COMPUTABILITY IN NONLINEAR SOLID MECHANICS

by

T. Belytschko[1] and K. Mish[2]

1. Walter P Murphy Professor of Computational Mechanics and Chair, Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208.
2. Director, Center for Computational Engineering, Lawrence Livermore National Laboratory, Livermore, CA

ABSTRACT

The computability of nonlinear problems in solid and structural mechanics problems is examined. Several factors which contribute to the level of difficulty of a simulation are discussed: the smoothness and stability of the response, the required resolution, the uncertainties in the load, boundary conditions and initial conditions and inadequacies and uncertainties in the constitutive equation. An abstract measure of the level of difficulty is proposed, and some examples of typical engineering simulations are classified by this measure. We have put particular emphasis on engineering calculations, where many of the factors that diminish computability play a prominent role.

1. INTRODUCTION

We address here the question of how well the mechanical response of solids can be computed, particularly at the edges of the performance envelope where failure of the structure or solid is possible. We will call this attribute of physical behavior its *computability*. The word computability is used in a sense that differs from its prevalent meaning in computer science. In computer science, the meaning of computability today is given in the sense of Turing: an abstract problem is computable if there exists an algorithm that can solve the problem in a finite number of steps. The most familiar manifestation of this principle is the Turing Test, which examines whether a computer program can imitate the behavior of a human mind, with the computer thus cast as a thinking machine. Our usage of the term *computability* is in a similar spirit, but applied to von Neumann's model of a computer as a numeric engine. We use the term computability to refer to how well a physical process can be computed, or in other words, how well a simulation can predict the physical response of a mechanical system. In this sense, computability refers to whether simulation can successfully mimic physical experiments, in the same way that a computer passing Turing's test mimics a human brain. Our meaning for the word computable is thus almost synonymous with predictable. We have chosen not to use the term predictable because the word "predict" has many connotations arising from its common usage, although we will sometimes use the phrase "predictive capability of a program."

It is often tacitly assumed that the main obstacle to computability is the lack of sufficiently powerful computers. Here we explore the alternative viewpoint that the behavior of solids and structures in certain regimes can only be computed with a limited degree of precision, regardless of the power of the computers applied to the problem, and we examine the factors that limit computability, i.e. the barriers to computability. These factors are combined through an abstract measure that we call the level of difficulty, which is inversely proportional to computability. The intent of this measure is to provide some guidance of the expected difficulty of computing a particular response of a structure or solid.

We have focused our discussion on problems of solid and structural mechanics, but some of the concepts are more generally applicable. For example, we will argue that the major impediments to computability are the degree of roughness and the degree of instability of a physical process. These factors are barriers to computability in many other fields, such as fluid dynamics and weather prediction. However, while these other fields are also challenged by the difficulties arising from chaotic phenomena, chaos has been of little concern in most solid mechanics problems. However, the difficulties arising from loss of stability, and the range of scales over which this occurs, pose severe difficulties in all of these fields.

The effectiveness of simulation software is currently evaluated in two steps:
1. verification
2. validation

Verification consists of establishing that the governing equations are correctly discretized the algorithms are correctly programmed. For example, verification examines whether the discrete equations are convergent, and that the many branches of the program which implement different features are correctly programmed and are compatible with each other. A very interesting recent treatment of verification is given by Roache (1999), who also examines the definitions of verification and validation and their origins; see also Oberkampf and Blottner(1998).

In practice, verification of all logical paths in a program is a gargantuan task, and the verification of modern software for nonlinear solid mechanics is seldom achieved even over the service life-time of the simulation application. A typical software package for nonlinear analysis of solids contains more than one million lines of code, and the number of execution paths is almost infinite. Many logical paths in a program have never been traversed even after a program has been in service for years, and obscure bugs tend to crop up regularly. While modern program source analysis tools can be utilized to help insure that all execution paths are tested, the general problem of completely verifiable code coverage is very difficult. Verification will not be considered further in this paper.

In validation, simulations are compared with physical experiments and tests to insure that the underlying algorithms accurately capture the essence of the associated physics. Today validation problem set for a solid mechanics program typically consists of the patch test, a set of benchmark problems, and perhaps 100 to 1000 other problems, most of them simple. These problems check the discretization and the algorithms rather than the performance of the program in the failure regimes of structures.

An example of benchmark problems is the MacNeal-Harder (1985) set for linear solid mechanics. In benchmarking, the results are compared against exact solutions, when available (as in the patch tests and beam problems) and against consensus results otherwise. Many of these benchmark problem sets are aimed more at verification than validation, since they primarily test whether the discretization is convergent and whether the algorithm is implemented correctly. Similarly, the problems called validation problems in most software manuals appear to be primarily aimed at verification: they check the various algorithms, but almost none of them are anywhere near the complexity of common engineering calculations.

Validation is closely related to computability. Implicit in validation is the assumption that problems in the domain of applicability of the computer program can be accurately computed if the program meets certain standards. However, the domain of problems to which nonlinear solid mechanics programs are applied includes many problems which cannot be computed with a high degree of precision, i.e. it is intrinsic in the response of the physical system that it is not computable. This aspect of nonlinear solid mechanics is often glossed over by many software developers and researchers. In fact, little is said about what makes a problem difficult and what level of accuracy to expect in a simulation. The intent of this paper is to examine the factors which contribute to the difficulty of solid mechanics problems, and which thus decrease computability. In addition, we will give some guidance as to what to expect in various types of problems and to the techniques that can be used in difficult problems.

The outline of this paper is as follows. Section 2 summarizes the governing equations for solid and structural mechanics and the major sources of uncertainties. Section 3 discusses the major barriers to computability: inadequacies in material laws, the roughness of the data and solutions and stability of a solution. Section 4 examines some methods with dealing with these difficulties. In Section 5 a measure of levels of difficulty is proposed and discussed; conclusions are given in Section 6.

2. GOVERNING EQUATIONS AND UNCERTAINTIES

Governing Equations and Data. To provide a foundation for our examination of computability, it is worthwhile to review the governing equations for a solid. The governing equations are (see Belytschko et al(2000) for different forms of these equations):

1. conservation of matter
2. conservation of linear and angular momentum (which becomes the equilibrium equation in static problems)
3. conservation of energy
4. strain-displacement equations
5. constitutive equations.

The conservation equations are based on first principles and their validity is beyond question. For a smooth, stable solution, when the constitutive equations and resolution are adequate, modern simulation methods will achieve convergence in these equations. The strain-displacement equations are a mathematical definition and are not a source of difficulties, since they are exact.

For a system with only mechanical response, the constitutive equation provides the closure for the system equations. Constitutive equations are not based on basic physical or mathematical principles. Instead, they are empirical approximations to the observed behavior of materials. In many cases, there is a rich and powerful theory that governs the structure of these equations, as in elasticity, plasticity and viscoelasticity, but these theories only provide a framework for the construction of the constitutive equation. The exact form of the equations is a matter of fitting experimental results, and even then there are no guarantees that a material will behave according to these theories.

For thermomechanical systems, the energy equation also needs to be solved. This adds the need for additional closure equations: heat conduction law, phase change laws and laws for the conversion of mechanical to thermal energy. Except for the heat conduction law, these often also pose difficulties.

Uncertainties. A second source of difficulty in nonlinear analysis are uncertainties in the data. The major data for a solid mechanics simulation are:

1. material data
2. loading, boundary conditions, and initial conditions
3. geometry

Uncertainties can be classified in two groups:

1. stochastic uncertainties (or irreducible uncertainties), that as indicated in the second name, are an inherent part of nature and cannot be reduced by obtaining more data.
2. epistemic uncertainties (or reducible uncertainty) which is due to insufficient measurements.

For example, the anticipated seismic loads on a structure are subject to stochastic uncertainties: more measurements will not reduce the uncertainties. The material properties on a geologic site are subject to epistemic uncertainties: more measurements will reduce the uncertainty of material properties, though as a practical matter the reduction in uncertainty by more measurements is quite limited. See Oberkampf et al (2000) for further discussion and application of these concepts.

Uncertainties in the data become particularly troublesome when the response is neither smooth nor stable. In linear finite element analysis, variability in the load data is not particularly troublesome unless the variability is comparable to the magnitudes of the loads, since the variability in response is linearly dependent on the variability in the data. This benign effect is lost in nonlinear simulations: small changes in data can have large effects on response, so even small uncertainties can be troublesome.

Uncertainties in loads, initial conditions and boundary conditions play an important role in the computability of a problem. While most analysts are aware of the stochastic nature of loads, most do not seem to realize how difficult it is to predict loads for nonlinear processes, even in carefully designed precision experiments. Even in a testing machine, the measured load is not the true load on the specimen, particularly in difficult stages such as the points of material instability. Compliance of the components of the machine can lead to significant errors, and these deviations from assumed perfect support rigidity should arguably be modeled as part of the physical response. The specification of loads for nonlinear tests is generally very difficult. For example, in pendulum tests of concrete beams, where a pendulum is used to apply an impulsive load, the loading is very complex and accurate predictions cannot be made without including the response of the pendulum and support fixture in the simulation. In crush tests of box beams or soil consolidation tests, the friction between the fixture and the specimen can have significant effects.

The loads in actual operations of systems are of course even more difficult to ascertain. In a failure analysis of an aircraft, nuclear reactor or a computer, loads depend on how the system responds, which in turn may depend on details of the design and response. For example, the response of a protective structure or a submarine to extreme loads may depend on the behavior of the welds: if several welds fail, the response, and in turn the loading, will change drastically. Seismic loads are generally based on records available from other sites: the actual load on a particular structure may be far different in spectral content and time history.

Uncertainties in initial conditions, boundary conditions, and geometry are similarly problematic. Most solids and structures are subject to initial stress: in the sheet metal of a car, these are due to the forming process, in the soil surrounding a tunnel, they are due to tectonic forces. These stresses are almost impossible to measure, and ascertaining them through analysis would require knowledge of the history of the entity that is simply not achievable. Similarly, the geometry of a solid is sometimes crucial to its response; this is discussed further subsequently.

3. BARRIERS TO COMPUTABILITY

Material Models. The constitutive equations in the nonlinear regime are a major source of difficulties in computability. Nonlinear constitutive equations are often very simplified approximations to material behavior. For example, the elastic-plastic laws that are used to model metals, soils and concrete are usually crude approximations to the actual response of these materials for complex stress paths. No material closely agrees with the Mises and Drucker-Prager yield functions, and the approximations employed to model rate-dependence, changes in internal variables (such as backstress), and interactions between phases often deviate significantly from actual material behavior. This becomes particularly pronounced with cyclic loading and complex stress paths, which are crucial in simulating many problems in solid mechanics; for example the seismic response of structures. Substantial errors can arise from inadequacies of the material model. Furthermore, it is common practice to apply constitutive laws well beyond their range of applicability: one example is when geological materials are idealized using total-stress measures that omit the effect of entrained fluid pressure, instead of more relevant (and accurate) effective-stress relations. Elastic-plastic models are often driven far beyond the range of available data.

Even when we have a good material model, the determination of the material properties of a given site or system is problematic. In most cases, there are severe limits on how well we can know material properties. For example, in the plastic behavior of metals, significant deviations are found in the response for different batches of the same material: 10% variations in the yield stress of the same batch of material is not uncommon. An example of the effect of material properties is shown in Fig.1, where the response of a fully annealed hexagonal container (hexcan) is compared to a typical manufactured specimen. This study was prompted by the inability to predict the response of the manufactured hexcan. Fortunately, an experienced engineer pointed out that the specimen was made by cold-working, which increased the yield stress around the corners. Figure 1 shows experimental results for a hexcan that was fully annealed after being manufactured as compared to a manufactured hexcan. It can be seen that the response is dramatically different. Such variations in yield stress are found in many manufactured goods and are seldom accounted for.

The material property variations of materials such as concrete, rock and soil are far greater. Measurements on several core samples from a geologic site generally exhibit large variations, and an engineer must make daring assumptions in ascribing material properties to the volumes for which no data is available.

In many cases, the difficulties do not just involve a matter of sufficient test data. For example, in an in-service reactor, the properties of the materials will depend on long histories of temperatures to which they have been subjected, local corrosion, stress histories, and initial manufacturing methods. The stress and temperature histories will vary from point to point in the same component, so the material properties will differ. In a geologic site, local features, such as a groundwater flows or faults, can occur throughout the site and yet be undetectable. Even in an experimental specimen, the assumed uniformity of material properties is at best a convenient fiction. All of these factors will affect the response in the unstable regime.

Smoothness. The principal message of this paper is that the computability of a problem depends on an abstract measure, which we call the level of difficulty of the problem. The level of difficulty is inversely proportional to its computability. Two primary factors that contribute to the level of difficulty of a problem are:

1. the smoothness of the model
2. the stability of the model

When we speak of the smoothness of the model, we are also referring to the smoothness of the data and response, for generally any roughness in the response results from roughness in the data and roughness that is inherent in the model.

By smoothness we refer to the differentiability of the response for a model of reasonable resolution. We will characterize the smoothness by C^n, where n is the number of times the function is continuously differentiable. Examples of C^1, C^0 and C^{-1} functions are shown in Fig. 2. A C^1 function is once continuously differentiable, i.e. its derivative is everywhere continuous. We will apply this notion both to continuity in space and in time, and note that all functions with which we are concerned are continuously differentiable over subintervals in time and space, which is implicit in the notion of continuum mechanics. Thus a C^0 function in time will posses discontinuities in its first derivatives with respect to time, but between these discontinuities it will be continuously differentiable.

We have added the proviso of "reasonable resolution" in the first sentence of the previous paragraph because many problems can be made smooth and differentiable if the time and spatial discretization are sufficient. An example is the elastic-plastic response of metals: a typical model uses the stress-strain curve shown in Fig. 3a. As can be seen, its derivative is discontinuous at the transition from elastic to plastic behavior and at the transition from plastic to elastic behavior. It is always possible to replace the curve in Fig. 3a by the curve shown in Fig. 3b, in which the transition between elastic and plastic behavior is smooth, which in fact corresponds more closely with experimental results. But for the model in Fig. 3b to appear smooth in a computation, the time steps and spatial discretization must be fine enough so that the transition from elastic to plastic behavior occurs over several discrete increments in space and time. This is impossible for most computations today. It is worth noting that this is not a fundamental barrier to computability, as it can be overcome by more powerful computers.

The computability of a problem depends on the smoothness through two avenues

1. the smoother the problem, the easier it is to obtain a converged solution
2. the smoother the problem, the less sensitive its response to uncertainties in the data

Generally, the rougher the response, the more difficult it is to make accurate computations. And as indicated above, intertwined with the numerical difficulties of roughness are difficulties in computability. When the response is rough, its final outcome does not vary continuously with the data. Thus small changes in the data can result in large changes in solutions.

One of the most important sources of roughness in engineering structural calculations is contact-impact. When impact occurs, the velocity is discontinuous, so the velocity is C^{-1} in time. Thus the computation of the response of a system with contact-impact is inherently more difficult than for a smooth response.

In addition, impact is an on-off process. In a car model such as shown in Fig. 4, impact between certain components can either occur or not occur. Depending on the way the car strikes the obstacle, a specific component may either impact an adjacent component or slide past it. This will dramatically change the response in many cases. The outcome of such a simulation is very sensitive to the details of the computations and the initial conditions of the model. For example, Hallquist (1998) found that results of crash simulations were changed markedly when performed on a parallel computer on two different occasions. The discrepancy was traced to the a reordering of the summation of the nodal forces in different runs of the same program and model, which led to tiny changes in the nodal forces but large differences in the results due to the on-off properties of impact.

Many other simulations of structural systems involve contact-impact, since most systems are assembled from components that are only joined along surfaces; the surfaces may be free or connected by welds, bolts or rivets. In a nonlinear simulation of the system under extreme static loads or dynamic loads, the interaction of components along these common surfaces plays a large role and they are usually included in simulation models. The models of the interfaces along which sliding and contact-impact can take place, are called sliding interfaces. For example, in a drop test of a laptop computer or a power saw, the many mechanisms and joints are modeled by sliding interfaces. These interactions play an important role in the response and must be included for predictive simulation. Figure 5, a finite element model of a power saw for a drop test simulation, shows the many distinct components in a typical consumer product. All of the components interact through sliding interfaces.

In a droptest simulation, these interfaces open and close several times. The closing due to impact is often associated with high frequency waves in the model, which can lead to very noisy results. Another example of a model that makes extensive use of sliding interfaces is the suspension model shown in Fig. 6. This model was used for durability tests and subjected to impact loads on the tires to model a car hitting a pothole in the road, as shown in Fig. 7. When the impact force of a pothole is applied to a tire, the complex structural response leads to various impacts in the sliding interfaces between the suspension components. Thus the response can be quite rough, and the computability is affected adversely.

Interfaces between components introduce many other thorny issues: friction on the interfaces, sensitivity to initial conditions (components interact differently depending on how they were assembled, prepared or manufactured). In dynamic simulations, component interfaces result in noisy simulations due to the many impacts. The models of friction are still often based on theoretical constructs that were more appropriate in classical rigid body dynamics; usually Coulomb models with a single friction coefficient are used. When transitions occur between static and dynamic friction, these can introduce instabilities similar to the material instabilities described later. In friction models and many other areas, considerable fundamental mechanics research is needed before the mechanical models match the sophistication of the simulation programs.

Discontinuities in space are also important contributors to the difficulty of a computation. One of the major causes of discontinuities in space are cracks. These are associated with material instabilities that are discussed later. The formation and propagation of a crack forces the computer program to track an evolving discontinuity that is generally not aligned with the edges of the initial mesh. The analyst is then forced to choose between simple algorithms that force the crack along element edges, or to model the crack by removing elements. More sophisticated techniques are under development but have not been incorporated in software (Moes et al (2000) and Belytschko et al (2001)). Moreover, these discontinuities in space introduce discontinuous responses in time with a roughness comparable to contact-impact. Most software today model cracks by unzipping edges or killing element (element erosion) which are even more discontinuous than the actual process. The fidelity of these models for tracking actual crack growth has not been studied mathematically and is not well understood.

Stability. One of the major impediments to computability are instabilities in the response. We consider stability in the customary sense: a system is stable if a small perturbation of the system results in a small change in the response of the system. This is a definition that goes back to the nineteenth century, and was explored by pioneers such as Liapunov and Banach. Banach coined the alternate word "contractive" for stability, which is a stronger notion. Consider a discrete model governed by an evolution equation such as the equation of motion or discrete heat conduction equation. Let the solution with the initial conditions $\mathbf{d}_A(0)=\mathbf{d}_A^0$ be denoted by $\mathbf{d}_A(t)$. Now consider additional solutions for initial conditions $\mathbf{d}_B(0)=\mathbf{d}_B^0$, where $\mathbf{d}_B^0$ are small perturbation of $\mathbf{d}_A^0$. This means that $\mathbf{d}_B^0$ is close to $\mathbf{d}_A^0$ in some norm, (to be specific we will use the ℓ_2 vector norm):

$$\left\|\mathbf{d}_A^0-\mathbf{d}_B^0\right\|_{\ell_2}\le\varepsilon \qquad (1)$$

A solution is stable if for all initial conditions that satisfy (6.5.1), the solutions satisfy

$$\left\|\mathbf{d}_A(t)-\mathbf{d}_B(t)\right\|_{\ell_2}\le C\varepsilon \quad \forall t>0 \qquad (2)$$

According to this definition, the solution is stable if all solutions $\mathbf{d}_B(t)$ lie in a ball around the solution $\mathbf{d}_A(t)$ for any time whenever $\mathbf{d}_B^0$ is a small perturbation of $\mathbf{d}_A^0$. This definition is illustrated for a system with two dependent variables in Fig. 8. The left-hand side shows the behavior of a stable system. Here we have only shown two solutions resulting from perturbations of the initial data, since it is impossible to show an infinite number of solutions. The right hand side shows an unstable system. It suffices for a single solution starting in the ball about $\mathbf{d}_A^0$ to diverge from $\mathbf{d}_A(t)$ to indicate an unstable system. Similar definitions can be developed for the underlying partial differential equations.

While the mathematical definition of stability is expressed in terms of changes in initial data for purposes of tractability, the effects of changes in other data on the response of an unstable system are similar. For example, for a unstable system, small changes in loads or boundary conditions may lead to large changes in response.

The deleterious effects of lack of stability on computability can be grasped easily from the definition of stability. If a system is unstable, small changes in data will result in large changes in response. Obviously, the computability of the response decreases . The standard mathematical definitions of stability do not completely describe the range of events that can be associated with instability. In some cases, very large motions of the system follow instability, and the response is very sensitive to data. In other cases, phenomena that are characterized as instabilities by mathematicians are comparatively benign. For example, in the buckling of a shell immersed in a fluid is often accompanied by very small motions. Thus as the field of computational mechanics progresses, more powerful definitions of stability will need to be developed because the classical definitions do not suffice to resolve the underlying issues of engineering interest.

Instability poses a fundamental barrier to computability in many other fields. For example, in weather prediction, it is widely accepted that predictions of more than a month will never be possible regardless of the power of computers, because of the many instabilities in atmospheric phenomena. When local conditions are particularly unstable, even a 24 hour prediction may be impossible. Similarly in quantum mechanics, the prediction of behavior near the dissociation point of atoms is very difficult, since the system and governing equations are unstable.

In solid mechanics problems, instabilities can be classified as follows:

1. geometric (or system) instabilities
2. material instabilities

We will discuss each of these in the following.

Geometric Instabilities. Geometric instabilities can roughly be described as those associated with buckling and snapthrough (called limit points in mathematics). They are associated with the growth of perturbations of the system, hence the name system instabilities. For certain classes of solids and structures, they are associated with loss of positive definiteness of the system Jacobian matrix (tangent stiffness matrix); see Belytschko et al (2000, Chapter 6 for more details). Geometric instabilities make a system more difficult to simulate since special techniques are needed to track the system in equilibrium analyses and both equilibrium and dynamic analyses are more sensitive to data. Consequently system instabilities decrease computability. However, methods for analyzing system instabilities in equilibrium solutions have achieved a high degree of maturity, and powerful parametrization techniques such as the arc-length method have been developed, see e.g. Riks (1972) and Crisfield (1980). If the analyst takes the time to analyze the system carefully, a very good understanding of the behavior of the system can be obtained.

But certain types of geometric instabilities substantially impair computability. Two examples are imperfection sensitivity, as in the buckling of cylindrical shells, and systems with many equilibrium branches. In many cases, when analysts are confronted with the task of examining the dynamic performance of a structure under extreme transient loads, they simply run a single dynamic simulation. However an understanding of the behavior of geometrically unstable system and its imperfection sensitivity cannot be gained by a single simulation. As an example of the dramatic effects of imperfections on the unstable response of a structure, we cite the application studied by Kirkpatrick and Holmes(1989). The objective was to make a pretest prediction of the response of a cylindrical-shell like structure. Pretest predictions were substantially in error and a good correlation with experiment was only obtained after incorporating in the model the imperfections measured in the test specimen. Of course cylindrical shells are well-known for their imperfection sensitivity, which is most common in very symmetric structures. Most in engineering designs do not have such extreme symmetry, but some imperfection sensitivity is found in many engineering components and structures. Computability is diminished significantly by instabilities in imperfection sensitive structures.

Furthermore, in many engineering calculations, the object to be simulated is not available for measurement, so imperfections have to be assumed.

Geometric instabilities also detract substantially from the computability of a problem. For example, in the buckling of a long frame member, the buckle under high impact loads can first appear at various locations along the member. Although the buckling load will depend very little depending on the exact location of the buckle, in modern design some of the response of interest will not be readily computable. For example, if a model is used for the design of accelerometer placement, as in the accelerometers used to deploy airbags, whether the buckle occurs in front of the accelerometer or behind the accelerometer will make a large difference in the acceleration record. Thus while some aspects of the response, such as the maximum axial force sustained in a member, are not sensitive to the exact location of a buckle, other aspects that are of engineering interest are sensitive to these features of the response.

Material Instabilities. A material is considered unstable when a perturbation applied to an infinite slab of the material in a uniform state of stress grows without bound. The definition is in a sense highly idealized, but it is a remarkable fact that materials which are unstable by this definition will manifest behavior such as shear banding or cracking. When the criteria for material stability are not met in any point of a computer model, difficulties are often encountered in computations. Further discussion of this topic can be found in deBorst (2001).

Two types of unstable material models are most common:
1. Materials with negative moduli, which is often called strain-softening; a stress-strain curve with strain softening is shown in Fig. 9.
2. Elastic-plastic models with nonassociative flow laws, i.e. flow laws in which the plastic strain rate is not normal to the yield curve; a nonassociative flow law is contrasted with an associated flow law in Fig. 10.

Material instabilities are associated with several important phenomena in the behavior of solids. When the instability is tensile, the instability is associated with the formation of cracks. When the instability is in shear, the instability is associated with the formation of shear bands. Shear bands have been observed in metals, rock and soils, and are an important failure mode. Sometimes, they are a precursor to failure by fracture.

Thus material with unstable behavior is needed to model cracking and shear bands; material instabilities are also associated with phenomena such as the liquefaction of soils and crazing of polymers. The modeling of these phenomena is crucial in many failure simulations. Engineering analysis seldom dealt with these phenomena several decades ago. Fracture analysis consisted of a linear solution with a crack, and if the stress intensity factor exceeded the toughness of the material and hence predicted the possibility of crack growth, the design was deemed unacceptable and was altered. However, in the simulation of droptests, failure of parts, weapons lethality and vulnerability, and extreme events in critical systems such as a nuclear reactors, it of great interest to understand the behavior of a system after material instabilities occur. Simulations are used to judge the implications of failure, and failure is simulated so that better designs can be developed.

Such simulations require material models that replicate the failure modes that actually occur.

Material instabilities are generally more difficult to compute than systems instabilities. Not only is the response of models with unstable materials sensitive to initial data, as in an imperfection-sensitive system instability, but material instabilities are accompanied by localization: the deformation localizes in a narrow band about the region of material instability, with very high strains and strain gradients in the band. This often introduces difficulties due to lack of resolution. Unless the mesh is very fine in the region of material instability, the localization cannot be resolved properly. This has led to the notion of mesh sensitivity: solutions involving material instability are very sensitive to the alignment and size of the mesh. In many cases, mesh sensitivity is simply a matter of insufficient resolution.

It is of interest that in the 1970's, considerable controversy arose as to whether unstable materials should ever be used in computer models. The advocates of stable material models, see Sandler and Wright (1984), argued that computer models with unstable material models were essentially ill-posed according to classical definitions such as given by the Lax equivalence theorem; see Richtmeyer and Morton (1967). Sandler and Wright showed that addition of damping appears to improve the situation. Incidentally, the definition of well-posedness given by Lax is equivalent to a definition of stability given above; the assumption of "well-posedness" was necessary to prove convergence, but the term is quite misleading to engineers. Bazant and Belytschko(1985) showed another shortcoming of rate-independent strain softening models: the deformation localizes to a set of measure zero, so that there is no energy associated with the localization. In other words, a rate independent strain-softening material can not be used to model fracture since the resulting fracture energy vanishes

However, it has become clear that shear banding and fracture, which are crucial in failure analysis, cannot be simulated without unstable material models. Thus a computer model which can compute these phenomena must include unstable material models. (There was some merit to the argument against unstable material models, since the instabilities often arose in ad hoc models where it was not intended.)

In the past decade, many regularization procedures have been developed for unstable materials, among them rate-dependent material models, gradient models and nonlocal material models; see Belytschko, Liu and Moran (2000) for a brief, general description. Suffice it to say that except for viscoplastic models, very few of these have been verified by comparison with experimental results. Many researchers have developed regularization methods solely with the objective of remedying mesh sensitivity and have judged their methodology correct when mesh sensitivity is absent. In the absence of experimental verification this approach appears quite meaningless.

The difficulties of material instabilities manifest themselves in many ways in engineering simulations. Examples are failures of welds, rivets and other connectors, tearing of sheet metal, and shear band formation. The details of imperfections that lead to failures of these components are difficult to ascertain, particularly in a structure that has not yet been built. Furthermore, the failure processes in connectors such as welds occur on a much smaller scale than the scale of the structure. The effects of these subscale imperfections can be dramatic: when a line of welds fails during an automobile crash the response can change by 30% or more. Yet this difference can originate from a microscopic flaw in a weld, which is one of hundreds of welds in the structure. Developing a methodology that can practically deal with this is truly mind-boggling, for one cannot assume a range of flaws in all of the welds and hope to compute the result in a lifetime. Yet a solution that intelligently accounts for weld failure is clearly needed if the simulations are to be predictive.

It is clear that material instabilities pose severe challenges to computability. Today they are considered more intractable than geometric instabilities because of the accompanying localization effects. However, perhaps material instabilities are fundamentally no more difficult than system instabilities and perhaps it is only a lack of suitable algorithms that makes them seem more difficult today. In many ways they are similar to shocks in fluid mechanics, and computational fluid dynamics is today very proficient in dealing with shocks.

4. METHODS FOR DEALING WITH UNCERTAINTIES

As can be seen from the preceding, a key challenge in obtaining useful simulations is to account for uncertainties and their effect on the response when it is rough and unstable. Several methods have been studied for dealing with the uncertainties. We will briefly comment on three of those

1. sensitivity analysis
2. uncertainty analysis
3. anti-optimization

Sensitivity analysis has been widely used in design of systems governed by linear equations. Some efforts have also been devoted in applying it to the nonlinear regime, Tsay and Arora (1990) and Kulkarni and Noor (1995). However, the effectiveness of these methods for rough, unstable response are not clear since sensitivity methods rely on the smoothness of the variations of response with data.

Probabilistic methods are beyond the scope of this paper, but again they have been challenging even in linear problems. Direct application of these methods to nonlinear problems is not envisaged in the near future

An interesting approach is antioptimization, which has been used in the context of optimum design by Elishakoff et al(1994). The application of these concepts to large-scale nonlinear simulations would entail setting bands on the data and then maximizing and minimizing the response of interest.

5. LEVELS OF DIFFICULTY

We will next provide some guidance as to how these factors contribute to the level of difficulty of a simulation, which is the inverse of its computability. We then give some guidelines as to the levels of difficulty for various problems that are currently treated by engineers and researchers.

Table 1. Weights W_i for factors contributing to levels of difficulty $\mathcal{L}$

W_1	Smoothness	0 to 1.0
W_2	Geometric stability	0.5 (imperf. insensitive) to 1.0 (imperf. sensitive
W_3	Material stability	1.0 (one instability) to 2.0 (many instabilities)
W_4	Effectivenss of constitutive equations	0 to 2
W_5	Variability in data	0 to 1
W_6	Resolution requirements	0 to 1

Table 1 gives the major contributors to barriers in computability and assigns some rough ranges of numerical values to the difficulty associated with these factors. The range of values is quite large and considerable judgment is needed to assign a useful value to these factors, but rough estimates can be made quickly by assigning an upper and lower bound to each weight. Before giving some background on the factors, we would like to stress that the landscape of nonlinear simulations is very complex, and a table such as this is at best a very rough guideline.

The weight W_1 depends on the smoothness in data and response. If there are many impacts and many cracks form in a system, the response will be very rough and a factor of 1.0 is appropriate. For moderate lack of smoothness, such as roughness in loads and one or two impacts, a factor of 0.5 is suggested.

The weight W_2 depends on the type of instability. Obviously, any system where the instability is imperfection sensitive is less computable. Imperfection sensitivity varies over a wide range. The classical problem of a cylindrical shell is an extreme example, where 1.0 is an appropriate weight. However, even beams in impulsive loading exhibit marked sensitivity to imperfection, see Lindberg and Florence(1978).

Material stability has been separated from system stability because it is inherently more difficult and the state of the art is not as advanced with respect to these problems. Since unstable material behavior is also usually imperfection sensitive, it constitutes a major barrier to computability. They also engender difficulties in resolution, since they are associated with localization of the deformation which leads to multiscale behavior. The ratio of scales can be very large, and if there are many localized bands, as in a penetration calculation, a sound computation in multi-dimensions is truly formidable.

The difficulties due to material laws depend on how well the material is understood and the complexity of the load path. Some materials are very well understood over a remarkably large load range, e.g. rubber, which agrees very well with the Mooney-Rivlin hyperelastic laws. Cyclic loading poses a major challenge in most materials: even in rubber, hysteretic behavior has not been characterized well, whereas in elasto-plastic metals after several cycles the computability deteriorates almost completely. Thus a major factor in assigning a weight to the material is the number of cycles expected. For many materials, $0.1 \le W_4 \le 0.5$ for a single load excursion, but increases to 1.0 for several cycles.

How these factors should be combined to obtain a level of difficulty for a particular problem is also an open question. One approach is to use a root-mean square law, where factors W_i are combined to obtain a level of difficulty $\mathcal{L}$ by

$$\mathcal{L} = \left(1 + W_5 + \alpha W_6\right)\left(1 + \sum_{i=1}^{4} W_i^2\right)^{\frac{1}{2}} \tag{3}$$

where W_i are defined in Table 1 and α is a factor between 0 and 1, depending on the computational resources available (obviously, if the problem requires so little resolution that a computation can be done in a minute on the available computer, resolution poses no barrier). The formula is designed so that problems range from $\mathcal{L} = 1$, the simplest and most computable, to $\mathcal{L} = 10$, the most difficult and least computable.

There are many cases where equation (3) does not apply at all. It is easy to construct particular cases where a single difficulty suffices to make a reliable simulation almost impossible today. For example, predicting the acoustic signature of a submarine is a linear problem that only requires great resolution. However, the complexity of a submarine is so great and the number of uncertainties in its internal configuration so overwhelming that this problem is almost unsolvable today. Similar difficulties are found in the prediction of the seismic excitation of a nuclear power plant. Such systems are so complex that computations provide only rough estimates of the response in the linear range. True predictions have probably never been compared with results, and undoubtedly they would not reproduce the time histories of the response very well.

The selection of a root-mean square measure of the total level of difficulty is quite arbitrary. We have made some trial evaluations with this measure, some of which we will describe. It seems to place the difficult problems that cannot be solved routinely at about $\mathcal{L} = 3$. Problems at a level of difficulty of 1 are very simple, while those about 2 require sophisticated users but are amenable to today's software.

We next give some examples of how various simulations fit into this scheme. The discussion is summarized in Table 2, which gives the range of weights for these problems.

The simplest simulations involve homogeneous system with smooth, stable response with well-characterized materials and loads. Examples are a rubber engine mount subjected to axial load or the simulations of single loading of a steel pressure vessel which brings it to the plastic range but not near the failure load. Here the primary difficulties are shortcomings in the material model and data, and these materials are among the best characterized.

We have given levels of difficulty for two types of crash analysis: (1) simplified analysis, such as commonly performed today where material instabilities are neglected, the entire car is steel and the material parameters have been tuned by previous test-simulation comparisons; (2) more complex models consisting of composites and other new materials with untuned parameters and the likelihood of material instabilities. The inclusion of composites in the second model decreases computability significantly, since in composites crush failure is very complex and material instabilities are pervasive. In the simplest models crash simulations, material instabilities are often ignored, and geometric instabilities are not difficult because triggers for buckling are often designed into the automotive frame.

Table 2. Level of difficulties for various example simulations

Problem description	Rough-ness	Geo. stab.	Mat. stab	Mat. law	Varia-bility	Level of difficulty
	W_1	W_2	W_3	W_4	W_5	L
Engine mount	0	0.5	0	0.3	0	1.2
Car crash-simplified	1	0.7	0	0.5	0.3	2.2
Car crash-complex with composite materials	1	0.9	2	1.5	0.3	3.9
Seismic response concrete Column with r-bars	0.5	0	2	1.5	0.5	4.1
Seismic response: foundation with liquefaction	0	0	2	2	1	6
Blast response of cylindrical buried structure (foreign site)	0.5	1	2	1.5	2	9.7
Lab test of shear failure	0	0	1	0.5	0.3	2.0
Response of pressure vessel with some yielding	0	0	0	0.3	0.3	1.4

In the problems listed in Table 2 involving blast or seismic response, severe difficulties are engendered by the intractability of the material modeling near failure and large variabilities in the data. In addition, computability is impaired the unstable nature of the response.

6. CONCLUSION

We have proposed an abstract measure of the level of difficulty for the computability of mechanics problems. The major factors we have considered are

1. lack of smoothness
2. lack of stability
3. uncertainties and inadequacies in the material law, boundary conditions and initial conditions
4. resolution requirements

The difficulty measure $\mathcal{L}$ combines these factors so that the relative computability of various problems can be assessed. To a large extent, difficulties in the first three aspects of a problem can not be overcome by simply applying more computational resources. Instead, fundamental advances are needed in our understandings of the effects of these factors on computability and new techniques need to be developed to provide more intelligent simulations of nonlinear phenomena.

Even the difficulties posed by the need for resolution are daunting. We often get the impression of a widespread belief that Moore's law will overcome difficulties of insufficient resolution. (Moore's states that computer power increases by a factor of two every eighteen months; in variants of the law, the period varies from 12 to 24 months). Moore's law has enabled rapid increases in resolution over the past three decades when problems were primarily two-dimensional or linear. However, as engineers rely increasingly on three-dimensional nonlinear analyses, the benefits of Moore's law become less dramatic. For a three dimensional explicit calculation of a hyperbolic partial differential equation, the computations required for a simulation with a mesh with n elements along each side is of order n^4. Over the next decade, Moore's law predicts that computer power will increase by 100. Yet the consequent increase in the numbers of elements along an edge for a three dimensional mesh is only 3.2, so that we can not expect even an order of magnitude increase in resolution in the next decade due to Moore's law.

This is one of the first attempts that we know of in assessing the computability of nonlinear mechanics problems. Thus it is quite rough and should be taken with a grain of salt. It is certainly not definitive, and we expect that as computability is examined more carefully in the future, other factors will be identified and the impact of the factors described here will be understood more precisely.

ACKNOWLEDGMENTS

The support of the Army Office of Research and the Office, the Office of Naval Research and the National Science Foundation to Ted Belytschko are gratefully acknowledged.

REFERENCES

Bazant ZP and Belytschko T (1985) Wave propagation in strain-softening bar: exact solution. *J. Engineering Mechanics*, 10, 243-251.

Belytschko T, Liu WK and Moran B (2000) *Nonlinear Finite Elements for Continua and Structures*, Wiley, Chichester.

Crisfield MA (1980) A fast incremental/iterative solution procedure that handles snapthrough, *Computers and Structures*, **13**, 55-62.

DeBorst R (2001) Some recent issues in computational failure mechanics, *Int. J. for Numerical Methods in Engineering*, ??same issue as this paper??

Lindberg HE and Florence AL (1978) Dynamic pulse buckling-theory and experiment, Report DNA 6503H, SRI International, Palo Alto, CA

Hallquist J (1998) Personal communication.

Kirkpatrick SW and Holmes BS, "Axial Buckling of a Thin Cylindrical Shell: Experiments and Calculations", in *Computational Experiments* ed by WK Liu, et al, PVP Vol 178, American Soc. of Mechanical Engineers.

Kulkarni M and Noor AK (1995) Sensitivity analysis of the nonlinear dynamic viscoplastic response of 2-D structures with respect to material parameters, *Int. J. for Numerical Methods in Engineering,* 38 (2): 183-198.

MacNeal RH and Harder RL (1985) "A Proposed Standard Set of Problems to Test Finite element Accuracy," *Finite Elements in Analysis and Design*, 3-20.

Oberkampf WL, DeLand SM, Rutherford BM, Diegert, KV and Alvin KF (2000) Estimation of Total Uncertainty in Modeling and Simulation, SAND2000-0824, Sandia Nat. Laboratories, Albuquerque NM

Oberkampf WL and Blottner FG (1998) Issues in computational fluid dynamics code verification and validation, *AIAA Journal*, **36**: 687-695.

Richtmeyer RD and Morton KW (1967) *Difference methods for Initial Value Problems*, Interscience, New York.

Riks E (1972) The application of Newton's method to the problems of elastic stability, *Journal of Applied Mechanics*, **39**, 1060-1066.

Roache PJ (1998) *Verification and Validation in Computational Science and Engineering*, Hermosa, Albuquerque, NM.

Sandler IS and Wright JP (1984) Strain softening. In Theoretical foundation for large-scale computations of nonlinear material behavior, ed by S. Nemat Nasser, RJAsaro and GA Hegemier, Martinus Nijhoff, 285-296.

Tsay JJ and Arora JS (1990) Nonlinear structural design sensitivity analysis, part1, general theory, *Comp methods in appl mechanics and eng*, 81, 183-208.

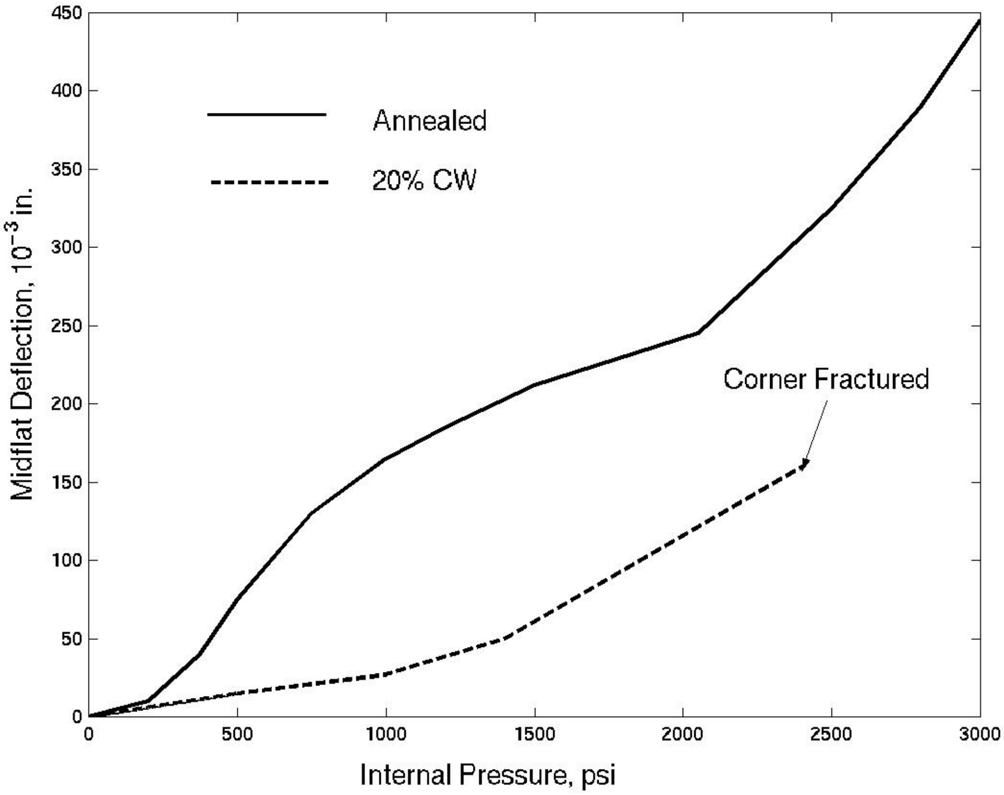


Figure 1. Experimental results for the midpoint deflection of a manufactured hexcan (20% coldworked) and a fully annealed hexcan showing the effect of change in material properties due to manufacturing process on the response

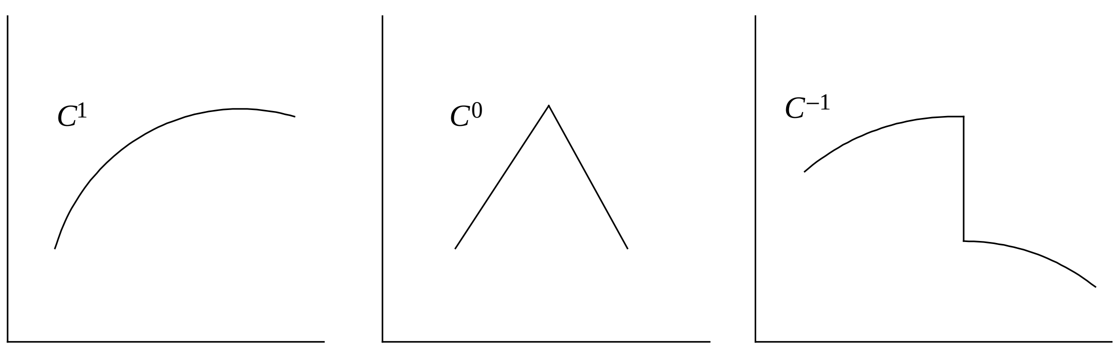
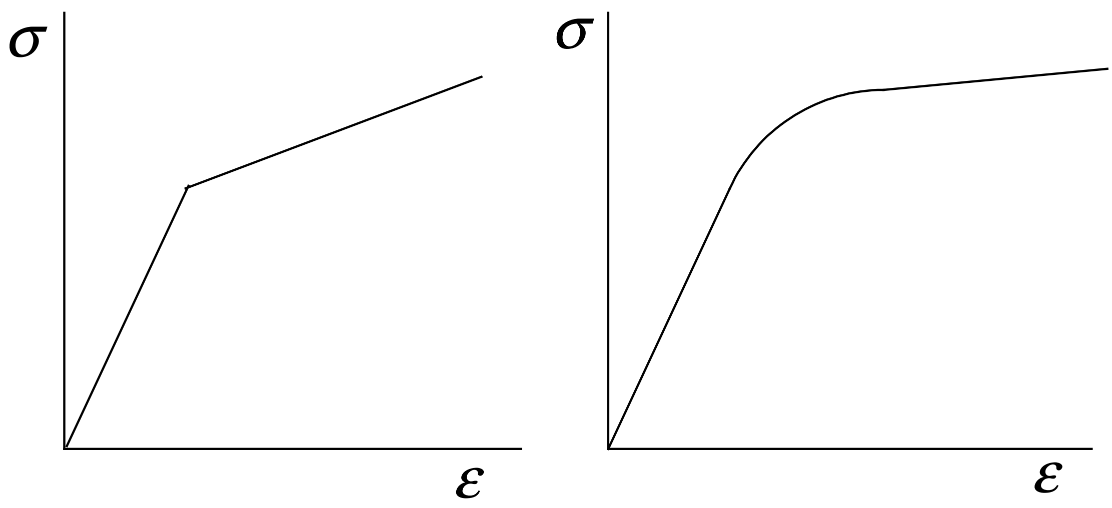


Figure 2. Examples of C^1, C^0 and C^{-1} functions; note that C^1 functions are distinguished from C^0 functions by the absence of kinks



(a) (b)

Figure 3 C^0 and C^1 representations of elastic-plastic response

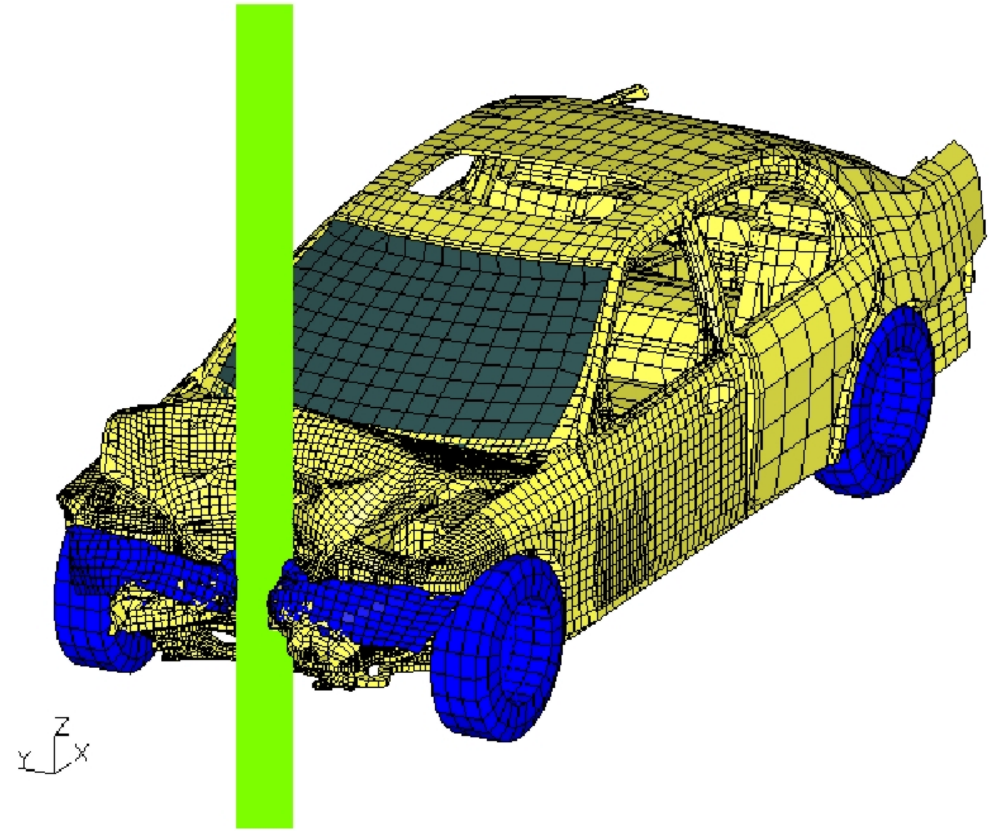


Figure 4. Finite element model of a car for crash simulation showing the many distinct components that can impact during an event.

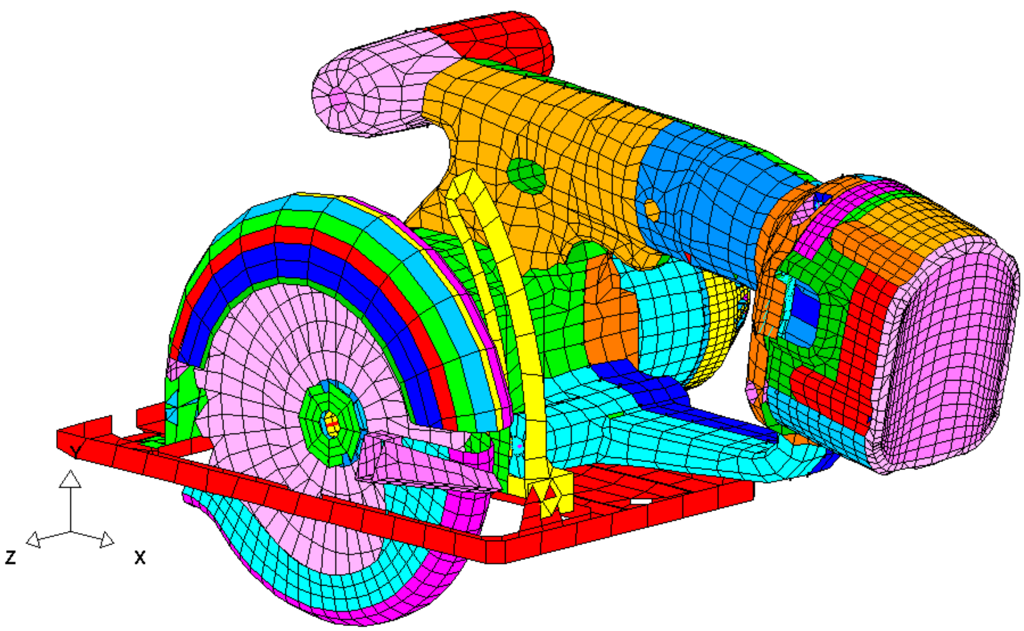


Figure 5. A model of a power saw used for droptest simulation; note the many distinct components which are separated by sliding interfaces.

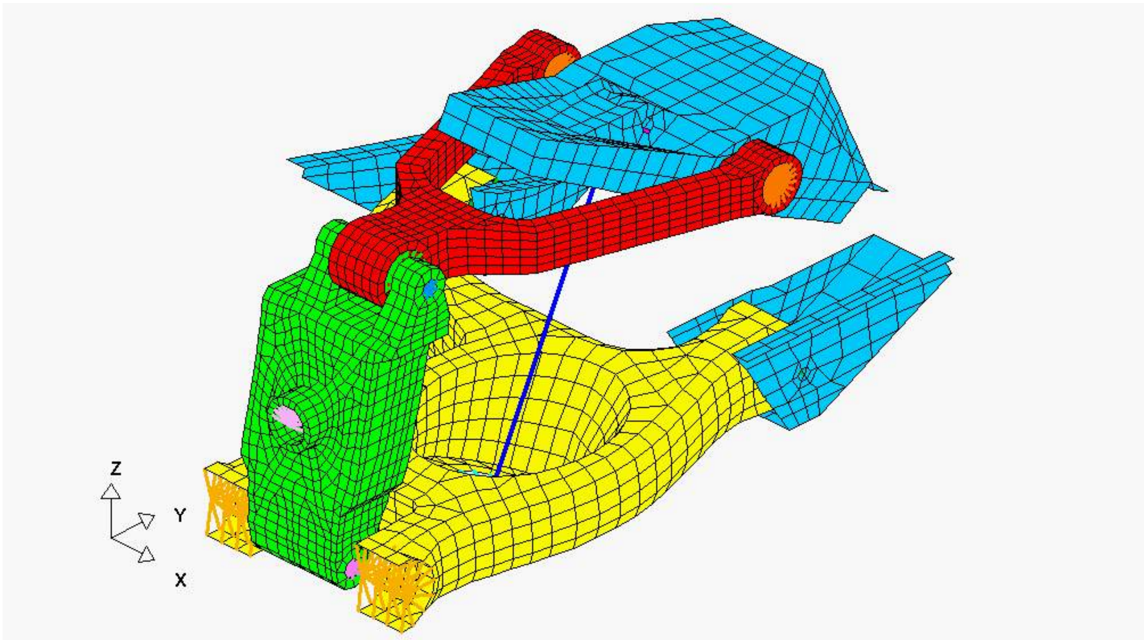


Figure 6. Model of a suspension for simulating effects of road roughness (courtesy of ETA Corp)

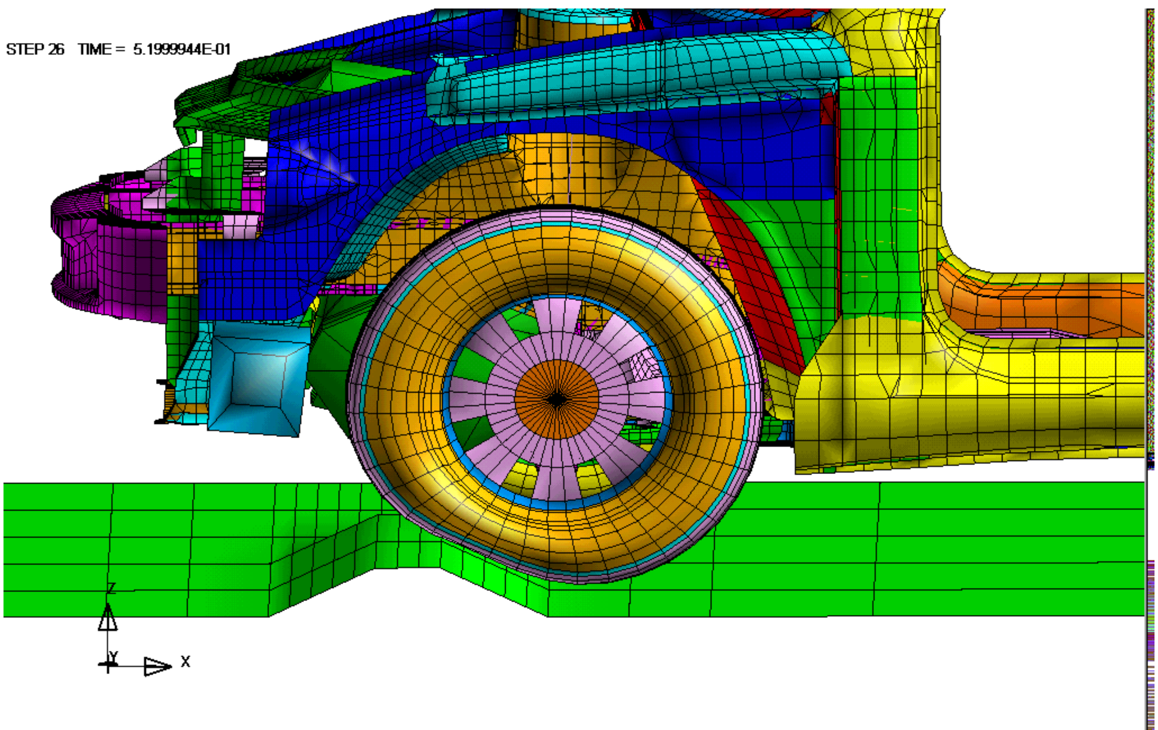


Figure 7. A car hitting a pothole; the suspension shown in Fig 6 is behind the wheel. (courtesy of ETA Corp.)

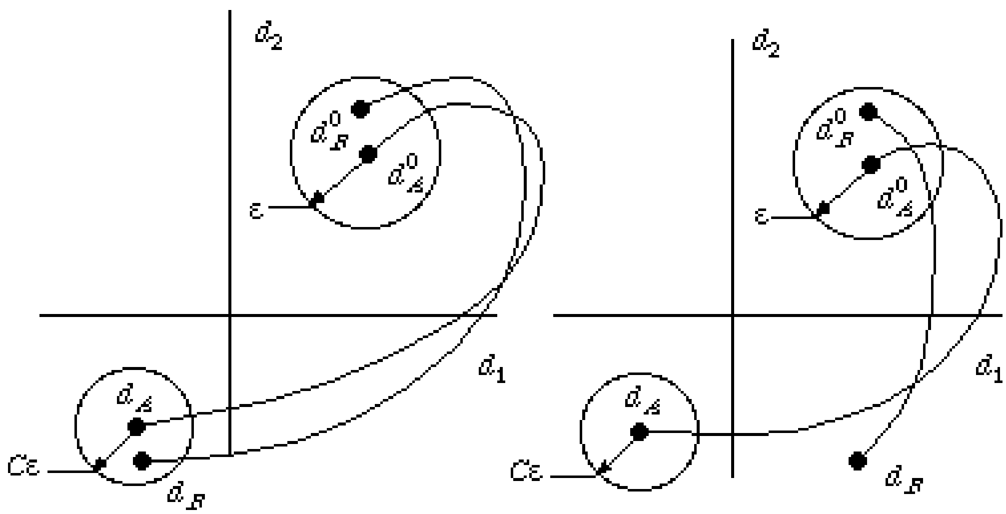


Figure 8. Trajectories for stable (on the left) and unstable (on the right) systems.

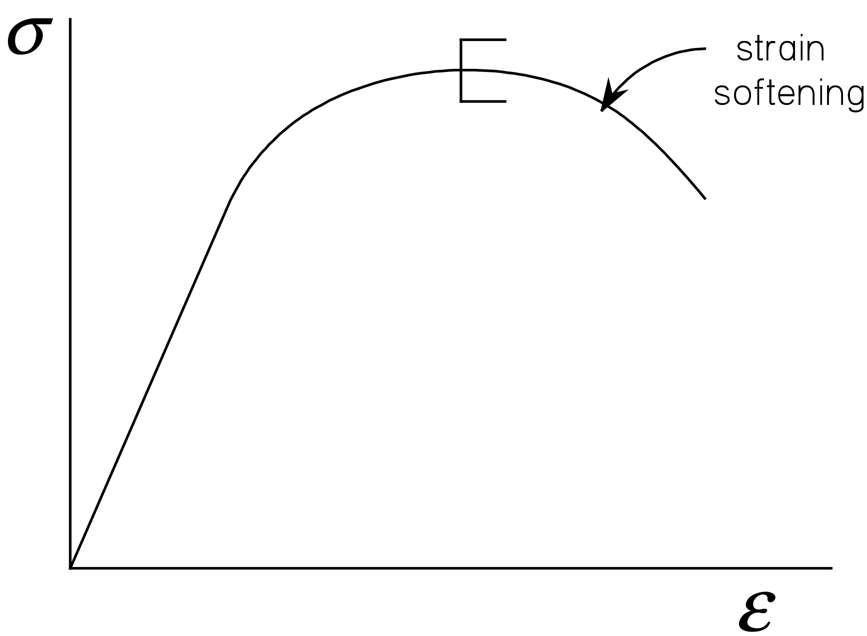


Figure 9. A uniaxial stress-strain in which material instability arises due to strain softening.

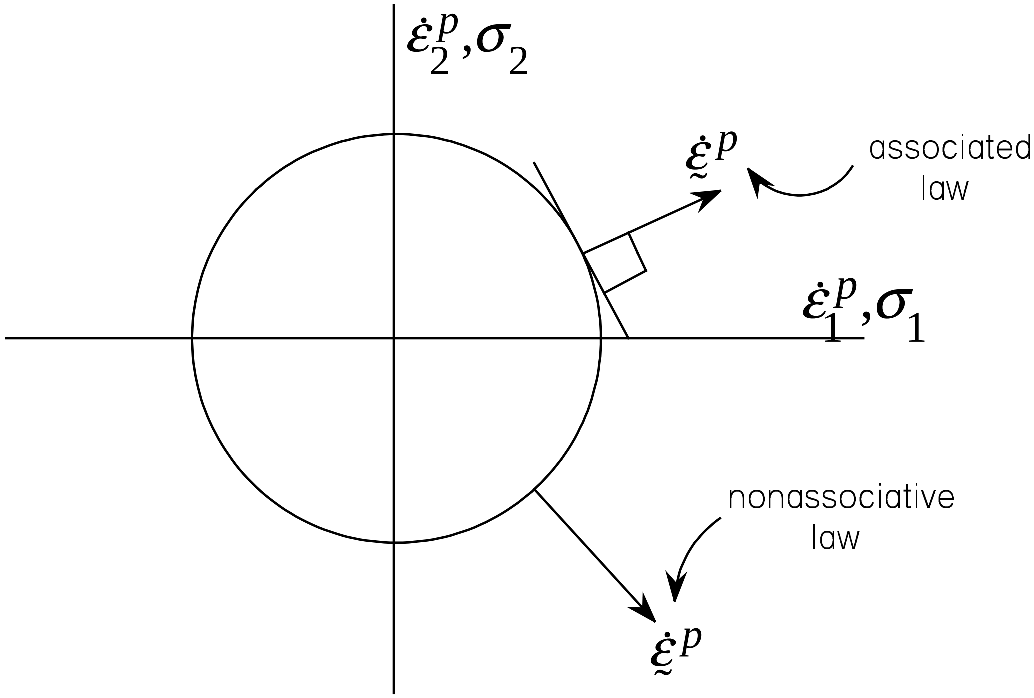


Figure 10. A yield surface with associated and nonassociated plastic flow laws; the latter can be unstable even with hardening.