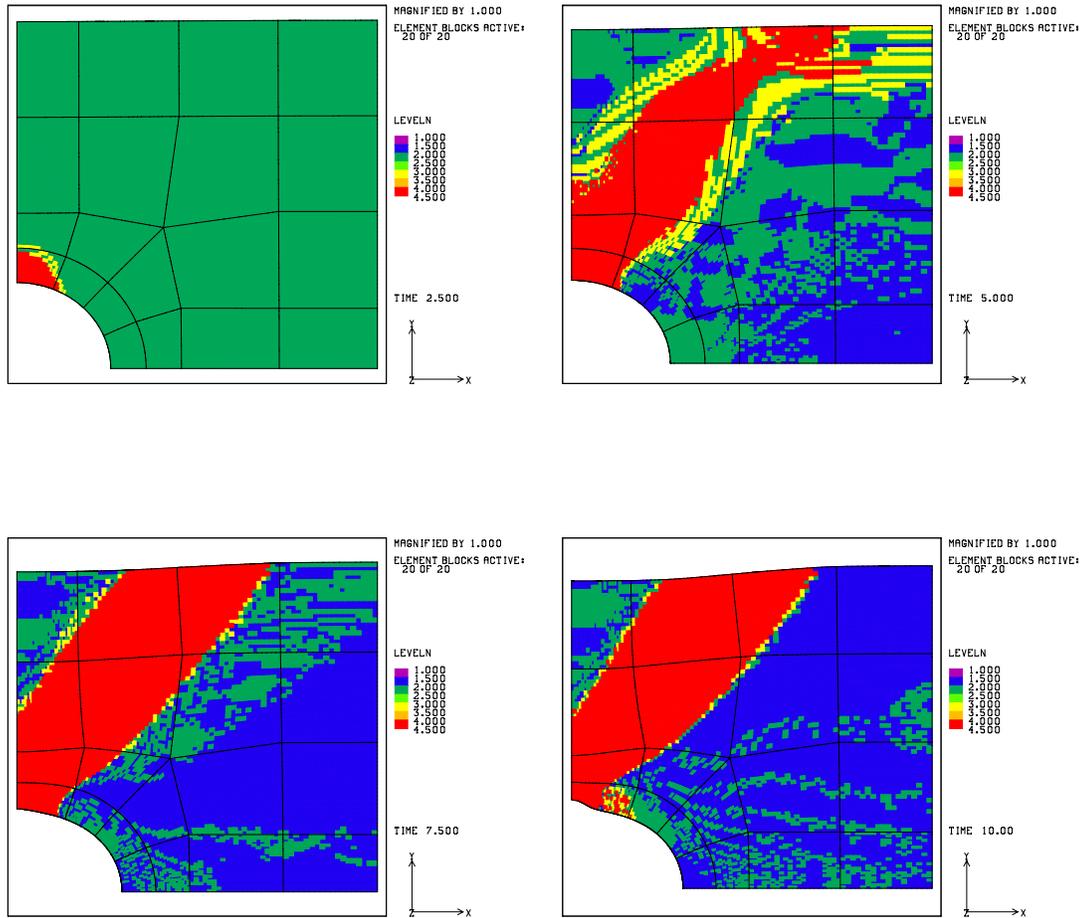
Figure 3.6: Models used for the time steps ending at $t = 2.5$, 5.0, 7.5 and 10 seconds (left to right, top to bottom) for ELDE with $TOL = 0.05$. (configurational energy (4B) = red, configurational entropy (3B) = yellow, WLF (2B) = green, pseudo elastic (1D) = blue, rubbery elastic (1B) = purple).
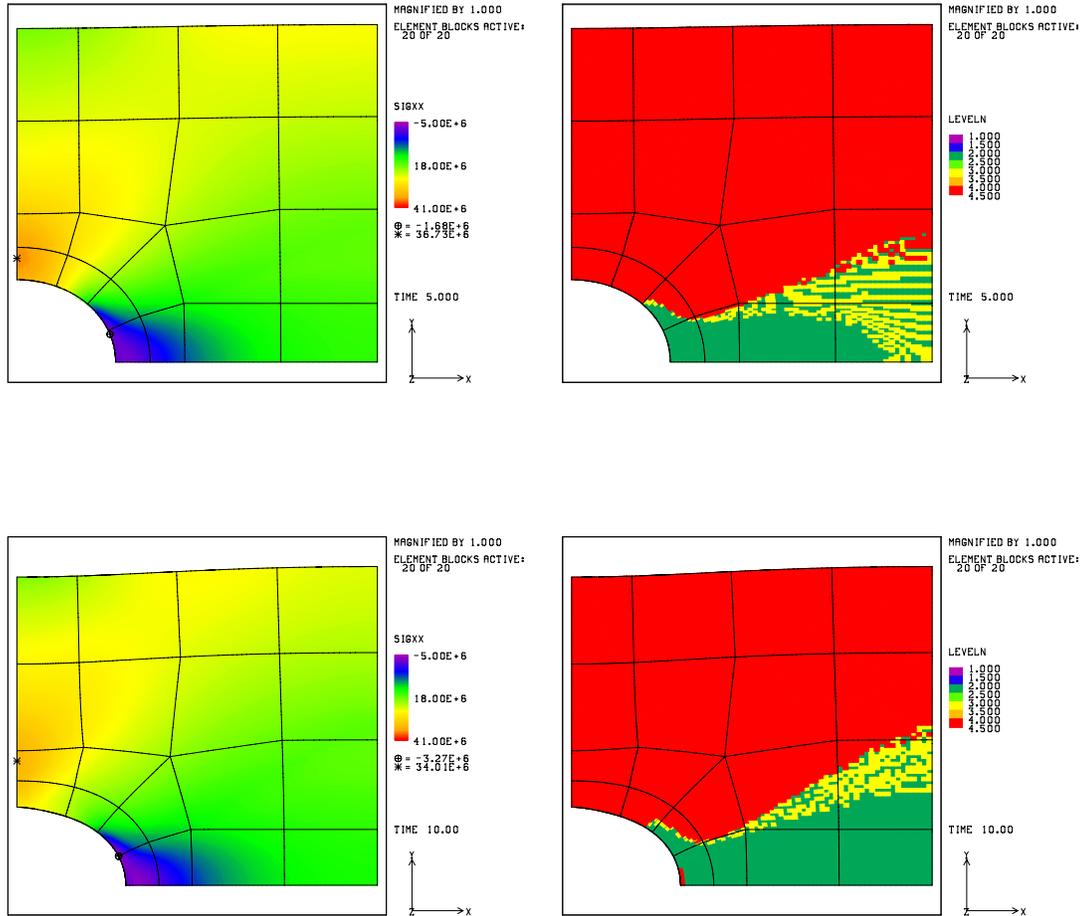
Figure 3.7: Cauchy stress σ_{xx} (left) and the models used over the time step (right) at $t = 5.0$ and 10.0 sec (configurational energy (4B) = red, configurational entropy (3B) = yellow, WLF (2B) = green, pseudo elastic (1D) = blue, rubbery elastic (1B) = purple) for the ELDE scheme with $TOL = 0.01$.
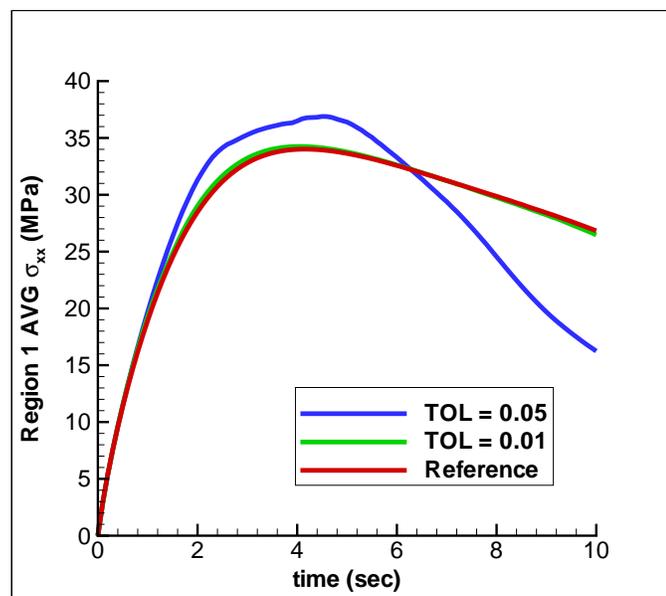
Figure 3.8: Average Cauchy stress σ_{xx} over region 1. Adaptive results corresponding to the ELDE algorithm along with the reference solution are shown.
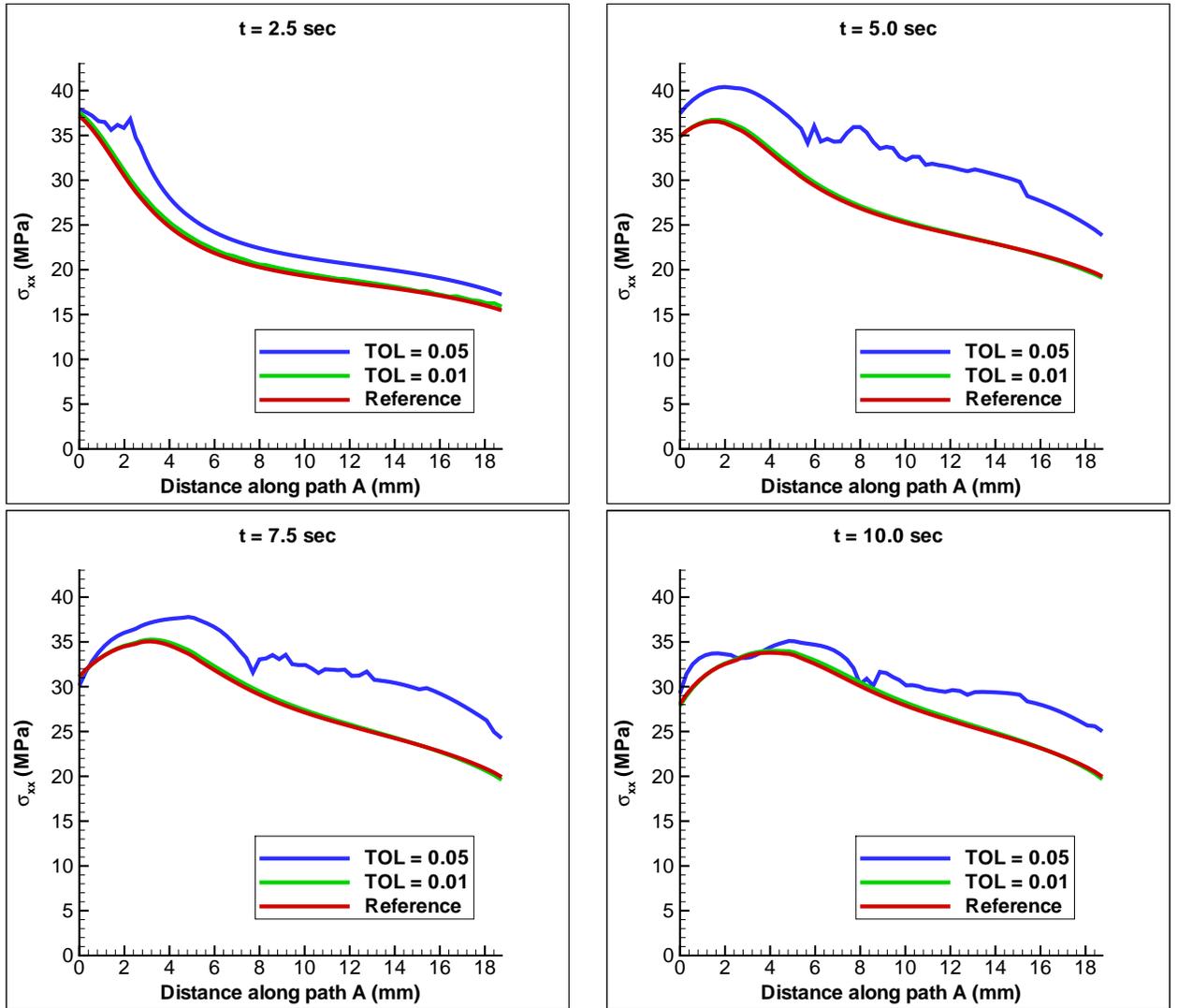
Figure 3.9: Cauchy stress σ_{xx} along path A at discrete times. The ELDE adaptive solutions using two tolerance levels along with the reference solution computed using the configurational energy model (4B) for all times are shown.

General Results and Observations

Although a number of results have been postprocessed for all of the adaptive analyses performed, only a small fraction of the generated information is given in this report. However, general observations based on these results will be presented subsequently. Tables 3.2 and 3.3 present data for each of the 35 adaptive runs performed using the fine mesh of 7803 elements. This information consists of the required effort (equilibrium iterations and CPU time) and a measure of the accuracy of the calculations (error in the maximum σ_{xx} over the domain at the end of the analysis). Hence, this information can be used as a general indication of the effectiveness of each adaptive simulation. Note that some runs failed to converge past a certain time step for the tolerances specified for the equilibrium calculations, whereas other analyses were stopped after it became clear that the results were highly erroneous and/or a large computational effort had already been spent. These tables do not present any results from the runs corresponding to using the coarse mesh of 867 elements.

All analyses indicate that tightening the tolerances used to answer the adaptivity questions of when?, where?, and how? resulted in higher accuracy in the results, as expected. Also in a number of cases, tightening these tolerances also lead to smaller computational effort being required. This likely results from the fact that using different spatial and temporal distribution of models leads to different equilibrium states being found. Of course, these differing equilibrium states require different amounts of computational effort to find. In fact, for a lot of the adaptive schemes using loose tolerances may lead to failure of the numerical solution procedure as the computed equilibrium states are vastly different from those in the reference solution. Although not presented in this report, the number of implicit iterations used by the RLSE, OLSE, OLDE, OGSE, and OGDE methods generally was smaller for the tighter tolerances for this same reason. For instance, using OGSE or OGDE with $\gamma = 0.1$ leads to no model adaptivity being performed until $t = 5.4$ sec, whereas using $\gamma = 0.01$ with these methods leads to adaptivity starting at $t = 1.0$ sec. Waiting so long before making any model changes, of course, leads to equilibrium states somewhat different than those in the reference solution. It may be prudent in future work to try other tolerance levels for each adaptive scheme for this problem.

Another closely related point is that any comparison model results are estimated using the displacement solution generated using the given equilibrium models over the domain. That is, no attempt is made to determine how the displacements would change if the comparison model is actually used. Of course, in the implicit calculations, equilibrium is converged again after any model

86

Table 3.2: Computational effort and $(\sigma_{xx})_{max}$ results for adaptive simulations with local use of error measure. All results correspond to using the fine mesh of 7803 elements.

Code	TOL	Equilibrium Iterations	CPU (hr)	% Error in Final $(\sigma_{xx})_{max}$
Ref	-	56,164	15.36	-
ELSE	0.05	82,995	16.28	5.47
ELSE	0.01	52,007	9.64	0.57
ELSF	0.05	71,892	10.04	0.96
ELSF	0.01	49,753	8.88	0.06
ELDE	0.05	77,573	11.32	3.86
ELDE	0.01	50,127	9.25	0.64
ELDF	0.05	62,718	7.59	1.49
ELDF	0.01	48,506	8.61	0.05
RLSE	0.05	$> 10^6$	281.02	4.84
RLSE	0.01	134,287	30.22	-1.84
OLSE	0.05	804,706°	unknown	8.43°°
OLSE	0.01	84,811	17.80	-1.75
OLDE	0.05	725,279[+]	119.33[+]	6.20[++]
OLDE	0.01	80,011	16.74	-1.73

° Total iterations when job was killed during $t = 9.1$ sec calculations
°° Calculated at $t = 9.0$ sec
[+] Iterations/CPU time to reach numerical failure for $t = 9.7$ sec calculations
[++] Calculated at $t = 9.6$ sec

Table 3.3: Computational effort and $(\sigma_{xx})_{max}$ results for adaptive simulations with global use of error measure. All results correspond to using the fine mesh of 7803 elements.

Code	γ	α	β_1	β_2	Equilibrium Iterations	CPU (hr)	% Error in Final $(\sigma_{xx})_{max}$
Ref	-	-	-	-	56,164	15.36	-
EGSE	0.1	0.8	0.8	0.2	115,601°	unknown	861.93$^{\circ\circ}$
EGSE	0.1	0.8	0.3	0.2	98,888$^+$	10.86$^+$	942.03^{++}
EGSE	0.01	0.7	0.7	0.1	197,871$^\triangle$	26.3$^\triangle$	59.96$^{\triangle\triangle}$
EGSE	0.01	0.7	0.7	0.0	51,792	8.81	1.10
EGSE	0.01	0.7	0.2	0.1	71,426	13.27	6.23
EGSF	0.1	0.8	0.8	0.2	42,995	4.83	8.19
EGSF	0.01	0.7	0.7	0.1	49,570	7.81	0.09
EGSF	0.01	0.7	0.2	0.1	49,522	7.81	0.09
EGDE	0.1	0.8	0.8	0.2	171,507	19.00	702.79
EGDE	0.1	0.8	0.3	0.2	157,787	17.48	1147.71
EGDE	0.01	0.7	0.7	0.1	175,182$^\nabla$	20.5$^\nabla$	562.83$^{\nabla\nabla}$
EGDE	0.01	0.7	0.7	0.0	51,361	8.93	0.98
EGDE	0.01	0.7	0.2	0.1	70,826	13.12	6.92
EGDF	0.1	0.8	0.8	0.2	42,995	4.79	8.19
EGDF	0.01	0.7	0.7	0.1	58,271	7.07	47.70
EGDF	0.01	0.7	0.7	0.0	49,805	8.10	0.01
EGDF	0.01	0.7	0.2	0.1	51,715	8.525	0.06
OGSE	0.1	0.8	0.3	0.2	306,410$^\diamond$	50.78$^\diamond$	47.06$^{\diamond\diamond}$
OGSE	0.01	0.7	0.2	0.1	94,215	19.31	-2.30
OGDE	0.1	0.8	0.3	0.2	$> 10^{6*}$	234.5*	38.46**
OGDE	0.01	0.7	0.2	0.1	82,700	16.04	-0.63

$^\circ$ Total iterations to reach numerical failure for $t = 7.7$ sec calculations
$^{\circ\circ}$Calculated at $t = 7.6$ sec
$^+$ Total iterations/CPU time to reach numerical failure for $t = 7.9$ sec
$^{++}$ Calculated at $t = 7.8$ sec
$^\triangle$ Total iterations/CPU time when job was killed during $t = 8.3$ sec calculations
$^{\triangle\triangle}$ Calculated at $t = 8.2$ sec
$^\nabla$Total iterations/CPU time when job was killed during $t = 6.6$ sec calculations
$^{\nabla\nabla}$Calculated at $t = 6.5$ sec
$^\diamond$ Total iterations/CPU time to reach numerical failure for $t = 7.0$ sec
$^{\diamond\diamond}$ Calculated at $t = 6.9$ sec
* Total iterations/CPU time when job was killed during $t = 8.3$ sec calculations
** Calculated at $t = 8.2$ sec

changes, whereas no such effort is made using the explicit methods. However, this approximation of the comparison model can still be a problem for the implicit methods when decisions are being made about whether to consider adaptivity or not using (3.20).

A potential pitfall for the algorithms that use the error measure globally is the normalizations used in (3.21), (3.22), and (3.23) to determine where refinement and coarsening are performed. The fact that the maximum $\zeta_{E/F}^K$ over the entire domain (or material block) and the maximum $\zeta_{E/C}^K$ from the elements not requiring refinement over this same region are used can lead to difficulties when these values are much larger than their respective means. That is, large error indicator values in just a few elements can make the error indicators in almost all other elements appear to be relatively small. For instance, consider the elements near the stress concentration in the present example. These elements usually use the fine model (configurational energy model (4B)), and, hence, do not need refinement. The coarse comparison model results from these elements are likely to be highly erroneous compared to the local fine model results, but could potentially be used for the maximum $\zeta_{E/C}^K$ in (3.23). This particular problem was observed for some of the numerical results from the various global schemes.

Although not evident from any of the presented results, a lot of chattering from time to time step occurred in the model selection process of each method. Differing amounts of chattering are expected from the different adaptive algorithms. For instance, the explicit methods are expected to result in more chattering in the final model selection from time step to time step than the implicit methods. Of course, the implicit methods also have chattering during the implicit calculations over a single time step. The smart model selection schemes are also expected to lead to more chattering than the dumb model selection schemes. However, no formal measure of chattering has been developed or implemented.

Chattering from time step to time step is not expected based on any of the physics being simulated in this numerical example. Rather, the chattering is a phenomenon of the implemented adaptivity schemes. Physically what is expected is distinct spatial and temporal regions where particular models are used. For example, one would expect that anywhere the stress is relatively high, the configurational energy model (4B) would be used. Hence, examining the stress contours from the reference solution, it is almost possible to guess where the configurational energy model (4B) should be used.

Of course, the chattering over the implicit calculations over a single time step differs in each type of implicit scheme. In fact, the differing rules for when coarsening is allowed were developed

partly to limit this chattering. Using the coarse mesh of 867 elements, it became clear for the pure implicit ILSE scheme that the model selection process can get stuck in a repeating loop. The only way to break such a loop is to reach the maximum number of implicit iterations allowed for a given time step.

Examining Tables 3.2 and 3.3, it is clear that much greater computational effort is expended using the implicit schemes rather than the explicit schemes. This results from the fact that each time new equilibrium models are selected using the implicit methods, the equilibrium state at the end of the current time step is reconverged to the same tolerance used in determining equilibrium in the explicit methods. One would naturally expect the number of iterations required for equilibrium to decrease as more and more implicit iterations are taken. This behavior was in fact observed. However, the number of equilibrium iterations after the first couple of implicit iterations usually was as large or larger as that in the initial equilibrium state determination for the time step. Also, the reduction in effort for equilibrium convergence in later implicit iterations was not as great as expected. Using some results generated using the coarse mesh of 867 elements, it was clear that even changing the models in only a few elements could raise the relative force imbalance by two orders of magnitude. Perhaps the best way to address this issue is to allow a coarser tolerance on the equilibrium calculations corresponding to the first couple of implicit iterations during a given time step.

The model selection process in the implicit methods encompasses not only chattering, but propagation as well. That is, in solving the boundary value problem, the model selection process in any given element is affected but what is occurring in other elements, especially those in the near vicinity. Hence, the use of a certain model may spatially propagate during the implicit iterations of a given time step. The number of implicit iterations is thus expected to rise as the mesh is refined, due to both more opportunities for chattering and more elements required to propagate the selection of a particular model a given spatial distance. For the implicit methods where calculations were also performed using the coarse mesh of 867 elements, fewer implicit iterations did result than when the fine mesh of 7803 elements was used.

The choice between using either an implicit or explicit scheme is likely to be problem dependent. That is, explicit methods are efficient and can be used in the case where the modeling error growth occurring in any given time step is relatively small and does not need to be corrected before proceeding to the next time step. For situations where this is not the case, an implicit method may in fact be a better alternative rather than just using smaller time steps with an explicit method.

Although no effort was made to examine approximation errors, these errors are, of course, present and do, in fact, affect the model selection process. That is, there are interactions between the modeling and approximation errors. This is evident when coarse mesh and fine mesh results for model selection using a particular method are examined. For this problem, the model selection process resulted in similar spatial and temporal distributions of constitutive models used in all of the cases examined. However, regions did exist where the models selected differed somewhat, especially during the later stages in the simulations.

Using the fine model state variable updates in the explicit methods appears to lead to good accuracy at a minimum cost. As noted earlier, for NLVE this corresponds to using the fine model stresses which are approximated using the displacements computed using the various equilibrium models. Although for explicit methods, these stresses do not correspond to equilibrium for the current time step, equilibrium is immediately enforced again in the next time step. Saving the fine model state variables may not work as well for other material model classes.

It should be noted that the data input into each model is assumed to be known precisely. That is, no attempt has been made to assess how any uncertainty in the data would affect model selection in the NLVE family. Any such data uncertainty may play a role in model selection as each model uses the material properties differently in determining elapsed time on the material time scale.

The first set of numerical results appear to be very encouraging. Indeed, adaptive solutions using different models in various regions of the computational domain have been produced that are almost as accurate as the solution obtained with the finest model. Nevertheless more numerical experiments are needed in order to better assess the performance of the various implemented algorithms for modeling adaptivity. Furthermore, thermodynamic and possibly other considerations must be taken into account so that a more robust adaptivity scheme can be developed for the nonlinear viscoelastic model studied here and other material models.

4. Concluding Comments

The concepts of modeling error, error estimation, and adaptive modeling provide a framework for systematically selecting appropriate models of physical phenomena. Several additional advances are needed if these methods are to have a useful role in model validation. First, the predictive qualities of computational models will always depend upon the goals of the simulation. In other words, the particular physical event(s) of interest must be clearly specified before it is meaningful to compare

the effectiveness of various models of it. Second, the modeling error is, as noted earlier, a random variable. Thus, the adaptive modeling process should be embedded in an appropriate stochastic framework or in something equivalent. Thirdly, modeling error and approximation error must be simultaneously estimated and controlled for a completed, verified and validated predictive tool to be created. Finally, adaptive modeling should be integrated into a larger framework that provides an interaction and feedback with physical experiments and tests to allow dynamic updating of the parameters that define models within a hierarchy of possible models.

REFERENCES

[1] M. Ainsworth and J.T. Oden. *A Posteriori Error Estimation in Finite Element Analysis*. Wiley-Interscience, New York, (2000).

[2] Bajaj, C. L., Pascucci, V., and Schikore, D. "The Contour Spectrum". In *Proceedings of the 1997 IEEE Visualization Conference*, pages 167–173. IEEE, 1997.

[3] Bajaj, C. L. and Pascucci, V. "Progressive IsoContouring". Technical Report TR 1999-36, Texas Institute for Computational and Applied Mathematics, Austin, Texas, 1999.

[4] Bensoussan, A., Lions, J. L., and Papanicolaou, G. *Asymptotic Analysis for Periodic Structures*, volume 5 of *Studies in Mathematics and its Applications*. North-Holland, Amsterdam, 1978.

[5] G.E.P. Box. Robustness in the Strategy of Scientific Model Building. In R.L. Launer and G.N. Wilkinson, editors, *Robustness in Statistics*, pages 201–236, Academic Press, New York, (1979).

[6] Blanford, M.L. JAS3D: A Multi-Strategy Iterative Code for Solid Mechanics Analysis. Sandia National Laboratories, (1998).

[7] R.S. Chambers, D.B. Adolf, and J.M. Caruthers. A Comprehensively Validated Theory of the Nonlinear Viscoelasticity of Glassy Polymers. *Sandia Report* (in preparation).

[8] Christensen, R. *An Introduction to the Mechanics of Composite Materials*. Wiley, New York, 1979.

[9] R.G. Easterling. Measuring the Predictive Capability of Computational Models: Principles and Methods, Issues and Illustrations. *Sandia Report*, SAND2001-0243, Albuquerque, February, (2001).

[10] Edwards, H. C. *A Parallel Infrastructure for Scalable Adaptive Finite Element Methods and its Application to Least Squares C-Infinity Collocation*. PhD thesis, The University of Texas at Austin, 1997.

[11] Ghosh, S. and Moorthy, S. "Particle Fracture Simulation in Non-Uniform Microstructures of Metal-Matrix Composites". *Acta mater.*, 46 (1998) 965–982.

[12] Gropp, W., Lusk, E., Doss, N., and Skjellum, A. "A high-performance, portable implementation of the MPI message passing interface standard". *Parallel Computing*, 22 (1996) 789–828.

[13] D.C. Hammerand, J. Tinsley Oden, S. Prudhomme, and K. Mieczyslaw. Modeling Error and Adaptivity in Nonlinear Continuum Mechanics. *Sandia Report*, SAND2001-0132, Albuquerque, February, (2001).

[14] Hashin, Z. "Analysis of Composite Materials, A Survey". *J. Appl. Mech.*, 50 (1983) 481–505.

[15] Jikov, V. V., Kozlov, S. M., and Oleinik, O. A. *Homogenization of Differential Operators and Integral Functionals*. Springer-Verlag, Heidelberg, 1994.

[16] Kak, A. C. and Slaney, M. *Principles of Computerized Tomographic Imaging*. IEEE Press, New York, 1988.

[17] Gropp, W. D. and Lusk, E. *User's Guide for* `mpich`*, a Portable Implementation of MPI*. Mathematics and Computer Science Division, Argonne National Laboratory, 1996. ANL-96/6.

[18] J. T. Oden and K.S. Vemaganti. Estimation of Local Modeling Error and Goal-Oriented Modeling of Heterogeneous Materials, Part I. *Journal of Computational Physics*, vol. 164, pp 22–47, (2000).

[19] J.T. Oden, S. Prudhomme, D.C. Hammerand, and M. Kuczma. Modeling Error and Adaptivity in Nonlinear Continuum Mechanics. *Computer Methods in Applied Mechanics and Engineering*, to appear.

[20] Oden, J. T. and Zohdi, T. I. "Analysis and Adaptive Modeling of Highly Heterogeneous Elastic Structures". *Comp. Meth. Applied Mech. Engng.*, 148 (1997) 367–391.

[21] Oden, J. T., Vemaganti, K., and Moës, N. "Hierarchical Modeling of Heterogeneous Solids". *Comp. Meth. Applied Mech. Engng.*, 172 (1999) 3–25.

[22] Oden, J. T. and Vemaganti, K. "Adaptive Modeling of Composite Structures : Modeling Error Estimation". *Int. J. Comp. Civil Str. Engrg.*, 1 (2000) 1–16.

[23] Sanchez-Palencia, E. *Non-homogeneous Media and Vibration Theory*. Number 127 in Lecture Notes in Physics. Springer-Verlag, Berlin, 1980.

[24] Patra, A. K. and Oden, J. T. "Problem Decomposition Strategies for Adaptive hp Finite Element Methods". *Computing Systems in Engrg.*, 6 (1995) 97–109.

[25] Rachowicz, W., Oden, J. T., and Demkowicz, L. F. "Towards a Universal *h–p* Finite Element Strategy. Part 3. Design of *h–p* Meshes". *Comp. Meth. Applied Mech. Engrg.*, 77 (1989) 181–212.

[26] P. Roache. *Verification and Validation in Computational Science and Engineering*. Hermosa Press, Albuquerque, (1998).

[27] Ashcraft, C., Pierce, D., Wah, D., Wu, J. *The Reference Manual for SPOOLES, Release 2.2: An Object Oriented Software Library for Solving Sparse Linear Systems of Equations*. Boeing Shared Services Group, 1999.
Available at www.netlib.org/linalg/spooles/spooles.2.2.html.

[28] Vemaganti, K. "Modeling Error Estimation and Goal-Oriented Adaptive Modeling of Perforated Materials". Technical Report TR 00-25, Texas Institute for Computational and Applied Mathematics, Austin, TX, 2000.

[29] K.S. Vemaganti and J.T. Oden. Estimation of Local Modeling Error and Goal-Oriented Modeling of Heterogeneous Materials, Part II. A Computational Environment for Adaptive Modeling of Heterogeneous Elastic Solids. *Computer Methods in Applied Mechanics and Engineering*, to appear.

[30] Zohdi, T. I., Oden, J. T., and Rodin, G. J. "Hierarchical Modeling of Heterogeneous Bodies". *Comp. Meth. Applied Mech. Engng.*, 138 (1996) 273–298.

[31] Chambers, R.S. The Numerical Integration and 3-D Finite Element Formulation of a Viscoelastic Model of Glass. *Sandia Report*, SAND94-1607, Albuquerque, (1994).

[32] Aklonis, J.J. and MacKnight, W.J. *Introduction to Polymer Viscoelasticity*. Wiley, New York, (1983).

<div align="center">

CHAPTER 3

The Treatment of Uncertainties in

Computational Mechanics

Ivo Babuška

</div>

1. Introduction

Mathematical models in computational mechanics are usually defined by a boundary value problem involving linear or nonlinear partial differential equations and a specific goal for the analysis. The boundary value problem consists of some input data such as the partial differential equations and their coefficients, the domain of definition of the solution, the boundary conditions, etc.

For a given physical phenomenon, we may introduce a hierarchical sequence of models with increasing complexity. The lower models are usually derived from the "highest" model by simplification of the equations or by choosing specific coefficients. Assuming that the "highest" model yields an accurate representation of reality, the solution of the lower models can be viewed as approximations of the solution of the highest model.

The computational analysis creates a mapping from the space of input data into the space of the desired data specified by the goal of the analysis. The goal typically aims at defining the data of interest within a given tolerance range. Note that the input data are never known perfectly. Moreover, using more sophisticated models requires additional data which in turn present more uncertainties. The treatment of uncertainties therefore consists of finding the relationship between the uncertainties in the output (defined by the goal of the analysis) and the uncertainties in the input data.

Characterization of uncertainties in the input data can be two-fold:

1. The range of uncertainties is known; for example the range of the coefficients.

2. The statistics of uncertainties is known; however the statistics, namely the probabilistic characterization, may involve additional information which is not perfectly known.

<div align="center">

96

</div>

In the first case, we are interested in the range of the output, while in the second, we wish to obtain the probabilistic characterization of the output or some information derived from it, like, for instance, the mean value or the standard deviation.

The mathematical problems should possess reasonable properties, namely, existence and uniqueness of solutions and continuous dependence of these solutions on the data, in a sense appropriate with the type of uncertainties. In the case where the continuous dependence property is not satisfied, the mathematical model should be carefully investigated as to whether it accomplishes the objective of the analysis. Because of increasing uncertainties in the description of more complex models (the "higher" models in the hierarchy), these models, beyond a certain level, are not necessarily more reliable. In other words, they do not necessarily provide more accurate results with respect to reality. Needless to say that the determination of such a threshold level is not an easy task.

We present in the following some aspects of the research directions which address the issues described above. In particular, we will discuss the problem of uncertainty in the domain of definition of the solution and the problem of stochastic input data.

2. Uncertainty in the Domain

Let Ω denote the domain on which the differential equation is defined. Uncertainties in the definition of Ω necessarily appear when it is generated using some scanning devices. The domain is then recovered from digital images and therefore strongly depends on the resolution level. In three dimensions for instance, the domain is reconstructed from two-dimensional slices which are usually spaced at a distance larger than the resolution level of the two-dimensional image.

The digital images obtained by scanning are postprocessed based on a threshold value of the "shadow" intensity. In the ideal two-dimensional setting, the value of the threshold parameter characterizes the volume fraction of the domain contained in the pixels. Therefore, we can recover a "pixel" domain which approximates the real domain Ω by selecting a particular value of the parameter. The pixel domain is defined as the union of all pixels whose volume fraction is at least larger than the prescribed value. We then can generate a monotonic (in the ideal case) sequences of pixel domains by selecting sequences of threshold values. In doing so, we conceptually obtain a pixel domain, denoted by Ω_{LOW}, which is a subset of Ω. The domain Ω_{LOW} is associated for instance with the highest value of the parameter. Likewise we can obtain a pixel domain, denoted by Ω_{UP}, which contains Ω. This time, Ω_{UP} is associated with the lowest parameter in the sequence.
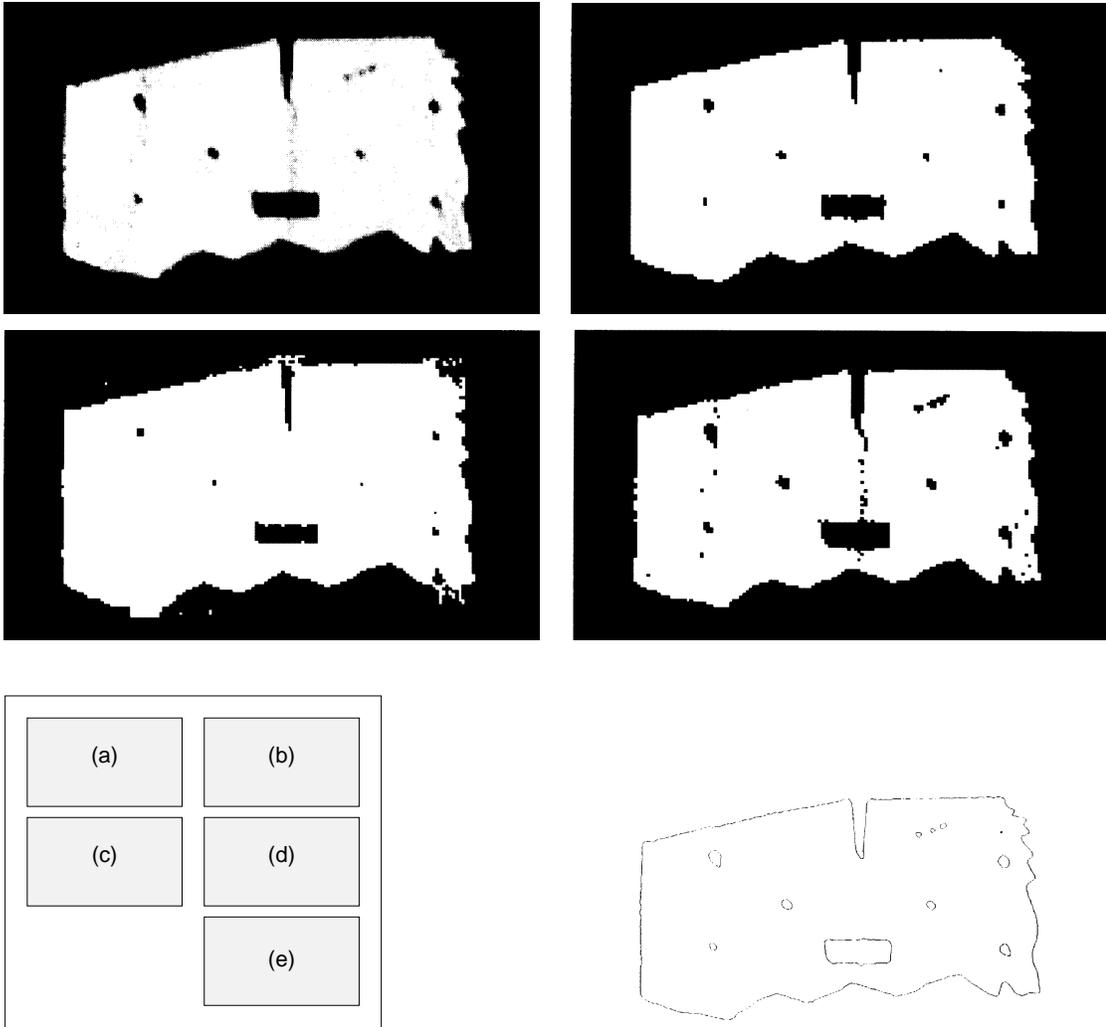
Figure 2.1: Original domain (a), pixel domains (b,c,d), and smoothed domain (e).

Unfortunately, this "ideal" situation is not always available because of various sources of "noise" in the digital images. There exist to date several, more or less heuristic, approaches in selecting an "optimal" value of the parameter; the corresponding pixel domain may be further postprocessed by smoothing the boundary while preserving known features of the domain. Various methods have been developed by and are currently being used in the computer graphics community.

To illustrate the problem, we show in Fig. 2.1 the digital photograph of a domain (a), three pixel domains for different values of the grading parameters (b,c,d) and the corresponding smoothed domain (e). In Fig. 2.2, we show the pixel approximation of fibers in a fiber composite for a given value of the parameter (diameter of the fiber is about 7μm). In this case, the input data are strongly dependent on the length separating fibers (especially when it is relatively small).
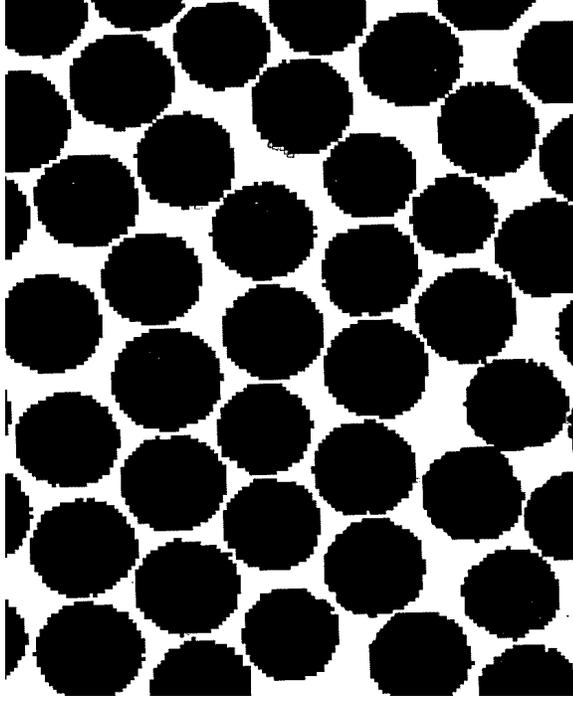
Figure 2.2: Fiber composite.

We proceed with our discussion by considering the following model problem:

$$-\Delta u + u = f, \qquad \text{in } \Omega \subset \mathbb{R}^2, \tag{2.1a}$$

$$\alpha \frac{\partial u}{\partial n} + \beta u = g, \qquad \text{on } \partial\Omega. \tag{2.1b}$$

We now distinguish three types of boundary conditions.

2.1 Dirichlet boundary condition

Setting $\alpha = 0$ and $\beta = 1$ in (2.1b), we obtain the Dirichlet boundary condition:

$$u = g, \qquad \text{on } \partial\Omega.$$

Because Ω is determined with some uncertainty, it is necessary to define g as the trace of a function $\psi \in H^1(\mathbb{R}^2)$, that is, $g = \psi|_{\partial\Omega}$. Likewise, f needs to be defined in $L^2(\mathbb{R}^2)$.

Assume now that $\Omega \subset \mathbb{R}^2$ is a bounded domain and $\overline{\Omega} = \overline{\mathbb{R}^2 - \overline{\Omega}}$ (i.e. there is no crack in Ω.) Further, let Ω_Δ denote a sequence of pixel domains with pixel size $\Delta \to 0$. Let us suppose that $\Omega_\Delta \to \Omega$, the convergence being understood in the following sense:

for every $x \in \Omega$, there exists $\Delta_0(x)$ such that $x \in \Omega_\Delta$, for all $\Delta \leq \Delta_0(x)$, and, for every $x \notin \overline{\Omega}$, there exists $\Delta_1(x)$ such that $x \notin \Omega_\Delta$, for all $\Delta \leq \Delta_1(x)$.

This assumption is satisfied when the pixel domains are constructed from digital images with a particular selection of the threshold parameter and increasing resolution ($\Delta \to 0$). Then we have

Theorem 1 (Stable and unstable domains) *Under the foregoing assumptions and definitions:*

1. *For every Ω_Δ, there exists a unique (mesh) solution $u_\Delta \in H^1(\Omega_\Delta)$.*

2. *There exists a family of unstable domains (with respect to the Dirichlet boundary condition) for which there is a sequence of Ω_Δ, such that u_Δ do not converge (in H^1) as $\Delta \to 0$.*

3. *There exists a family of domains Ω_Δ, called stable domains, for which $u_\Delta \to u$ in H^1 for any value of the parameter. By convergence, we mean convergence of u_Δ which is extended by $\psi \in H^1(\mathbb{R}^2)$ outside Ω_Δ.* ∎

For more details, see [1].

Remark: Most domains in practice are stable with respect to the Dirichlet boundary condition. ∎

Let us now assume that, for a particular Δ, the pixel domains Ω_{LOW} and Ω_{UP} are known. Then it is possible to estimate *a priori* or *a posteriori* the norms $\|u_{\widetilde{\Delta}} - u_{\text{LOW}}\|_{H^1}$ and $\|u_{\widetilde{\Delta}} - u_{\text{UP}}\|_{H^1}$ for any $\widetilde{\Delta} < \Delta$. This also provides bounds on $\|u - u_{\text{LOW}}\|_{H^1}$ and $\|u - u_{\text{UP}}\|_{H^1}$ for stable domains. We refer the reader to [2] for more details.

2.2 Neumann boundary condition

Taking $\alpha = 1$ and $\beta = 0$ in (2.1b), we now consider the Neumann boundary condition:

$$\frac{\partial u}{\partial n} = g.$$

We consider the particular case where $g = 1$, $f = 0$ (as before, g is the trace of a function ψ on $\partial\Omega$), and Ω is a unit circle centered at the origin. Then we have (cf. [3]):

Theorem 2 *Let u_Δ denote the solutions on the pixel domains Ω_Δ. $\Delta \to 0$ and u the solution on Ω. Then*

$$u_\Delta(0) \not\to u(0) \ \ as \ \ \Delta \to 0,$$

and for any Δ,

$$\frac{|u_\Delta(0) - u(0)|}{|u(0)|} \geq 0.2.$$

∎

Theorem 2 shows that Problem 2.1a with boundary condition $\dfrac{\partial u}{\partial n} = 1$ is not a well-posed problem in the case where we use pixel approximations of a circle. Hence the present mathematical formulation has to be used with utmost caution. The essence of the paradox mentioned in Theorem 2 is that the length of the boundary $\partial \Omega_\Delta$ does not converge to the length of $\partial \Omega$.

Remark: We observe in the literature that pixel meshes are sometimes used as finite element meshes after proper coarsening. This may yield erroneous results. Note that it is possible to re-formulate the boundary condition so that convergence is established (see [3]). This model can then be used in the computations. Also estimates of the error in the approximations on pixel domains can be obtained. ∎

Remark: If $g = 0$, then the problem mentioned in Theorem 2 does not occur. ∎

Remark: For Neumann problems, there exists a family of unstable and stable domains with respect to the Neumann boundary condition. Note that the families for the Dirichlet and Neumann boundary conditions are different. ∎

2.3 Newton boundary condition

The Newton boundary condition reads:

$$\alpha \frac{\partial u}{\partial n} + \beta u = g, \qquad \alpha > 0, \ \beta > 0.$$

This type of boundary condition takes into account convection (radiation) phenomena on $\partial \Omega$. Similarly to the Neumann boundary condition case, convergence cannot be established.

Domain uncertainty enters the category of examples where only the range of input data is available.

3. The Problem of Stochastic Input Data

Let D denote a domain with boundary ∂D. We now consider the model problem

$$-\nabla \cdot a(x, \omega) \nabla u \;=\; f(x, \omega), \qquad \text{in } D, \tag{3.1a}$$

$$u \;=\; 0, \qquad \text{on } \partial D. \tag{3.1b}$$

where $a(x, \omega)$ are $f(x, \omega)$ are stochastic functions.

For simplicity in the exposition, we will distinguish in the following two cases:

1. $a(x, \omega)$ is a stochastic function while $f(x)$ is a deterministic function.

2. $a(x)$ is a deterministic function while $f(x, \omega)$ is a stochastic function.

3.1 Stochastic a and deterministic f

In order to study the model problem 2, it is necessary to introduce a mathematical formulation in a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Here, Ω, $\mathcal{F}$, $\mathcal{P}$ denote the set of random events, the σ-algebra of subsets of Ω and the applicable probability measure respectively. If X is a real random variable in $(\Omega, \mathcal{F}, \mathcal{P})$ with $X \in L^1(\Omega)$, we denote its expected value by

$$E[X] = \int_\Omega X(\omega) \, dP(\omega) = \int_\mathbb{R} X \, d\mu(X).$$

Here μ is the distribution probability measure for X, defined on the Borel set B and $\mathbb{R}$, and given by

$$\mu(B) = P\left(X^{-1}(P)\right).$$

We will assume that $\mu(B)$ is absolutely continuous with respect to the Lebesgue measure; then there exists a density function for X, $\rho \colon \mathbb{R} \to \mathbb{R}^+$, such that

$$E[X] = \int_\mathbb{R} X\rho(x) \, dx.$$

We now consider the random functions $v \colon D \times \Omega \equiv \widetilde{D} \to \mathbb{R}$, where $D \subset \mathbb{R}^2$. We suppose that $v \in L^2(\widetilde{D})$, that is, the functions v satisfy $\|v\| < \infty$ with

$$\|v\|^2 = E\left[\int_D v^2(x) dx\right].$$

Remark: Because we wanted to keep the standard notation Ω for the set of random events, we introduce the notation D for the domain in $\mathbb{R}^2$. This should not be confused with the notation used in the previous section where $\Omega \subset \mathbb{R}^2$ denoted the domain. ■

We assume that $a(x, \omega) = \widetilde{a}(x)$ satisfies the uniform ellipticity condition, that is, there exists positive constants α_1 and α_2 such that

$$0 < \alpha_1 \leq a(x) \leq \alpha_2 < \infty \quad \text{a.e. in } \widetilde{D}. \tag{3.2}$$

It is obvious that

$$0 < \alpha_1 \leq E[a](x) \leq \alpha_2 < \infty. \tag{3.3}$$

Let V be the Hilbert space

$$V = \left\{ v \in \widetilde{D} \to \mathbb{R}; \|v\|_V^2 = E\left[\int_D a|\nabla v|^2 dx\right] < \infty; v(x, \omega) = 0 \text{ on } \partial D \right\}$$

and let $B(\cdot, \cdot)$ be the bilinear form on $V \times V$

$$B(v, w) = E\left[\int_{\widetilde{D}} a(x, \cdot)\nabla v(x, \cdot)\nabla w(x, \cdot)\, dx\right]. \tag{3.4}$$

Obviously, V is a Hilbert space of random functions and in view of the Lax-Milgram theorem, there exists a unique solution $u \in V$ of (2) such that

$$B(u, v) = E\left[\int_D fv\, dx\right], \qquad \forall v \in V. \tag{3.5}$$

Note that f is assumed to be a deterministic function. We shall also assume that $f \in L^2(D)$.

The stochastic function $a(x, \omega)$ can be written in the form

$$a(x, \omega) = E[a](x) + \sum_{n=1}^{\infty} a_n(x)X_n(\omega) \tag{3.6}$$

where $X_n(\omega)$ are mutually independent random variables with $E[X_n] = 0$ and $E[X_n^2] = 1$, $n \geq 1$. We will assume that X_n have bounded images $\Gamma_n = X_n(\Omega)$ with $\Gamma_n = (-\gamma_n, \gamma_n) \subset \mathbb{R}$, $\gamma_n > 0$ and consider the probability density function $\rho_n \colon \Gamma_n \to \mathbb{R}$ such that $0 < \beta_1 \leq \rho_n \leq \beta_2 < \infty$.

The expression (3.6) is the Karhunen-Loeve expansion (see e.g. [4], p. 478). We will assume that in (3.6), $n = 1, \cdots, N_0$ instead of $1, \cdots \infty$. Then we can directly use the Doob-Dynken lemma (see e.g. [5], p. 9) and write

$$a(x, \omega) = a(x, y) = E[a](x) + \sum_{n=1}^{N_0} a_n(x)y_n \tag{3.7}$$

where $y_n = X_n(\omega) \in \Gamma_n$ and $y = (y_1, \cdots, y_{N_0}) \in \Gamma = \Gamma_1 \times \cdots \times \Gamma_{N_0}$.

As an illustration, we show in Fig. 3.1 eight samples using (3.7) with $N_0 = 10$ and $x \in (-0.5, 0.5)$ where the value of $a(x, \omega)$ is ranging in the interval $(7, 9)$ and $E[a](x) = 8$.

We introduce the inner product

$$(v, w)_\rho = \int_\Gamma \rho(y)\int_D v(x, y)w(x, y)\, dxdy$$

with $\rho(y) = \rho_1(y)\cdots\rho_{N_0}(y_{1D})$ and denote by $\|\cdot\|$ the induced norm. We then define

$$L_\rho(\widetilde{D}) = L_\rho(D, \Gamma) = \{v \colon D \times \Gamma \to \mathbb{R}; \|v\| < \infty\}$$

$$W(D, \Gamma) = \{v \in L_\rho; \|w\|_W < \infty; v|_{\partial D \times \Gamma} = 0\}$$

103
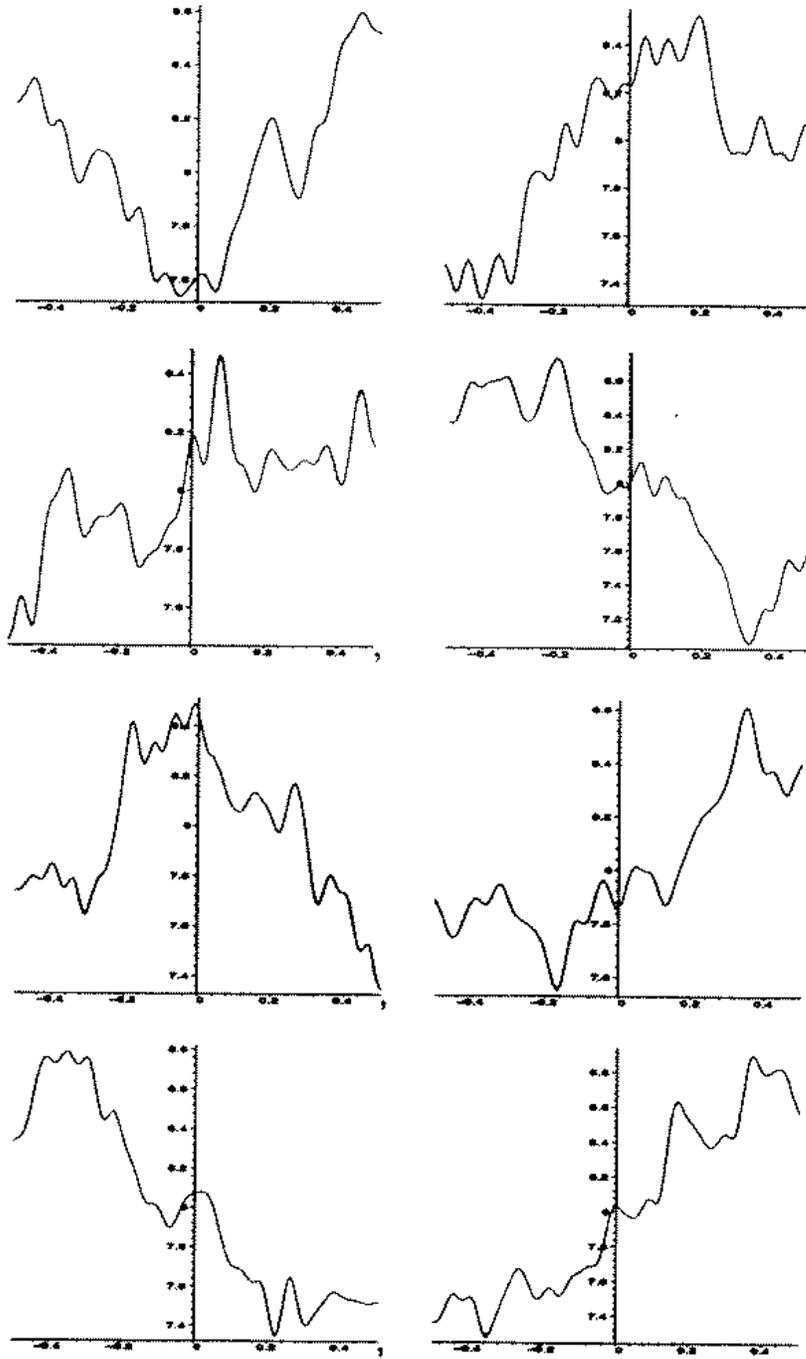
Figure 3.1: Samples of $a(x, \omega)$ with $N_0 = 10$ and $x \in (-0.5, 0.5)$ where the value of $a(x, \omega)$ is ranging in the interval $(7, 9)$ and $E[a](x) = 8$.

where

$$\|w\|_W^2 = \int_\Gamma \rho(y) \int_D a(x,y)|\nabla_x v(x,y)|^2 \, \mathrm{d}x\mathrm{d}y.$$

The space $W(D,\Gamma)$ is equivalent to the space V.

An equivalent formulation to (3.5) reads

$$\mathcal{B}(u,v) = (f,v)_\rho, \qquad \forall\, v \in W, \tag{3.8}$$

where

$$\mathcal{B}(v,w) = \int_\Gamma \rho(y) \int_D a(x,y)\nabla_x(x,y)\nabla_x w(x,y) \, \mathrm{d}x\mathrm{d}y. \tag{3.9}$$

The formulation (3.8) and (3.9) transforms the stochastic problem into a deterministic one. However, one major problem is the large dimension of Γ. For more details on the subject, we refer the reader to [6, 7].

The form (3.8) is very similar to the finite element method formulation. This was exploited in [6] in which finite elements are used. Results on the rate of convergence with respect to the norm $\|\cdot\|_W$ were obtained as well. Similarly, as in the usual finite element method, it was found that the rate of convergence for the functional is higher than that of the convergence in the norm $\|\cdot\|_W$.

Because of computational complexity, it is often necessary to approximate (3.7) using a value of N_0 as small as possible and then estimate the error due to that approximation. This was done among others in [7].

There basically exist four approaches to solve Problem (3.8), namely, using

1. The Monte Carlo or Quasi-Monte Carlo Method;

2. The h, p or h-p versions of the Finite Element Method;

3. The method of successive approximations;

4. A combination of all of the above methods.

We considered in [6] the Finite Element Method while we developed in [7] the theory of successive approximations with estimation of the error when approximating (3.7) using a smaller value of N_0. Issues associated with large values of N_0 will be addressed in the future.

3.2 Deterministic a and stochastic f

We have addressed so far the case of $a(x, \omega)$ being stochastic and $f(x)$ being deterministic. Let us now discuss the case when $a(x)$ is deterministic while $f(x, \omega)$ is stochastic. This case can be cast into a deterministic form in a similar manner as before.

Suppose that one is only interested in the mean value and the covariance of the solution. It can be shown that the mean value is the solution of a deterministic problem on D with the mean value of f on the right-hand side. Moreover, the covariance of the solution can be determined directly from the covariance of the right-hand side. It is actually the solution of a fourth-order partial differential equation on $D \times D$ with mixed order for the leading coefficients. This was shown in [6, 8]. The authors in [9] used this approach and proposed a sophisticated numerical solution which is not influenced by the high dimension of the problem. Note that the problem for the Laplace equation in three dimensions leads to a problem in six dimensions.

Although we discussed only the cases for stochastic coefficients and right-hand side and linear equations, the approach transforming the stochastic problem in a deterministic one for example for nonlinear equations, stochastic boundary, etc.

The major issue is the high dimensionality of the problem. Very likely, the construction of the probability field from the experimental data, respectively its characterization by (3.7), should be accomplished after a careful analysis of the data so that minimal dimensionality is achieved. If the stochastic functions are almost without correlation (i.e. nearly white noise), then the approach described here cannot be applied due to the necessity of using large N_0. We refer to [10] for more details. How to treat this case is an open problem which needs to be addressed.

Remark: The Karhunen-Loeve expansion is also used by Ghanem (cf. [11, 12]). ∎

4. Conclusions

The existence of uncertainties in the input data to solve problems of computational mechanics is almost inevitable: moreover, these uncertainties are sometimes important and can have a dramatic effect on the computational predictions.

The present report addresses some aspects of the treatment of uncertainties. The influence of the uncertainty of the domain on the solution of elliptic PDE's was examined using a worst scenario approach. This worst scenario approach characterizes the solution set which includes the image of the input data set. It was found that stable and unstable domains exist for both Dirichlet and

Neumann boundary conditions where the domains are described by scanning. Here stability of the domain refers to the convergence of the numerical solution for the displacements. For the Dirichlet boundary condition, most domains in practice are, in fact, stable.

The other portion of the work described in this chapter dealt with the stochastic solution of elliptic PDE's which have stochastic coefficients with variance. *A priori* error analysis of the projection and perturbation solution method was performed.

The issues associated with uncertainties and their treatment are a subject of great concern in the research community, and are directly related to validation. Very few rigorous results seem to be available in the literature to date, although many engineering papers have been published on the subject. We believe that a systematic research for the treatment of uncertainties is now needed.

REFERENCES

[1] I. Babuŝka, Stability of domains with respect to the basic problems in the theory of partial differential equations, mainly the theory of elasticity, I & II, *Czechoslovak Math.* **11**(86), 1961, 76–105, 165–203 (in Russian).

[2] I. Babuŝka and J. Chleboun, Effects of uncertainties in the domain on the solution of Dirichlet boundary value problems, TICAM Preprint, 2001.

[3] I. Babuŝka and J. Chleboun, Effects of uncertainties in the domain on the solution of Neumann boundary value problems in two spatial dimensions, *Math. Comp.* (to appear).

[4] M. Loeve, *Probability Theory*, 4th Ed., Springer-Verlag, New York, 1997.

[5] B. Oksendal, *Stochastic Differential Equations – An Introduction with Applications*, Springer-Verlag, 1998.

[6] M. Deb, I. Babuŝka, and J.T. Oden, Solution of stochastic partial differential equations using Galerkin finite element techniques, *Computer Methods in Applied Mechanics and Engineering*, (to appear).

[7] I. Babuŝka and D. Chatzipantelidus, On solving elliptic stochastic partial differential equations, TICAM Preprint, 2001.

[8] I. Babuŝka, On Randomized Solution of Laplace's Equation, *Cas. Pest. Mat.* **86**, 1961, 269–275.

[9] Ch. Schwab and R.A. Todor, Finite element method for elliptic problems with stochastic data, ETH Preprint, 2001.

[10] H. Holden, B. Oksendal, J. Uboe, and T.S. Zhang, *Stochastic Partial Differential Equations – A Modeling White Noise Functional Approach*, Birkhauser, 1966.

[11] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, 1991.

[12] R. Ghanem, Ingredients for a General Purpose Stochastic Finite Elements Implementation, *Computer Methods in Applied Mechanics and Engineering* **168**, (1-4), 1999, 19–33.

CHAPTER 4

Geometry and Visualization Support

Chandrajit Bajaj

1. Introduction

Our goals were to provide geometry and visualization support for both boundary element and three-dimensional finite element formulations of linear to nonlinear viscoelasticity computations on hierarchical models of heterogeneous materials. The domains were all assumed to be volumetrically scanned by a CT (computer tomography) device which yields close to micron feature size separation.

As a first attempt to generate boundary and 3D finite element meshes from volumetric tomographic imaging data, we utilized the results of our earlier work on the contour spectrum [1]. The contour spectrum allows the development of an adaptive ability using multiple global signature functions to separate interesting isovalues of the imaging function to yield good initial segmentation. Effective global signatures include surface area and volume fractions as functions of image intensity. This coupled with our fast isocontouring methods yielded fairly accurate boundary and 3D finite element, triangular meshes. The spectrum signature graphs as well as the interactive display of extracted boundary meshes provide good exploratory visualization of volumetric imaging data.

Our subsequent attempts on geometry and visualization tools led to the research and development of better and more automatic segmentation techniques for two material domains, with given volume fraction of the two materials. A progressive contouring algorithm coupled to a multiresolution representation of such segmented boundaries, are the contents of paper [2]. An evolution (anisotropic diffusion) time based method [7], starting from initial segmentation, generated using spectrum signatures [1], is currently being experimented with to provide multi-scale and smooth (higher order spline) representations of segmented boundaries in CT images.
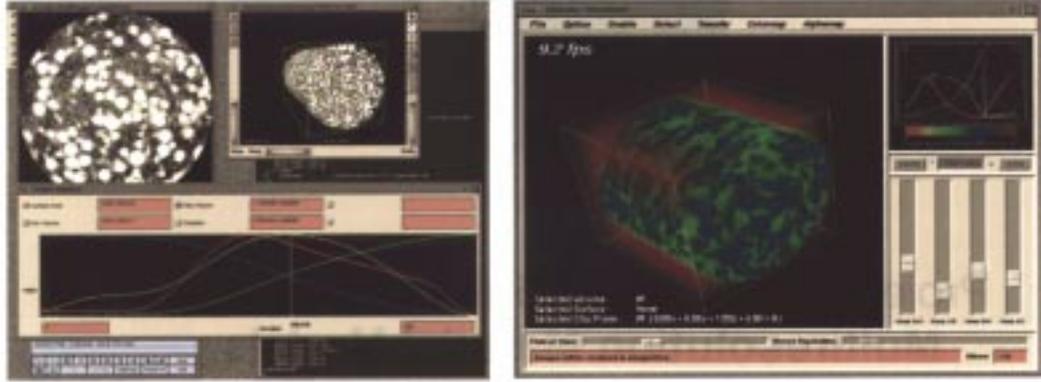
Figure 2.1: Using the contour spectrum and volume fraction graphs to segment and visualize the speckled inner surface from the embedding matrix material.

A new hexahedral mesh generator, based on a trivariate subdivision scheme, and complete with an interactive mesh visualization tool was developed. Several CT scans of an epoxy material with spherical glass inclusions were generated. The above segmentation and visualization tools were applied to generate the precise location of the spherical inclusions. The results of the hexahedral mesh are summarized in report [5]. The results of the anisotropic diffusion method [6] are also being currently experimented with to yield a more sophisticated hexahedral element quality improvement scheme [7].

2. Segmentation and Visualization

2.1 Approach based on the Contour Spectrum

In addition to computational and space complexity issues, user interfaces have a tremendous impact on the interactivity of a visualization environment. A *contour spectrum* consists of computed metrics over the volumetric imaging dataset (in general, a scalar field). On the basis of such metrics we can define a set of functions which provide a useful tool to enhance the interactive query of the dataset. One primary advantage of the contour spectrum interface is that it allows one to display in a 2D image a "global" view of the examined scalar field, independent of its dimension.

Consider a 2D scalar field composed of triangles t_i and vertices v_i. We build and display the spline function $L(w)$ whose value $L(w_0)$ is the length of the isocontour of height w_0. $L(w)$ can be computed as the sum of all the contributions $L_i(w)$ given by each cell c_i to the length of the

110

contours:

$$L(w) = \sum_i L_i(w)$$

Thus, we can concentrate on the computation of the generic term $L_i(w)$ associated with the triangle t_i. Triangle t_i has vertices v_1, v_2 and v_3 with height values $\mathcal{F}(v_1) \leq \mathcal{F}(v_2) \leq \mathcal{F}(v_3)$. Given the equation $f(x, y, w) = 0$ of the plane containing t_i, the value $L_i(w_0)$ is the length of the intersection between t_i' (projection of t_i onto the mesh space) and the 2D line of equation $f(x, y, w_0) = 0$. As we change the value of w_0 we obtain the measure of all the slices parallel to the line $f(x, y, 0) = 0$. In general it is know from spline theory that given a d-simplex in $\mathbb{R}^d$ the function that gives the measure of all the parallel slices of such simplex (that is the measure of the intersection with a set of parallel hyperplanes) is a degree $d - 1$, C^{d-2} continuous, B-spline function.

In the 2D case the B-spline is simply a piecewise linear C^0 function. Hence we need only compute the length of the segment for $w = \mathcal{F}(v_2)$ and connect it with the other two extremes for which the length is 0. Note that the B-spline formulation of the length is also useful to automatically handle the eventual degenerate cases. For example a portion of the terrain at height w can be a flat parallel to the x, y plane (a lake). In this case there occurs a definition problem, in determining the length of an isocontour which is partially a 1-dimensional curve and partially a 2D surface. The natural solution is to remove the flat region to regularize the dimension of the contour. The consequence is that the function that computes the contour length is only C^{-1} at the height w. Using the B-spline approach no special care must be taken for this case since the knot vectors of the flat triangles are $\mathcal{F}(v_1) = \mathcal{F}(v_2) = \mathcal{F}(v_3)$ resulting in "valid" splines which shrink to a point as expected.

As already pointed out, the above spline function can be computed for simplices of any dimension. For the 3D case of a tetrahedron (v_1, v_2, v_3, v_4) with scalar function values $(\mathcal{F}(v_1) \leq \mathcal{F}(v_2) \leq \mathcal{F}(v_3) \leq \mathcal{F}(v_4))$ we have a degree two C^1 B-spline. In this case the determination of the control polygon is as follows: Again for each cell we obtain a spline function. The sum of the splines associated to each cell is a single spline that gives the contour area for any isovalue.

While length and area are important metrics to report, in many cases they are not sufficient to guide the user in choosing appropriate isovalues. In many situations the user is interested in finding and displaying prominent surfaces in the data. Toward this end we have designed a metric which is based on the *slope* or *gradient* of the function. The difficulty with the gradient measure is to define it properly, since along a particular contour the gradient of the scalar filed is not (usually) constant. To

compute a consistent (single valued) gradient function we resort to the spline decomposition of the contour length/area function. For each triangle/tetrahedron of the mesh we have a spline function which gives the length of any contour within that triangle/tetrahedron. Moreover, by piecewise linear approximation, within each triangle/tetrahedron the gradient of the scalar field is constant. Hence to determine the contribution to the gradient function of the contours within a single triangle we just need to multiply the length function by the absolute value of the (constant) gradient. Again the sum of the splines defined in each triangle/tetrahedron gives a single global spline function which defines the gradient integral of any isocontour in the scalar field. The maximum of the gradient (yellow function plot) corresponds to the isocontour (red contour on top figure) bounding the relevant portion of the data. Note how the maximum of the contour surface (red function plot) is attained for a lower height value of the field. It captures the noisy part of the data that has a large contour length due to the numerous components.

While the display of contour metrics is both helpful and informative, there is clearly a lack of global structural information in the metrics described. For example, there is no indication of features such as local maxima and minima of the field. For this purpose we introduce the use of the *contour tree* as a tool for assisting the user in interaction with complex scalar fields. A contour tree captures the global changes in contour topology of the scalar field defined on the input the mesh. For example, in the display of an isosurface (threshold surface) of a CT image, one contour component maybe be hidden inside another. If we associate the isocontour display with the contour tree it becomes immediately clear that the current isosurface is composed of two components and hence we might need a clipping plane to look inside the current surface. It has been used before in image processing and GIS research. Another name in use is the *topographic change tree*, and it is related to the *Reeb graph* used in Morse Theory.

The user interface for presenting the contour spectrum takes on two forms. For static imaging data, a window presents a selected subset of the computed data characteristics in 1D plots. The horizontal axis represents the isovalue dimension. The vertical axis represents the range of each function, all of which are normalized for overlapping display. See Fig. 2.1 for an example. The user may select a subrange of the isovalues for display in order to enhance the local detail in the computed metrics. Vertical bars represent the current isovalues, which the user may change with a familiar click-and-drag operation. With time-varying data, it is desirable that the user have the ability to quickly browse all parameters of the visualization. In this case we use the vertical dimension of the interface as an index into the time step of the data. Of course, while we use $time$ here as an example,
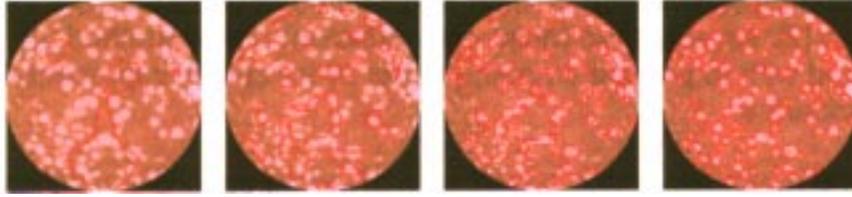
Figure 2.2: Adaptive and Multiresolution approximation of the boundary of inclusions of a different material, from the base embedding material. This multiresolution segmentation approach is based on progressive isocontouring refined by edge bi-section.

other parameters may be varied similarly, such as input parameters to a numerical simulation. Using this interface, each point in the 2D display maps to a number of functions. We selectively display one function at a time by pseudocoloring of the function values over the 2D grid.

2.2 Hierarchical approach

Multi-resolution representations are a key tool used in image processing and visualization to achieve real-time interaction with large data sets. A great deal of research has been focused on the off-line construction of such representations mostly using decimation schemes. Drawbacks of this class of approaches include: (i) the inability to maintain interactivity when the displayed surface changes frequently, (ii) inability to control the correct embedding (no self-intersections) of any approximated level of detail of the output surface. In our paper [2], we introduce a technique for on-line construction and smoothing of progressive isocontours. Our hybrid approach combines the flexibility of a progressive multi-resolution representation with the advantages of a recursive subdivision scheme. Our main contributions are: (i) a progressive algorithm that builds a multiresolution surface by successive refinements so that a coarse representation of the input is provided, (ii) application of the same scheme to smooth the surface by mean of a 3D recursive subdivision rule, (iii) a multi-resolution representation where any adaptively selected level of detail is guaranteed to be consistently embedded in 3D space (no self-intersections).

3. Mesh Generation and Visualization

In a landmark paper, Catmull and Clark [3] described a simple generalization of the subdivision rules for bi-cubic B-splines to arbitrary quadrilateral surface meshes. This smooth subdivision scheme has become a mainstay of surface modeling systems. MacCracken and Joy [4] described a general-

Figure 3.1: A ring modeled using MLCA subdivision with creases.

ization of this surface scheme to volume meshes. Unfortunately, little is known about the smoothness and regularity of this scheme due to the complexity of the subdivision rules. Our paper [5] presents an alternative subdivision scheme for hexahedral volume meshes that consists of a simple split and average algorithm (MLCA = multi-linear cell averaging). Along extraordinary edges of the volume mesh, the scheme provably converges to a smooth limit volume. At extraordinary vertices, we supply strong experimental evidence that the scheme also converges to a smooth limit volume. The scheme automatically produces reasonable rules for non-manifold topology and can easily be extended to incorporate boundaries and embedded creases expressed as Catmull-Clark surfaces and B-spline curves.

Our given scanned and segmented CT specimen (epoxy matrix with embedded glass spherical inclusions) can be considered a cuboid in $\mathbb{R}^3$ with several spheres in the cuboid as inner boundaries. We assume that these spheres do not intersect each other (see Fig. 3.2). The aim is to construct
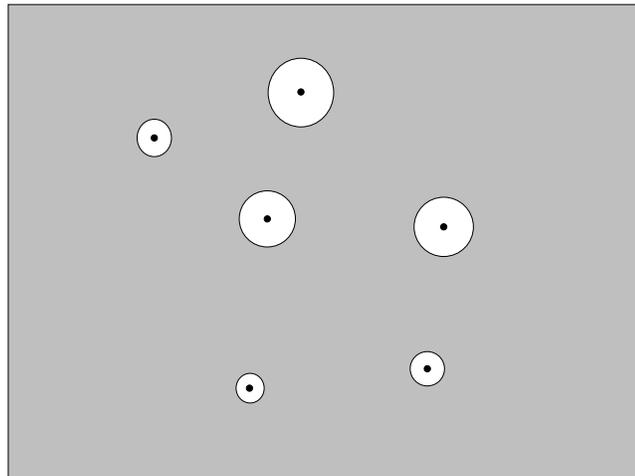
Figure 3.2: The initial scanned and segmented specimen volume with sphere inclusions as inner boundaries

hexahedral meshes that conform with these spheres and have certain specified adaptive features. For instance, the part of the constructed hexahedral mesh that is close to the spheres may be required

denser than the part that is away from the spheres.

Our algorithm utilizes the following theorem on hexahedral mesh construction.

Theorem 1 – *Any simply connected 3D domain with an even number of quadrilateral boundary faces can be partitioned into a hexahedral mesh respecting the boundary.*

3.1 Mesh construction steps

In this section, we describe steps of the mesh construction with the help of 2D figures.

Step 1. Construct the 3D weighted Voronoi diagram from the center points of the spheres (see Fig. 3.3). The weight of a sphere center is chosen to be proportional to the radius of the sphere.
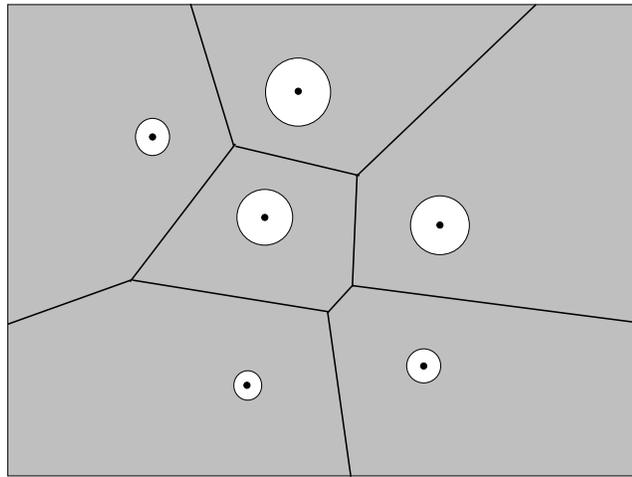
Figure 3.3: Step 1: Partition the initial cuboidal volume by Voronoi cells. Each cell is a polyhedron in 3D.

Step 2. Merge each of the short edges into one vertex. An edge is regarded as short if its length is less than a user specified threshold value.

Note that this merging may eliminate triangles (see Fig. 3.5). Also note that this merging may lead to the intersection between the boundaries of polyhedral cells and the surface of the spheres. Hence, it is necessary to check that if the merging leads to the intersection. If such an intersection happens when merging an edge, we do not carry out this edge merging. The aim of this step is to avoid producing tiny hexahedral elements. Hence this step is optional.

Step 3. Quadrilateralize each face of Voronoi volume cells (see Fig. 3.6). For each Voronoi cell, connect the center with the vertices and then form a pyramidal partition of the cell. Each pyramid
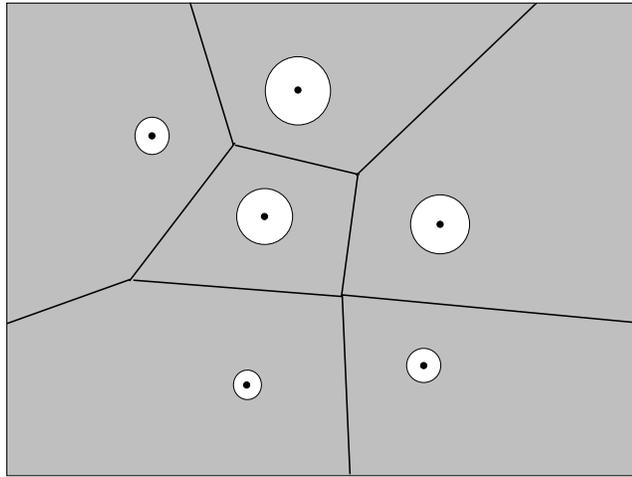
Figure 3.4: Step 2: Partition each Voronoi cell into pyramids.
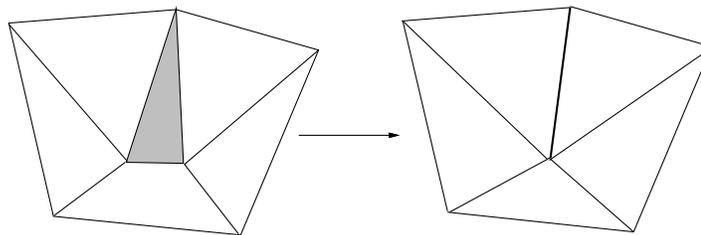
Figure 3.5: Step 2: Merging an edge leads to a collapse of a triangle.
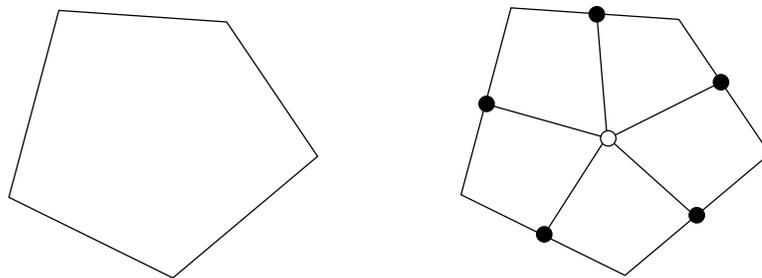
Figure 3.6: Step 3: Subdivide each face polygon of the Voronoi cell into quadrilaterals. In the figure, the empty dot is the centroid of the polygon, while the red dots are the mid-points of the edges.

is further subdivided into a hexahedral and another pyramidal cell by the surface of the sphere (see Fig. 3.7). After this step, the surface of each Voronoi cells is a triangulated polygon. The inner region of the sphere is partitioned into tetrahedra, the outer region of the sphere is partitioned into prisms.
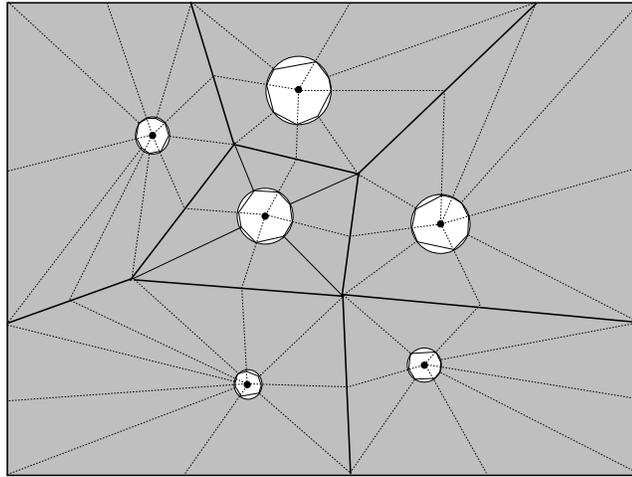

Figure 3.7: Step 3: Partitioning each Voronoi volume cell into prisms and tetrahedra.

Step 4. Adaptive subdivision of each cell in the radial direction. In order to avoid self-intersection, the points around spheres may need to be moved outwards a little. The aim of this step is to obtain adaptive meshes. Hence it is optional.
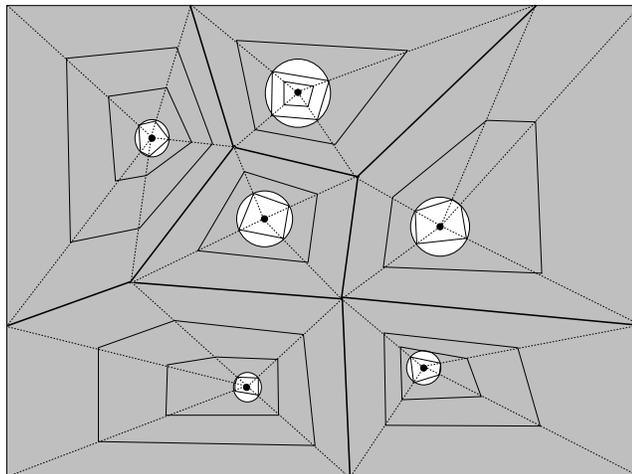

Figure 3.8: Step 4: Adaptive subdivision in the radial direction.

Step 5. Subdivide each prism into 3 hexahedra and split the tetrahedron into a hexahedron by

117

adding three vertices on the edges and one vertex on the face (see Fig. 3.9–3.10). After this step, each cell is partitioned into hexahedra (see Fig. 3.11)
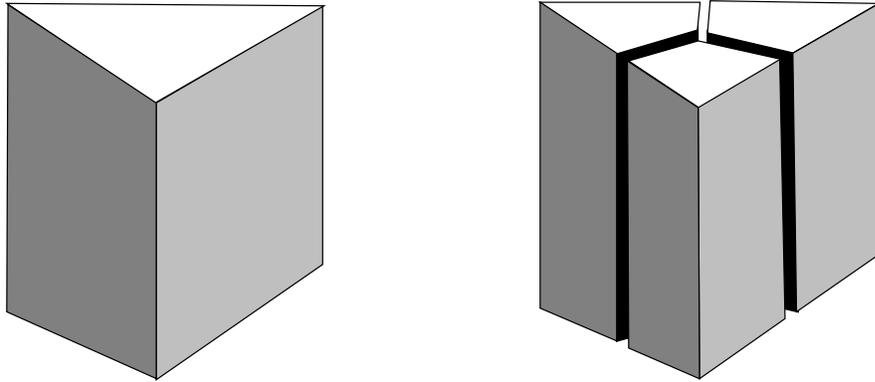

Figure 3.9: Step 5: Subdivide a prism into 3 hexahedra and a tetrahedron into four hexahedra.
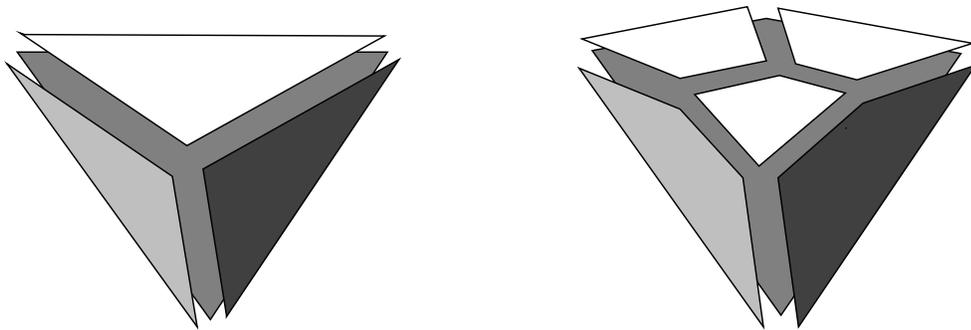

Figure 3.10: Step 5: Split a tetrahedron into a hexahedron by adding three vertices on the edges and vertex on the face. The figure shows the three triangular faces of a tetrahedron and six quadrilateral faces of a hexahedron.

Step 6. Smoothing the mesh produced by the last step using the MLCA recursive subdivision scheme with restriction of keeping the shape of inner and outer boundaries. The smoothing is conducted in three sub-steps.

   A. Smoothing on the creases of the outer boundary.

   B. Smoothing the inner and outer boundaries by 2D smoothing scheme with fixed vertices on the creases.

   C. Smoothing the volume with fixed boundary vertices.

118
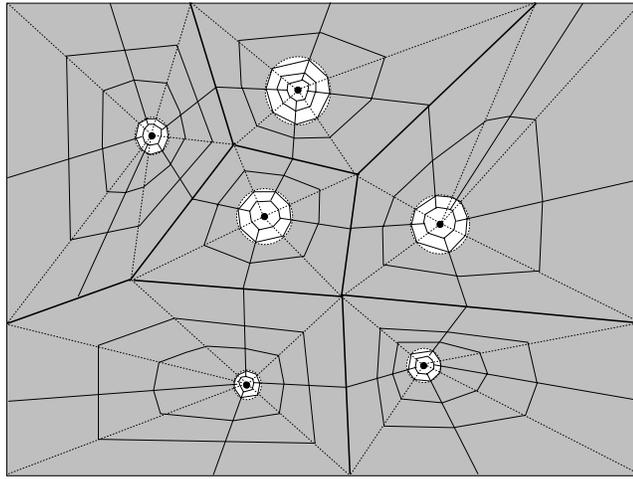
Figure 3.11: Step 5: Subdivide each prism into 3 hexahedra and change each tetrahedron into one hexahedron accordingly.
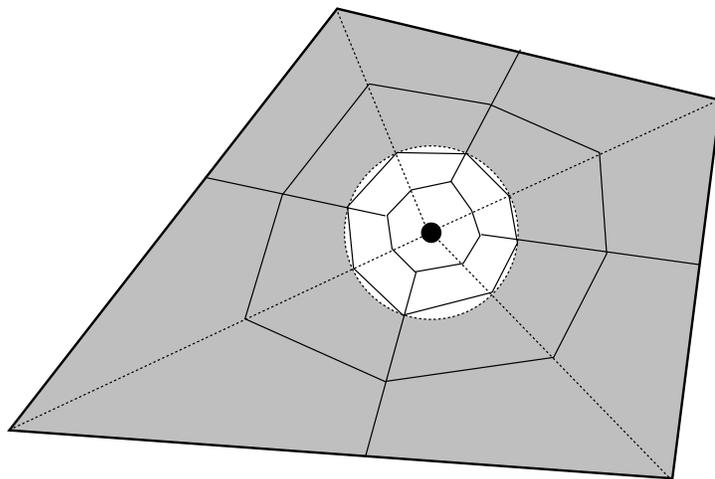

Figure 3.12: Step 5: Enlarged view of the center cell of Fig. 3.9.

119
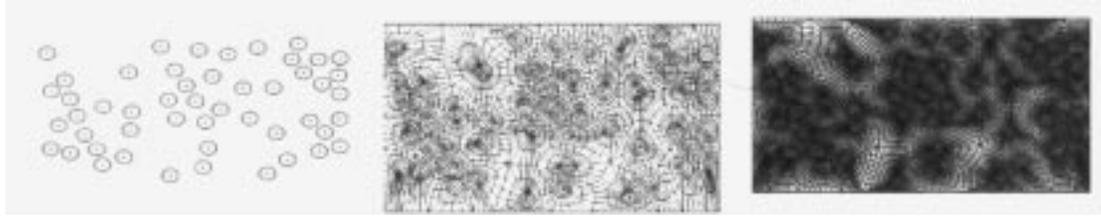
Figure 3.13: A final 2D quadrilateral mesh of the segmented image object with circular inclusions.
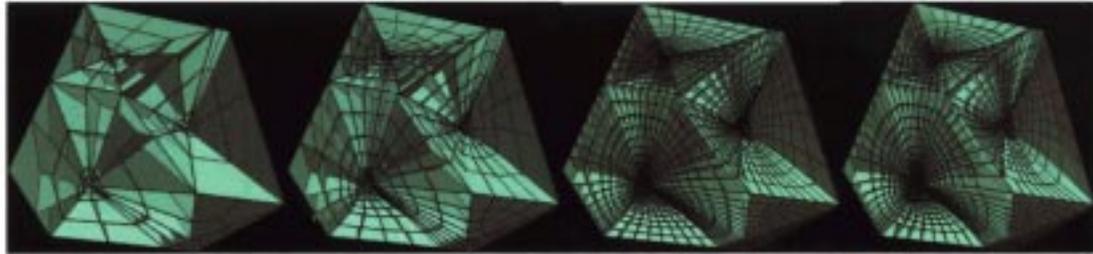

Figure 3.14: Cutaways of a 3D hexahedral mesh of a cuboidal domain of epoxy material with spherical glass inclusions.

The aim of this step is to produce an improved quality finite element mesh. The smoothing could be iterated several times. Since the shape of the inner boundaries is preserved, the mesh after smoothing is still adaptive.

Step 7. Further refine the mesh using MLCA subdivision scheme with the constraint of keeping the shape of the inner and outer boundaries. The aim of this step is to produce a sequence of meshes that have hierarchical nature. From this set of meshes, one can choose the one that has the ideal resolution.

Features of the Constructed Mesh. The mesh constructed is adaptive in the sense that it is denser in all regions that are close to the sphere. The mesh density varies linearly decreasing away from the spheres.

4. New Techniques

Our primary goal is to filter the noise from noisy manifolds, noisy manifold functions [6] and 3D imaging data [7] at multiple scales, so as to improve segmentation and visualization. Our secondary goal is to construct continuous (non-discretized) multi-scale representations for smoothed surface geometries and segmented volumetric imaging function data.

In paper [6] we treat discrete surface data (2-manifolds) in $\mathbb{R}^3$ and function vector data defined on the surfaces. A surface and a $\kappa - 3$ dimensional function vector data on the surface can be considered as a discretization of a 2-manifold embedded in $\mathbb{R}^\kappa$. We establish a unified anisotropic diffusion model for such manifolds aiming at smoothing (fairing) out noise both in the 2-manifold in $\mathbb{R}^3$ and the 2-manifold in $\mathbb{R}^\kappa$, while enhancing curve features on both 2-manifolds. We combine the C^1 limit function representation of Loop's subdivision for triangular surface meshes and vector functions on the surface mesh with anisotropic diffusion in a parameterized time setting, to arrive at a sparse linear system of equations. Iteratively, solving the sparse linear system, yields a sequence of faired (smoothed) meshes as well as faired functions with specified feature curves, enhanced. See Figure below.

Figure 4.1: Smoothing the geometry of the model head of Picard (146,036 triangles). The second and third figures are the meshes after 1 and 4 steps of smoothing. The time step is 0.001.

Our recent geometry driven anisotropic diffusion work on volumetric imaging data [7], is directly related with Preuber and Rumpfs level set method for anisotropic geometric diffusion in 3D image processing. The core of their method is an evolution driven by geometric diffusion of level surfaces. They construct anisotropic diffusion tensors based on prefiltered principal curvatures and principal direction of curvature. Their main idea is to decrease diffusivity in the dominant principal curvature. Our method presented in [7] can achieve both feature enhancement and coherence enhancement on the local structure, while decreasing the energy of the smooth image as well.

In [7], we construct 3D diffusion tensors by taking into account three directions of local structure.

The terms of the diffusion tensor are created with the following basic rules: If noise along the normal direction is to be eliminated as well, this may be accomplished by encouraging diffusion along the normal direction. The diffusion is larger at those locations which have larger likelihood
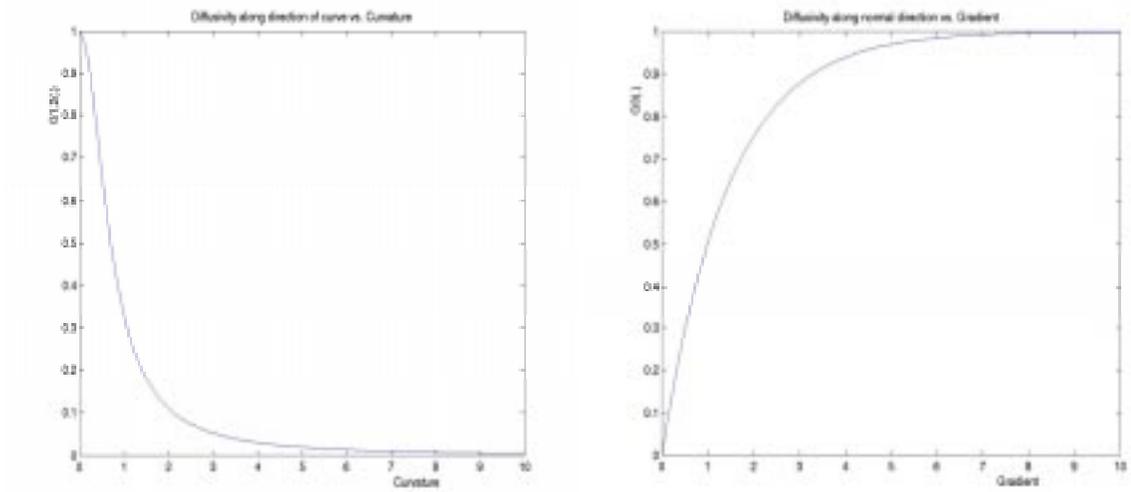
Figure 4.2: (Left) The diffusion along the curve direction is a function of curvature of the local structure. (Right) The diffusion along the normal direction is a function of gradient of local structure.

to be noisy, where the likelihood is measured by the gradient and choice of a suitable function controlling the diffusion rate.

We are now attempting to apply the above 3D combined, filtering and segmentation, approach to the scanned volumetric CT specimen (epoxy matrix with embedded glass spherical inclusions) to yield better boundary and 3D finite-element meshes.

5. Concluding Remarks

We have developed and experimented with several 3D image processing, mesh generation and visualization support for CT imaging data as part of our effort to provide computational infrastructure for the hierarchical models of heterogeneous materials project. Details of the intricacies of the developed segmentation, meshing and visualization algorithms as well as results of our experimentation have been summarized above. Further research is needed to determine full applicability of using evolutionary time-based PDE approaches to both segmentation and noise reduction schemes. The initial results look very promising.
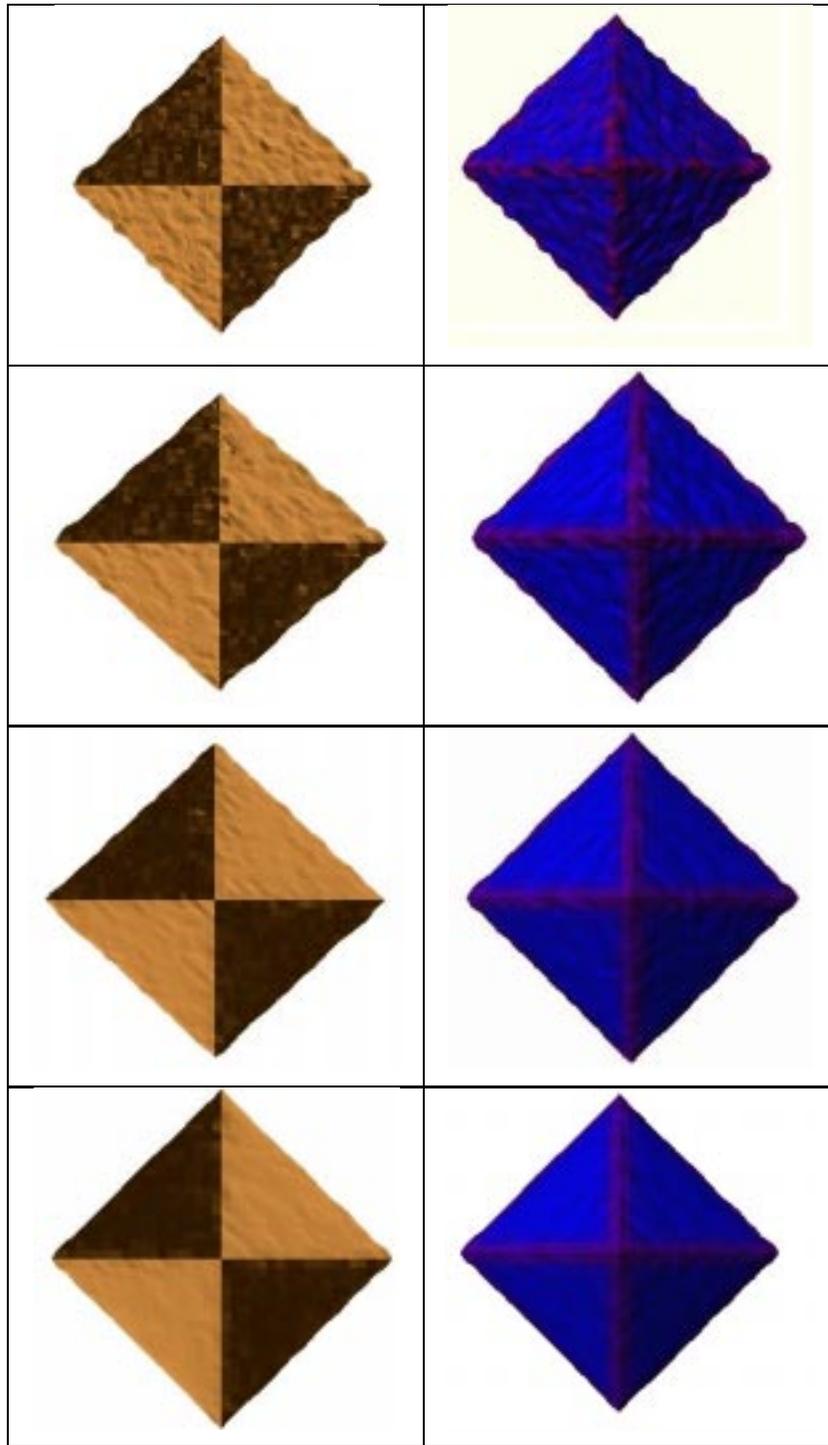
Figure 4.3: The left column shows an icosahedral isosurface of the function $\mathrm{abs}(x) + \mathrm{abs}(y) + \mathrm{abs}(z)$ + noise, at different steps of the anisotropic diffusion process. The right column is the color ramp from blue to red indicating the dominant curvature value.

<u>REFERENCES</u>

[1] C. Bajaj, V. Pascucci, and D. Schikore, The Contour Spectrum, *Proc. of the IEEE Visualization Conference*, Phoenix, Arizona, pp. 167–173, 1997.

[2] C. Bajaj and V. Pascucci, Time Critical Adaptive Refinement and Smoothing *Proceedings of the ACM/IEEE Volume Visualization and Graphics Symposium*, pp. 33–42, 2000.

[3] E. Catmull and J. Clark, Recursively Generated B-spline Surfaces on Arbitrary Topological Meshes, *Computer Aided Design*, Vol. 10, pp. 350-355, 1978.

[4] R. MacCracken and K. Joy, Free-form Deformations with Lattices of Arbitrary Topology, *Proceedings of the 23rd International Conference on Computer Graphics and Interactive Techniques*, New Orleans, Louisana, pp. 181-188, 1996.

[5] C. Bajaj, J. Warren, and G. Xu, A Smooth Subdivision Scheme for Hexahedral Meshes, *The Journal of Visual Computer*, Special Issue on Recursive Subdivision, to appear, 2001.

[6] C. Bajaj, and G. Xu, Anisotropic Diffusion of Noisy Surfaces and Noisy Functions on Surface, *TICAM Report 01-07*, The University of Texas at Austin, February 2001.

[7] C. Bajaj, Q. Wu, and G. Xu, Anisotropic Diffusion of Noisy Volumetric Images, *TICAM Report 01-xx*, The University of Texas at Austin, July 2001.

DISTRIBUTION:

10   TICAM
The University of Texas at Austin
Attn: J. T. Oden
        I. Babuška
        G. Rodin
        C. Bajaj
        S. Prudhomme
        K. Vemaganti
201 E. 24th Street, ACE 6.412
Austin, TX 78712

1   Poznan University of Technology
Attn: Mieczyslaw S. Kuczma
5 Maria Sk^{3}odowska-Curie Sq.
60-965 Poznan, Poland

1	MS 0316	M.D. Rintoul, 9209
1	MS 0321	W.J. Camp, 9200
1	MS 0555	M.S. Garrett, 9122
1	MS 0557	T.J. Baca, 9125
1	MS 0824	J.L. Moya, 9130
1		A.C. Ratzel, 9110
1	MS 0826	K.D. Copps, 9143
1		J.R. Stewart, 9143
1	MS 0835	J.M. McGlaun, 9140
1	MS 0841	T.C. Bickel, 9100
1	MS 0847	C.R. Dohrmann, 9226
1		M.W. Heinstein, 9142
1		J. Jung, 9126
5		S.W. Key, 9142
1		R.W. Leland, 9226
1		R.A. May, 9126

1	MS 0847	J.A. Mitchell, 9142
1		H.S. Morgan, 9120
1		J.M. Redmond, 9124
1		J.R. Red-Horse, 9211
1	MS 0888	D.B. Adolf, 1811
1	MS 0893	R.M. Brannon, 9123
1		R.S. Chambers, 9123
10		D.C. Hammerand, 9123
1		W.M. Scherzinger, 9123
1	MS 9042	M.F. Horstemeyer, 8728
1	MS 9405	D.J. Bammann, 8726
1	MS 9161	E.P. Chen, 8726
1	MS 0188	Laboratory Directed Research & Development
1	MS 9018	Central Technical Files, 8945-1
2	MS 0899	Technical Library, 9616
1	MS 0612	Review & Approval Desk, 9612 For DOE/OSTI