Introduction and Overview of Hydrocodes for Impact Cratering Modeling

Robert H. C. Moir
Western University
Department of Applied Mathematics
robert@moir.net

Contents

1 Numerical Modeling of Impact Cratering

1.1 introduction

The purpose of this document is to provide a brief but reasonably clear and intuitive introduction to the physical processes and numerical procedures used in hydrocode modeling of impact cratering phenomena. Because explaining these processes and procedures in detail requires a good deal of understanding of physics and applied mathematics, the presentation here relies on intuitive derivations of equations in continuum mechanics and simplified descriptions of numerical procedures. There are many excellent rigorous developments of continuum mechanics available for those seeking greater detail and rigor, as well as detailed resources on hydrocode modeling, including Zukas (2004) and (Benson, 1992), which can be sought for further details.

1.1.1 what hydrocodes are

Hydrocodes are code packages that can be used for the numerical solution of mathematical models of high energy wave propagation phenomena. This includes the ability to accurately model shock waves. Hydrocodes also include facilitility to solve mathematical models of structural dynamics. This makes them useful for modeling a wide variety of physical phenomena, including fluid-structure interactions, vehicular collisions, explosions, shock hardening, and lunar and planetary impacts. Of course, the particular interest here is in the use of hydrocodes to model the various processes involved in the formation of impact craters from small fast-moving objects striking planets and moons.

1.1.2 what hydrocodes do

Hydrocodes compute approximate solutions to a dynamical material model, which can involve numerous interacting materials with very different physical properties. This model is formulated using the equations of continuum mechanics, the branch of classical mechanics that deals with the statics and dynamics of continuous materials, augmented with models of shock formation and dynamics. Thus, the user is responsible for constructing an appropriate model of some target phenomenon and for specifying equations that govern how the materials in the system behave under the range of conditions experienced by the simulated system.

1.1.3 what hydrocodes don't do

Hydrocodes do not, however, provide a black box that instructions can be fed to and out pops a simulation or explanation. These code packages require a great deal from the user in terms of understanding of the underlying physics, the associated mathematics and the numerical methods used to perform calculations. Hydrocodes also do not wear their limitations on their

sleeves. Consequently, a good deal of time, experimentation and experience is required to be able to use a hydrocode package to accurately model shock wave phenomena.

1.1.4 how hydrocodes work

The subject of this document, then, is to give a very brief outline of the physics and numerical methods needed to understand how hydrocode packages solve dynamical material equations. We begin in the next section with an intuitive treatment of the equations of continuum mechanics that are used to model macroscopic physical phenomena. This section is followed by a section explaining the numerical methods used to solve continuum mechanics equations, as well as some of the features of the numerical methods employed specifically by hydrocode packages. We conclude this document with a brief survey of the use of hydrocode packages to model various features of the impact cratering process.

It is hoped that someone reading this who is unfamiliar with the physics, numerics in general and hydrocodes in particular can take away a basic level of understanding of hydrocodes and their use to study impact cratering phenomena. Enough in terms of introduction, then, let us begin.

2 Mathematical Modeling of Continuous Media

In order to accurately model the dynamical behaviour of materials involved in impact cratering processes it is necessary to select a set of physical laws that govern the behaviour of the materials involved in the process and the forces and conditions that they are subject to. This amounts to the construction of a number of interdependent equations that determine how the physical quantities that characterize the relevant behaviour evolve over time. In this section we will consider the both the physical quantities that are relevant in impact cratering, which thus need to be included in models, as well as the types of equations that are used to model their dynamical behaviour.

Since geophysical processes generally involve the behaviour of large quantities of matter, a means of accurately modeling macroscopic quantities of materials is required. Since materials are composed of large numbers of atoms, the most detailed model of a geophysical system would require tracking the dynamical evolution of each of the atoms in a material. Not only would such a modeling approach be futile both in terms of specifying an initial state of the material and in terms of the complexity of the calculations, the solutions to such a model would be useless. What we want to know is how the materials *behave* in terms of macroscopic properties of the material that we can observe.

One way of dealing with macroscopic quantities of materials is to use a *statistical* approach, by working with average behaviour of the many atoms that we cannot track individually. This is the approach taken by statistical mechanics. But this is not the approach that is generally used because we have better methods that work by treating macroscopic materials as if they were *purely continuous* media. It turns out that it is possible to treat the behaviour of many macroscopic materials in terms of the evolution of a small number of measurable parameters that are subject to precise dynamical laws. Thus, we can use parameters like veloctiy, density, viscosity, temperature, *etc.*, to model the behaviour materials in geophysical processes.

Modeling macroscopic materials as continua is possible because of a physical phenomenon called *scale separation*, which is where the dynamical processes of a system at one scale are effectively *independent* of processes at other scales. It turns out that this is true of most fluids and solids and under are wide variety of conditions. Thus, we can avoid dealing with the description of materials in terms of their component atoms and treat them as if they were pure continua. In fact, in many cases changing the molecular composition entirely does not affect the macroscopic behaviour of a material. To be sure, the *values* of physical parameters that characterize a material, like the bulk modulus we will see below, will change with

a change of material, but the dynamical *laws* governing the behaviour will stay the same. It is for this reason that so many macroscopic materials *behave as if* they were pure continua.

Thus, we see that the most sensible strategy for modeling many geophysical processes involves treating materials as if they were continuous media. Constructing a valid model, then, *i.e.*, a model that correctly describes the behaviour of a real geophysical process, requires that we determine the physical laws that govern the dominant behaviour of the materials we wish to model. The branch of physics that is concerned with the construction, manipulation, analysis and solution of systems of such physical laws is *continum mechanics*. Thus, we now turn to consider the basic approach of continuum mechanics and the manner in which the behaviour of macroscopic materials is characterized.

2.1 continuum mechanics

The assumption that macroscopic materials can be treated as continuous is called the *continuum hypothesis*. We therefore see that the validity of the continuum hypothesis relies on the scale separation between macroscopic and microscopic behaviour of materials. The hypothesis involves more than simply this, however. For the hypothesis to be valid, it must be the case that macroscopic materials continue to behave like continua down to scales much smaller than the bulk medium. When this happens, it is valid to treat a macroscopic medium as being composed of tiny, "infinitesimal" *material parcels*. Since matter is not actually continuous, the smallest scale at which the medium behaves like a continuum must be much larger than the atomic scale. Thus, at this intermediate scale, small volumes of material are hardly infinitesimal in the mathematical sense, since they have a finite volume much greater than the atomic scale. Nevertheless, the validity of the continuum hypothesis allows the construction of dynamical equations that capture the dominant behaviour of a real material by treating material parcels *as if* they were infinitesimal.

One technical point that is important to recognize is that material parcels do not consist of the same set of molecules over time. This can be seen simply from the fact that for a fluid at rest the fluid (material) parcels do not move, but the molcules making up the fluid are in constant thermal motion and will travel from one "fluid parcel" to another, even for a stationary fluid. Thus, the velocity of a material parcel cannot be identified with the average velocity of a particular collection molecules. Rather, it is the average velocity of the molecules in a given tiny part of the material *at a given time*, which generally are a different set of molecules from one moment to the next. We regard, therefore, the material parcels as *tiny parts of the continuous material*. In this way, the physical quantities, such as velocity, density, pressure, temperature, viscosity, *etc.*, that we associate with a medium at a point come to represent the ideal values of physical quantities that we would *measure* at a given point in a medium at a given time.[1]

With an understanding of the difference between a material parcel and a tiny volume of molecules in a real material, we can consider how the equations of continuum mechanics are constructed and what they mean. The basic idea is that the equations governing the behaviour of the medium are the expressions of conservation laws, such as conservation of

[1] Whether or not the values provided by the model agree with what we would actually measure depends on whether or not the model is an accurate representation of the behaviour of the material. This is the question of the *validity* of the model, and will be discussed in section 3.4 below.

mass, energy or momentum, and Newton's laws of motion *applied to infinitesimal material parcels*. Since the quantities of interest (generally) vary continuously over a volume of material and over time, constructing the equations we need requires specifying how the properties of a material parcel change with respect to small changes in space and time. The need to consider variations in more than one dimension means that the kind of laws that result are *partial differential equations*, in the sense that the equations involve partial derivatives of quantities of interest.

Before we turn to consider the physical motivation for these laws, we must raise an important issue concerning the choice of *coordinates* in continuum modeling. First of all, we will use Cartesian coordinates $\mathbf{r} = (x, y, z)$, where $r = |\mathbf{r}|$ is the distance from a chosen origin, for the location of a point in space and the coordinate t for the time. There are then two basic ways to describe the motion of a material.

The first way of describing a material's motion is to track the way that the material is moving at a particular point in space. In this case, we imagine that we are sitting at a particular spot and continuously measuring the velocity of the material at that spot over time. In this case, we describe the velocity of the material in motion as

$$\mathbf{v}(\mathbf{r}, t) = \mathbf{v}(x, y, z, t).$$

This is called the *Eulerian* specification of the motion. It is also called the *spatial* approach, since we are working with a fixed space and tracking the motion of the material within that space. The coordinates $\mathbf{r}$ are called *Eulerian coordinates* of the material motion.

The second way of describing a material's motion is to track the way that a *particular material parcel* is moving over time. In this case, we imagine that we are travelling along with a material parcel as it moves and continuously measuring its position in space as we go along with it. In this case, we need a new coordinate $\mathbf{p} = (p_x, p_y, p_z)$ to pick out each parcel. This may be taken to be the centre of mass of a given parcel at some initial time t_0. The motion of the material is then described as

$$\mathbf{x}(\mathbf{p}, t) = \mathbf{x}(p_x, p_y, p_z, t),$$

where $\mathbf{x}$ is the position in space of material parcel $\mathbf{p}$ a time t_0. This is to say that at t_0 we divide up the material into material parcels and then track the motion of each of them over time. This is called the *Lagrangian* specification of the motion. It is also called the *material* approach, since we are tracking motion by tracking the location of individual parcels over time. The coordinates $\mathbf{p}$ are called the *Lagrangian coordinates* of the material motion.

For the purposes of this section we will work with the Eulerian specification, since it is simpler. But we will see the importance of the Lagrangian specification in section 3.3 when we consider numerical methods for solving the dynamical equations.

2.1.1 conservation laws

We will now restrict attention to fluid motion, as opposed to the general material we have been considering up to this point. This reduces generality, but simplifies the treatment, and is sufficiently detailed for our purposes. We will treat a fluid parcel as a small parallelepiped, the shape of which may vary as a result of applied forces. In the case of an *infinitesimal*

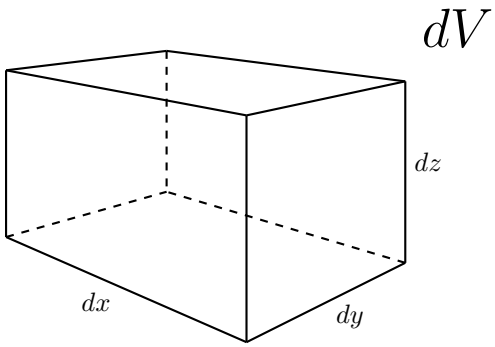


Figure 2.1: An infinitesimal fluid parcel

parcel, any physical quantity, such as velocity, density, *etc.*, will be considered to be *constant* over the parcel over an infinitesimal period of time dt. There can be changes between a given parcel and an adjacent parcel, but these changes must be infinitesimal.

The simplest equation arises from the formulation of conservation of mass for fluid parcels. Recall that we are using the Eulerian specification, so that the parcels are fixed in space and the fluid moves through them. Consider an infinitesimal fluid parcel dV located at $\mathbf{r} = (x, y, z)$ at time t, and let $\rho(x, y, z)$ be the density in the parcel and $\mathbf{v}(x, y, z, t) = (v_x, v_y, v_z)$ the velocity in the parcel at t. In the absence of sources or sinks of fluid, the only way that the amount of mass in the parcel can change is due to flow across the boundary of the parcel. We can break this down into contributions to the change of mass in V due flow in the x, y and z directions.

Consider the x direction. The y and z directions will be similar. Suppose that we can measure the velocity of the fluid on the left and right boundaries along the x-axis. If an equal amount of fluid leaves one boundary as enters the other, there will be no contribution to a change in mass of the parcel from flow in the x-direction. Thus, there is only a change in mass if the velocity of parcels at the two boundaries in the x direction are *different*. Let us now consider a step-by-step derivation of the change of mass in the parcel due to flow in the x-direction.

We will suppose that the near and far boundaries in the x-direction are the same size and that the velocities of fluid can be different at the two boundaries. Let us first, then, consider how to calculate the amount of material leaving the far boundary in the x-direction. Suppose that the velocity at that boundary is v_{x^+}, the x-component of the fluid velocity there. This means that v_{x^+} is the distance dr_{x^+} the fluid travels in the time interval dt divided by the time interval, *i.e.*,

$$v_{x^+} = \frac{dr_{x^+}}{dt}.$$

Note that if v_{x^+} is positive then fluid is travelling out of the parcel across the far boundary.

Now, the volume of fluid leaving the far boundary is the distance the fluid travels times the area of the surface it passes though. So, the volume leaving the far boundary is $dr_{x^+}dA_x$, where dA_x is the area of the boundary in the x-direction. Since ρ is the density (mass/volume) of the fluid, the *mass* of fluid leaving the far boundary is $dm_{x^+} = \rho dr_{x^+}dA_{x^+}$. By noting that

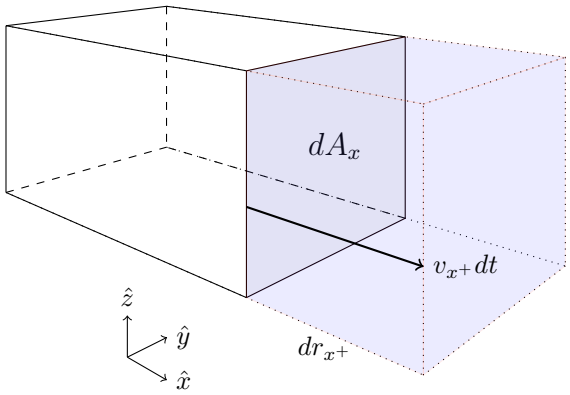


Figure 2.2: An infinitesimal flow of fluid out of an infinitesimal parcel in the x-direction. The volume of fluid flowing across the far x-boundary is $dr_{x^+}dA_x = v_{x^+}dA_x dt$, in which case the mass of fluid flowing across this bounrday is $\rho v_{x^+}dA_x dt$.

$dr_{x^+} = \frac{dr_{x^+}}{dt}dt = v_{x^+}dt$, we therefore obtain

$$dm_{x^+} = \rho\left(\frac{dr_{x^+}}{dt}dt\right)dA_x = \rho v_{x^+}dA_x dt.$$

Similarly, the amount of mass leaving across the near boundary over dt is $dm_{x^-} = -\rho v_{x^-}dA_x dt$, where the minus sign appears because a negative velocity will result in fluid leaving the parcel.

Thus, the total x-direction contribution to the *change* of mass dm of the parcel is

$$dm_x = -(dm_{x^+} + dm_{x^-}) = -(\rho v_{x^+} - \rho v_{x^+})dA_x dt = -d(\rho v_x)dA_x dt, \tag{2.1}$$

where $d(\rho v_x)$ represents the infinitesimal change in the quantity ρv_x across the parcel and the minus sign here appears becuase fluid *leaving* the parcel, which is what we just calculated, *decreases* the mass of the parcel. Therefore, the total contribution to the change in *density* $d\rho$ from x-direction flow is dm_x/dV, where $dV = dxdA_x$ is the volume of the parcel, which from eqaution (2.1) gives us

$$d\rho = \frac{d(\rho v_x)}{dx}dt,$$

which upon rearranging gives

$$\frac{d\rho}{dt} = -\frac{d(\rho v_x)}{dx}. \tag{2.2}$$

Now, the equation (2.2) that we just derived involves *ordinary* derivatives, *i.e.*, involves only measures of the *total* change in some dependent quantity under changes in some independent quantity. The dependent quantities here being density ρ and denstiy times velocity ρv_x, and the respective independent quantities being time t and length x. Because in the derivation of (2.2) we actually considered that *part* of the changes to mass and density contributed by changes occuring in the x-direction, holding any changes in the y- and z-directions fixed, the derivatives in (2.2) are actually *partial derivatives*, *i.e.*, measures of *partial* change in a quantity contributed by changes in one of several dependent quantities *while holding those other quantities fixed*. Thus, we must change the notation of (2.2) to reflect this, which yields

$$\frac{\partial\rho}{\partial t} = -\frac{\partial(\rho v_x)}{\partial x}.$$

Now, the argument we developed over the last few paragraphs for the contribution to density change in the x-direction is exactly similar for the y- and z-directions, so the overall change in density over the time interval dt is the sum of the contributions from the three directions, which gives us

$$\frac{\partial \rho}{\partial t} = -\left(\frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z} \right). \tag{2.3}$$

Using the special notation $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ to denote the *spatial differental operator*,[2] *i.e.*, the operator that when applied to a function measures how it changes with small changes in *spatial location*, we can write equation (2.3) in the more standard vector notation form

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}), \tag{2.4}$$

where the dot '$\cdot$' indicates the vector dot-product. This equation says that the change of density in time is equal to the negative of the *divergence* of the density flow, which we may think of intuitively as the degree to which density is *flowing out*, of a fluid parcel, hence diverging, at a given point. Equation (2.4) is called the *continuity equation*, since it expresses the condition that there are no discontinuous changes in mass, *i.e.*, that mass is conserved. An advantage of writing it in the vector form (2.4) is that, although we derived this equation in Cartesian coordinates, it is valid using any system of spatial coordinates.

This demonstrates how a partial differential equation arises out of the consideration of ininitesimal units of flow or change. We will not consider the construction of further equations in such detail now that the basic ideas have been made clear. We may simply note that equations for other conservation laws, like conservation of energy and conservation of momentum can be formulated in a similar way.

2.1.2 dynamical equations

Equations that express conservation laws like (2.4) are called *kinematic* because they specify how quantities vary in time prior to any consideration of the forces that are applied and the nature of a particular material. They are basic constraints that any continuous system must satisfy. In order to model how a material behaves over time, we need to model both the forces that are applied to the material and how the material responds to those forces. The basic relation used for this purpose is Newton's second law, which is used to express the conservation of momentum in dynamical phenomena. The equation expressing the law will be familiar from basic physics as

$$\mathbf{F} = m\mathbf{a},$$

which says that when a total force $\mathbf{F}$ is applied to a body, the acceleration $\mathbf{a}$ is in the direction of the force and is equal in measure to the force divided by the mass m of the body. So, larger force means larger acceleration and larger mass means smaller acceleration. By considering infinitesimal parcels of material we can apply this same principle to parcels, which results in dynamical equations for materials.

[2] The operator '∇' is typically called 'del' but is sometimes also called 'nabla'.

To obtain Newton's second law in a form suitable for continuum mechanics, we may recognize that it can be expressed in the form[3]

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt},$$

where $\mathbf{p} = m\mathbf{v}$ is the momentum, *i.e.*, the total force is equal to the rate of change of momentum. Let us apply this principle to an infinitesimal Eulerian fluid parcel. As we saw above, the infinitesimal mass of a fluid parcel can be expressed as

$$dm = \rho dV.$$

Now consider some body force $\mathbf{F}$, like gravity, applied to this parcel. Then Newton's second law implies that

$$\frac{\partial(dm\mathbf{v})}{\partial t} = \frac{\partial(\rho\mathbf{v})}{\partial t}dV = \mathbf{F}.$$

Then replacing the force $\mathbf{F}$ on the parcel with $\mathbf{f}dV$, where $\mathbf{f}$ is the *force per unit volume*, then we have, after cancelling the infinitesimal volume dV on both sides,

$$\frac{\partial(\rho\mathbf{v})}{\partial t} = \mathbf{f}.$$

This is just Newton's second law "per unit volume", *i.e.*, the change in momentum per unit volume is the force per unit volume.

This equation is too simple as it is, however, because the total change in momentum changes not only because of forces applied to the fluid parcel but also because of flow of material into and out of the fluid parcel. From our construction of the continuity equation, we know that the change in *mass* density ρ due to material flow is $-\nabla \cdot (\rho\mathbf{v})$. It follows, then, that the change in *momentum* density, where momentum is mass times velocity, due to material flow is $-\nabla \cdot (\rho\mathbf{v})\mathbf{v}$, which we can write as $-\mathbf{v}\nabla \cdot (\rho\mathbf{v})$ to make it clear that the velocity *vector* is multiplied by the *scalar* divergence of the momentum density. Adding this contribution to the change in momentum density gives us

$$\frac{\partial(\rho\mathbf{v})}{\partial t} = -\mathbf{v}\nabla \cdot (\rho\mathbf{v}) + \mathbf{f}. \tag{2.5}$$

We still have one other basic contribution to consider. We have included the body forces applied to the parcel, but we have not considered the *contact forces* applied by nearby parcels.

The basic contact force is the *pressure* $p(x, y, z, t)$ exerted by nearby parcels.[4] An acceleration due to pressure in the x direction will occur if there is an unbalanced pressure force in the x direction, *i.e.*, if the pressure *changes* from one side of the parcel to the other. From similar considerations as we used above, we find that the infinitesimal change in pressure force across the parcel in the x direction is $(dp)_x dA_x$, where $(dp)_x$ is the infinitesimal

[3]The meaning of the symbol $\mathbf{p}$, which here represents momenum, is not to be confused with the meaning intended above, where it was used within the Lagrangian specification to refer to the coordinates of a fluid parcel at some reference or initial time t_0.

[4]We do not need to explicitly represent the direction of the isotropic pressure force since it is always directed inward on the boundary of a volume. Thus, we do not represent simple isotropic pressure as a vector field.

change in pressure across the parcel in the x direction. When we divide this by the volume $dV = dxdA_x$ of the parcel, we see that our contribution to the force density is

$$-\frac{(dp)_x}{dx},$$

where the minus sign appears because an increase in pressure in the positive x direction results in a force in the opposite direction. Recognizing that restriction to changes in the x direction makes this a partial derivative, we obtain

$$-\frac{\partial p}{\partial x}.$$

Adding the force density contributions from the three directions, taking account of the fact that the three forces are directed along the coordinate axes, the force density contribution from pressure is

$$-\nabla p = -\left(\frac{\partial p}{\partial x}, \frac{\partial p}{\partial x}, \frac{\partial p}{\partial x}\right).$$

This indicates that the force density due to pressure is the negative of the *gradient* of the pressure, which is the spatial rate of change of the pressure, *i.e.*, a directional measure of how the pressure changes as we change position at a particular time. Addiing this to the right hand side of equation (2.5) and rearranging gives the equation of motion

$$\frac{\partial(\rho\mathbf{v})}{\partial t} + \mathbf{v}\nabla\cdot(\rho\mathbf{v}) = -\nabla p + \mathbf{f}. \tag{2.6}$$

This is essentially Euler's equation for fluid flow.

There is one other important force contribution in the case of fluids—viscosity. Viscosity results from the friction between nearby parcels. Thinking in terms of parallelepiped parcels as we have been, in any given direction (x, y or z) there are two independent frictional forces between surfaces since a surface of one parcel and slide over another in two different directions (see figure 2.3 for a depiction of the shear stress forces that result in viscous friction). Thus, for adjacent surfaces in the x direction, we can break down the frictional force into a y-component and a z-component of the frictional force. The amount of friction depends on how fast parcel surfaces are moving relative to one another. Thus, for surfaces in the x direction, the frictional forces are proportional to how fast the the fluid velocity $\mathbf{v}$ changes in the y and z directions, *i.e.*, on

$$\frac{\partial\mathbf{v}}{\partial y} \quad \text{and} \quad \frac{\partial\mathbf{v}}{\partial z}.$$

A simple model of the resulting viscous stresses, is thus

$$\tau_{xy} = \mu\frac{\partial\mathbf{v}}{\partial y}\hat{y} \quad \text{and} \quad \tau_{xz} = \mu\frac{\partial\mathbf{v}}{\partial z}\hat{z},$$

where μ is a constant parameter called the *dynamical viscosity* that measures the amount of friction between surfaces and $\hat{y}$ and $\hat{z}$ are unit vectors in the y and z directions indicating the direction of the forces. Since the net force on a fluid parcel is exerted by the *change* in stress across a fluid parcel, the total force per unit volume in the x direction from stress applied to the x-surface is

$$-\frac{\partial p}{\partial x} + \mu\frac{\partial^2\mathbf{v}}{\partial y^2} + +\mu\frac{\partial^2\mathbf{v}}{\partial z^2}.$$

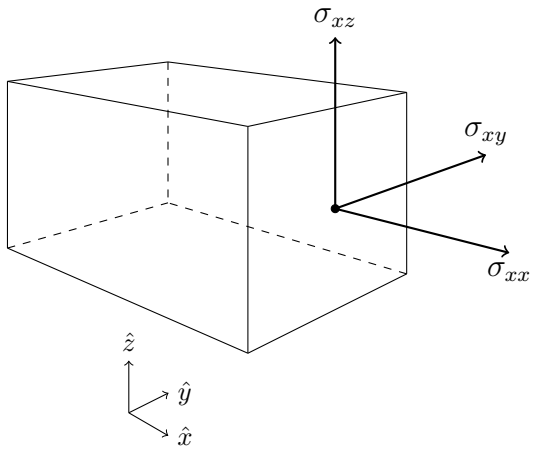


Figure 2.3: A depiction of the three principal stresses on the surface of an infinitesimal fluid parcel in the the x-direction. The stress σ_{xx} that points in the x-direction is a compression/expansion force. The other two stresses, σ_{xy} and σ_{xz}, which point in the y- and z-directions but are applied to the x-surface, are shear forces, which can result in viscous friction between parcels, or strain deformation or rotation of the fluid parcel. The three principal stresses in the y- and z-directions are similar.

Fluids for which this simple viscosity model is valid are called *Newtonian fluids*.

Rather than consider viscosity in more detail, it will suit our purposes better to consider a more general case for materials that need not be fluids. We consider the case where for any given surface of an infinitesimal material parcel there can be forces in three independent directions: shear forces that are parallel to the surface; and tension or compression forces that are perpendicular to the surface. Thus, viscosity is an example of a shear force and pressure is an example of a compression force. This gives rise to a *stress tensor* $\boldsymbol{\sigma}$, which has components σ_{ij} corresponding to the stress in the direction $\hat{j}$ on a surface in the direction $\hat{i}$, where $i, j \in \{x, y, z\}$ (see figure 2.3 for a depiction of the stresses σ_{xj} acting on the x-surface of a fluid parcel; the stresses σ_{xyj} and σ_{zj} on the y- and z-surfaces, respectively are defined similarly). This allows us to regard the stress tensor as a 3×3 matrix, where the diagonal terms are tension-compression stresses and the off-diagonal terms are shear stresses.

Now, folllowing a similar argument to that used to derive the continuity equation, we find that the contribution of a given component of the stress tensor to the change in momentum density is equal to the *rate of change of stress* in the direction the force acts. Consider the x direction. Then, the tension-compression force density is $\frac{\partial \sigma_{xx}}{\partial x}$, much as before except that the diagonal component σ_{xx} of the stress tensor takes the place of the pressure. The other components to stress in the x direction arise from shear forces on surfaces in the y and z directions. Similarly, the resulting force densities are $\frac{\partial \sigma_{yx}}{\partial y}$ and $\frac{\partial \sigma_{zx}}{\partial z}$, since accelerations will result if the shear forces are *different* on opposite sides of the material parcel. Thus, the total x direction contribution to the force density becomes

$$\nabla \cdot \boldsymbol{\sigma}_x = \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{yx}}{\partial y} + \frac{\partial \sigma_{zx}}{\partial z},$$

where $\boldsymbol{\sigma}_x = (\sigma_{xx}, \sigma_{xy}, \sigma_{xz})^{\mathrm{T}}$ is the first column of the stress tensor matrix. Therefore, the Euler equation of motion gets replaced by the Navier-Stokes equation

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + \mathbf{v} \nabla \cdot (\rho \mathbf{v}) = \nabla \cdot \boldsymbol{\sigma} + \mathbf{f}, \tag{2.7}$$

where the "tensor divergence-gradient" $\boldsymbol{\nabla} \cdot \boldsymbol{\sigma} = (\nabla \cdot \boldsymbol{\sigma}_x, \nabla \cdot \boldsymbol{\sigma}_y, \nabla \cdot \boldsymbol{\sigma}_z)$ (indicated by the bold del operator) is convenient notation.

Now, if we shift to a Lagrangian material parcel that can be stretched or compressed by stresses, then we can differentiate between three basic forms of stress, *viz.*, those that change volume, those that change shape, and those that rotate. To do this we first separate the stress tensor into two parts,

$$\boldsymbol{\sigma} = \boldsymbol{s} + p\boldsymbol{I},$$

where p is an isotropic, scalar pressure, responsible for changes in volume, with $\boldsymbol{I}$ the identity matrix, and $\boldsymbol{s}$ is the *deviatoric stress tensor*, which represents forces that change the shape or rotate material parcels (but preserve volume). Now, the deviatoric stress tensor can also be separated into two parts,

$$\boldsymbol{s} = \boldsymbol{s}_s + \boldsymbol{s}_a,$$

one symmetric and the other antisymmetric.

Let us consider the symmetric case, the antisymmetric case should then be clear. For the symmetric part of the deviatoric stress, the components satisfy the condition

$$s_{ij} = s_{ji},$$

which is just what it means to be symmetric. What this means geometrically is illustrated in figure **??**, which shows the meaning of the condition $s_{xz} = s_{zx}$ where both are positive. This corresponds to a strain that deforms the fluid particle by collapsing the x-face and z-face into one another. If both stresses are negative, the vectors shown in figure **??** point in the opposite direction and the strain deforms the particle by pulling the faces apart. Generalizing this to s_{xy} and s_{yz} as well, this shows that the symmetric part of the deviatoric stress tensor specifies the forces that cause deformation of fluid parcels. And this only requires the specification of three principal stresses, *e.g.*, s_{xy}, s_{xz}, and s_{yz}. Once this symmetric case is understood, the antisymmetric case, where $s_{ij} = -s_{ji}$, is easily understood to specify stresses that rotate fluid particles, which is necessary to model vorticity (see Tritton, 1988, 81, for a definition and discussion of vorticity).

Since vorticity is significant only in fluid motion and we are primarly interested in continuous and fractured solids, we will assume that the deviatoric stress tensor $\boldsymbol{s}$ is symmetric. In this case, the Navier-Stokes equations reduce to the form[5]

$$\frac{\partial(\rho\mathbf{v})}{\partial t} + \mathbf{v}\nabla \cdot (\rho\mathbf{v}) = -\nabla p + \boldsymbol{\nabla} \cdot \boldsymbol{s} + \mathbf{f}, \tag{2.8}$$

which clarifies the manner in which the Navier-Stokes equations are a generalization of the Euler equations.

Although this equation might appear intimidating, it is important to recognize the *physical meaning* that the equation clearly expresses given our understanding of the physical origin of each term. Each term in the equation represents either a rate of change in momentum density or a force density. The two terms on the left hand side represent, respectively, the

[5]Note that the vector divergence $\boldsymbol{\nabla} \cdot p\boldsymbol{I}$ reduces to the gradient of p since, *e.g.*, the x component satisfies $\nabla \cdot (p\boldsymbol{I})_x = \frac{\partial p}{\partial x}$, where $(p\boldsymbol{I})_x$ is the first column of the diagonal matrix $p\boldsymbol{I} = \mathrm{diag}(p, p, p)$.

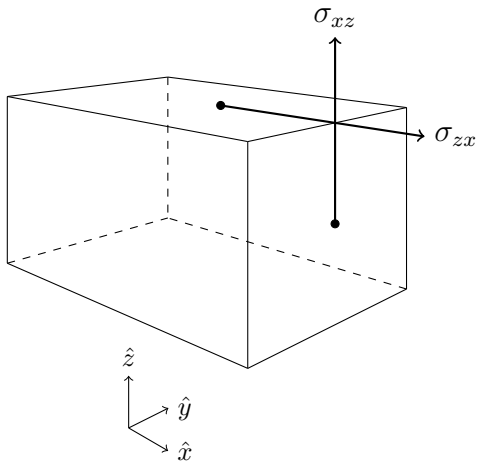


Figure 2.4: A depiction of the condition that the deviatoric stress tensor by symmetric for the s_{xz} and s_{zx} terms. The condition implies that the stresses act contrary to one another, leading to deformation of the fluid parcel. The case where s_{xz} is positive is shown, leading to collapse of the x- and z-surfaces into one another. For negative s_{xz} the vectors point in opposite directions, resulting in opening of the surfaces. In the *antisymmetric* case, the stresses act in conjunction with each other, leading to rotation of the fluid particle about the y-axis (counterclockwise for positive s_{xz}, clockwise for negative s_{xz}).

overall rate of change in momentum density and the change in momentum density specifically as a result of fluid motion. The three terms on the right hand side then represent, respectively, contact forces that change the volume, contact forces that change the shape, and external body forces. In this way, the Navier-Stokes equations are simply the expression of Newton's second law of motion for a continuous material.

Now that we have defined the deviatoric stress tensor, we can express another important conservation law, that of conservation of energy. In this case, the quantity that we are interested in tracking changes in is the *internal energy density* U of a material parcel. Since the internal energy can change for a number of reasons, including the flow of material in and out of the parcel, we obtain a term $\frac{\partial(\rho U)}{\partial t} + \mathbf{v} \cdot \nabla(\rho U)$ on the left hand side similar to that of (2.8), with the notational shift reflecting that we are tracking the motion of a *scalar* field (energy) rather than a *vector* field (velocity) in the case of (2.8).

The other sources of change in internal energy result from the heat generated by the stresses modeled by $\boldsymbol{\sigma}$. We can divide this into volume work done by the pressure and dissipation of heat due to viscous (frictional) stresses. Suppose that we require the parcel to have a fixed mass, then what we previously considered to be flow of material across the boundary equates to a change in volume of the parcel in order to keep the mass constant. Thus, the divergence $\nabla \cdot \mathbf{v}$ is a measure of the change in volume due to pressure forces (positive divergence means outward flow or decrease in volume). Since the volume work is given by $-pdV$, the contribution to change in internal energy from heat generated by pressure forces is $-p\nabla \cdot \mathbf{v}$.

The *strain* due to deformation of the material parcel is measured by a dimensionless tensor quantity $\boldsymbol{\varepsilon}$. The component ε_{ij} is the percentage change in the shape of a surface of a material parcel in comparison to an undeformed parcel. Given that in our model the deformation is caused by the deviatoric stress, changes in the component ε_{ij} of the state of deformation, are the result of the s_{ij} component of the deviatoric stress. The infinitesimal

amount of energy stored or released during an infinitesimal period of time is therefore the product of the stress and the time rate of change of the strain, *i.e.*, $s_{ij}\dot{\varepsilon}_{ij}$. This rate of change of the strain is called the *deviatoric strain rate*. Thus, we may write the conservation of energy equation as

$$\frac{\partial(\rho U)}{\partial t} + \mathbf{v}\nabla\cdot(\rho U) = -p\nabla\cdot\mathbf{v} + \boldsymbol{s}\cdot\boldsymbol{\varepsilon} + \mathbf{h}, \tag{2.9}$$

where $\mathbf{h}$ represents sources or sinks of heat and the "tensor dot product" (indicated by the bold dot) of $\boldsymbol{s}$ and $\boldsymbol{\varepsilon}$ is understood as the sum of the products of corresponding components of the two tensors, *i.e.*, the sum of all the terms $s_{ij}\varepsilon_{ij}$.

Once again, we end up with an intimidating looking equation. But, once again, the important thing to recognize is the physical meaning that the equation clearly expresses, given that we now understand the physical origin of each term. The two terms on the left hand side represent, respectively, the overall change in energy density and the change in energy density specifically as a result of fluid motion. The three terms on the right hand side then represent, respectively, the heat from the work done by the pressure to change the volume, the heat from the work done by the change in strain (deformation) effected by volume preserving stresses and external sources or sinks of heat.

Finally, before we continue, we may note that the same kind of derivatives appear on the left hand sides of equations (2.8) and (2.9) are the same. The common differential operator is often called the *material derivative* and is abbreviated $\frac{D}{Dt}$. In this notation, then, our three partial differential equations for the dynamics of the material are:[6]

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot(\rho\mathbf{v}); \tag{2.10}$$

$$\frac{D(\rho U)}{Dt} = -p\nabla\cdot\mathbf{v} + \boldsymbol{s}\cdot\boldsymbol{\varepsilon} + \mathbf{h}. \tag{2.11}$$

$$\frac{D(\rho\mathbf{v})}{Dt} = -\nabla p + \nabla\cdot\boldsymbol{s} + \mathbf{f}. \tag{2.12}$$

2.1.3 constitutive equations and the material model

The three equations (2.10), (2.12) and (2.11) constitute the specification of a (fairly) general theory of the behaviour of continuous materials in motion. Perhaps despite appearances, however, these equations on their own do not describe the behaviour of any physical system. Indeed, they are compatible with very different sorts of behaviour under nominally the same conditions, *ceteris paribus*. This is because these equations are too abstract—to get definite behaviour we need to supply information about the properties of the *kind* of material that we are modeling. Such information will tell us how the quantities of pressure, density, energy, stress and strain are related to each other. Since the relations between these quantities are, in general, different for different materials, the relations characterize the behavioural characteristics of a kind of material. The equations that specify these relations are called *constitutive equations*. These equations are often determined from empirical models based on measurements of the behaviour of real materials.

Since the unknown variables we have are ρ, $\mathbf{v}$, U, p, $\boldsymbol{s}$ and $\boldsymbol{\varepsilon}$, and we have only three equations in these six unknowns we require three additional equations to define definite be-

[6] This list of equations is not exhaustive. A more complete list is provided by (Zukas, 2004, 107, *ff.*).

haviour. These equations we need are of two basic types.

The first of these are thermodynamic equations that relates the pressure, density and either the internal energy or the temperature, which can be interconverted. These equations are called *equations of state* because they specify the state of a given quantity given the states of the other quantities related by the equation. The most common equation of state in continuum mechanics, so common that it is often not stated explicitly, is

$$\rho = \text{constant}.$$

This approximation simplifies the equations of motion of the last section considerably because $\frac{\partial \rho}{\partial t} = 0$ and ρ can be taken outside of a derivative operator. This together with the continuity equation implies that the material is *incompressible*, *i.e.*, has zero divergence. This is often a good approximation for fluids since they are known to change their density even with quite large changes in pressure (Tritton, 1988, 65). When this approximation is not valid, however, an equation of state of the form

$$\rho(p, U) = f_\rho(p, U),$$

for some function f_ρ of the pressure and internal energy, must be specified. We may call this equation the *density model*. In any case, we also require a second equation of state to relate the three quantities. Since the internal energy is usually computed directly in most impact models (Osinski and Pierazzo, 2012, 256), the second equation of state required is typically of the form

$$p(\rho, U) = f_p(\rho, U).$$

We may call this equation of state the *pressure model*.[7]

The equations of state determine thermodynamic properties of a material such as thermal expansion/compression, heat capacity, wave speeds, *etc.* Importantly, however, equations of state may also describe phase changes of a material where discontinuities in thermodynamic properties occur. Examples of phase changes are the familiar solid-liquid, liquid-gas and solid-gas transitions, but also include solid-solid transitions. Figure 2.5 shows a well-known equation of state for water, including the variety of distinct solid phases of ice, as a function of temperature, specific volume, and pressure. Since impacts produce melt rock as a result of the thermal energy released in a collision, considering phase changes is very important for accurate modeling. Phase transitions are also involved in more detailed effects like shock metamorphism, so careful choice and handling of phase transitions is an important part of constructing valid models of impact phenomena. An example of a pressure-temperature equation of state for several types of rock at the pressure scales and temperatures seen in impact phenomena is shown in figure 2.6.

The second type of equation we need to make the theory descriptive relates the deviatoric stress to the strain and strain rate, as well as the pressure and temperature or internal energy. Whereas the equations of state specify the response of the material to volume changes, these equations relating stress and strain specify the behaviour of the material to *shape changes*.

[7] Note that in general the density model and pressure model may be specified implicitly as $F_\rho(\rho, p, U) = 0$ and $F_p(p, \rho, U) = 0$, respectively. In such a case the density and pressure may not be single-valued as a function of the other variables. An instance where this situation occurs is in the phenomenon of *hysteresis*. In such a case the value of the density or pressure is *path-dependent*.

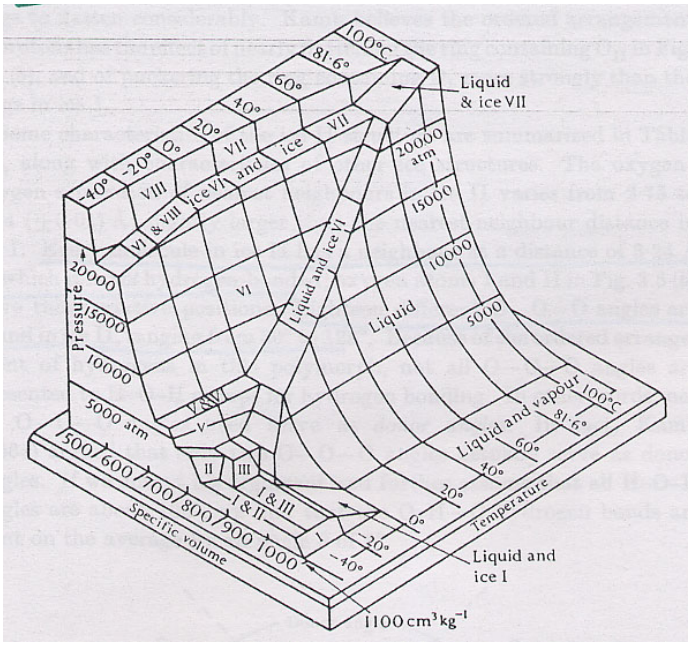


Figure 2.5: Graphical representation of the temperature-specific-volume-pressure equation of state for ice.

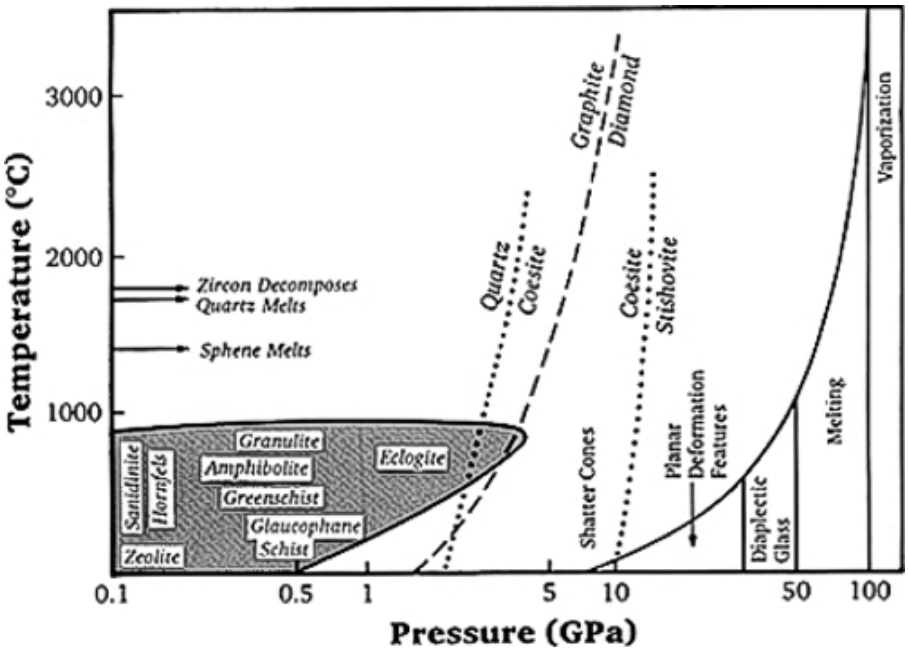


Figure 2.6: Graphical representation of the pressure-temperature equation of state for different types of rock under conditions of shock pressures and temperatures seen in an impact event.

These equations are called the *deviatoric stress-strain model*.[8] It is defined by an equation of the form

$$\boldsymbol{s} = \boldsymbol{f_s}(\boldsymbol{\varepsilon}, \dot{\boldsymbol{\varepsilon}}, p, U),$$

where $\boldsymbol{s}$ depends *component-wise* on $\boldsymbol{\varepsilon}$ and $\dot{\boldsymbol{\varepsilon}}$, *i.e.*, a given component of the stress tensor depends only on the corresponding components of the strain and strain rate tensors.

The simplest such equation is simply

$$\boldsymbol{s} = 0,$$

i.e., the material does not support any strain (deformation) in response to stress. This is called the *hydrodynamic approximation*, since the only contact force affecting the motion is the hydrostatic force from the pressure gradient. Although this is strictly false of any real material, it can be a valid approximation at very high pressures, where the contribution to the total force density from the pressure gradient overwhelms the contribution from the deviatoric stress. Accordingly, the hydrodynamic approximation is valid during the early stages of the impact process where the pressures are very large. For this reason the earliest numerical impact models ignored the strength of materials and adopted the hydrodynamic approximation, which is why they came to be called 'hydrocodes' (Osinski and Pierazzo, 2012, 265).

Just as viscosity must be taken into account for general fluid motion, the strength of materials must be taken into account for general solid motion. For solids, not only can there be viscous dissipation of heat, there can also be energy stored in the strain that a material can tolerate from applied stress. Up to a given point the strain is elastic, *i.e.*, the removal of the stress causes the energy to be released and the material returns to its prior unstrained state. This point is called the *elastic limit*. Beyond this limit the material becomes permanently deformed and the deformation is said to be *plastic* rather than elastic. This is a familiar phenomenon to any one who has stretched a string too much. The amount of strain that a material can support before a plastic defomation occurs is called the *yield strength*. There are two main ways that plastic deformation can occur depending on the type of material and its physical state. In *ductile flow*, the strain is uniformly distributed and the material permanently changes shape without breaking. In *brittle fracturing* or breaking, the strain is localized. In such a case a model of the resulting discontinuity, called a *failure model*, is required to represent the effects of breaking. Since the formation of shatter cones from impact shock waves are conical fractures in the country rock, modeling their formation requires this type of modeling.

There are two important general properties of the strength of rock to note here. The first is that the strength increases with increasing pressure as a result of the granular structure of rock. Increasing pressure presses the granules closer together, which makes it harder for granules to move over one another, thereby increasing the strength of the rock. The second is that the strength of rock decreases with increasing temperature. As the temperature

[8] Although both the equations of state and the deviatoric stress-strain model are constitutive equations, equations specifying the behaviour of specific classes of materials, the stress-strain model is sometimes referred to as the constitutive model or strength model of the material (*cf.*, Osinski and Pierazzo, 2012, 256). Note that Osinski and Pierazzo (2012) call the stress-strain equations the *deviatoric stress model*, I include the '-strain' to make clear that the model specifies how applied stress and the resulting strain in the material are related.

approaches the melting point of the rock, the ability of the rock to store energy as strain vanishes. As a result, the deviatoric stress term in the Navier-Stokes equations (2.12) reduces to a viscous term and the deviatoric contribution to the change in energy in equation (2.11) reduces to viscous dissipation. For more details on the strength of materials in the context of hydrocode modeling of impacts, including the effects of porosity, fractured materials, and sample size, see (Osinski and Pierazzo, 2012, 265-7) and the references included therein.

The equations of state together with the deviatoric stress model constitute the *material model*, which may be summarized as:

$$\rho = f_\rho(p, U); \tag{2.13}$$

$$p = f_p(\rho, U); \tag{2.14}$$

$$\boldsymbol{s} = \boldsymbol{f_s}(\boldsymbol{\varepsilon}, \dot{\boldsymbol{\varepsilon}}, p, U). \tag{2.15}$$

We will refer to the overall dynamical model specified by the six equations (2.10-2.15) as the *dynamical material model*.

2.1.4 boundary conditions

One final note before we turn to the consideration of numerical methods concerns the solutions of the dynamical equations together with the material model. Although these equations together describes definite physical behaviour, these equations are compatible with a wide variety of specific behaviour depending on the physical conditions that the material is subject to. The additional conditions are *boundary conditions*.

In the case of impact modeling, the boundaries that must be specified are the spatial boundaries of the different materials being modeled. In cases where we are dealing with interfaces between two large samples of different material, the boundary conditions simply specify the velocity and thermodynamic conditions that obtain at boundaries. In the case of porous or mixed materials, however, this becomes enormously more complicated. Strictly speaking, similar velocity and thermodynamic conditions must be specified *wherever a boundary between different materials exists*. For porous and mixed materials, these boundaries can exist throughout the bulk material and at very small scales. One type of approach determines the effective thermodynamic behaviour of the mixture of materials. Another, more accurate, approach resolves the material mixture in detail and the proper velocity and thermodynamic boundary conditions are applied. This more accurate approach can be pursued algebraically by multiscale methodsm which use special averaging and homogenization techniques to accurately determine the effective macroscopic behaviour from the known behaviour at smaller scales (see Pavliotis and Stuart, 2008). The more accurate approach to modeling mixed materials can also be pursued numerically by resolving the mixed material in detail and computing the effects of material boundaries. For more details on this case see (Osinski and Pierazzo, 2012, 264-5) and the references included therein.

There is also an important boundary in time since the state of the material system must be specified at some initial time t_0. This involves the specification of the initial values of the spatial configuration, velocity, density, pressure, internal energy, stress, strain and strain rate. When these initial conditions are applied together with appropriate spatial boundary condtions, the dynamical model of the material system will determine definite behaviour, at least locally to some range of the values of the physical quantities.

3 Numerical Modeling of Continua

The availability of cheap computing devices with fast processing and large amounts of fast memory makes the direct numerical solution of the equations of a model possible for increasingly many continuum mechanics models. The reliance on numerical solutions rather than algebraic solutions, however, brings with it new challenges and difficulties. Solving equations numerically introduces new sources of error that must be carefully accounted for to ensure that the results of a computation are reliable. In this section we will see that how concerns about reliability can be handled and how in some cases we can regard a numerical solution as an *exact solution to a nearby material model* using the techniques of *backward error analysis*.

3.1 convergence, stability and conditioning

The two basic conditions for any numerical algorithm to be reliable are *convergence*, which ensures that increasing the precision of the calculation converges on the exact solution, and *numerical stability*, which ensures that the error in the computed solution is "small" in a relevant sense. Before considering specific methods of solution of partial differential equations in the next section, we will here consider these two concepts in general terms.

There are two basic kinds of error in an numerical computation of the solution to a mathematical problem: forward error; and backward error. It will be helpful to explain these concepts in terms of a simple example. Consider the problem where we are to solve the equation

$$x^2 - 2 = 0 \tag{3.16}$$

using some numerical algorithm. In this case we know the exact solution is $x = \sqrt{2}$. Suppose that we use a computer to find an approximate solution and the result is $\hat{x} = 1.4142$. Then the *forward error* is the difference $x - \hat{x}$ between the exact solution x to equation (3.16) and the computed solution $\hat{x}$. In this case the forward error is approximately $1.36 \cdot 10^{-5}$, or 0.001% of the exact solution, a quite accurate result by many measures.

In this example we are able to calculate the forward error directly because we know the exact solution. In general, we do not know the exact solution and have no way of finding it. Indeed, for differential equtions, not being able to find the exact solution is the generic case. There is another kind of error, however, that can be computed or estimated and used to judge the quality of a numerical solution. This is called the *backward error*, which is the smallest change we can make to the data defining the problem so that the computed solution is the *exact solution to a modified problem*.

Let us explain this using the same example as before. If we substitute our computed

solution $\hat{x} = 1.4142$ to equation (3.16) back into the equation we find the result that

$$\hat{x}^2 - 2 = -3.83 \cdot 10^{-5}, \tag{3.17}$$

which can be rearranged to give

$$\hat{x}^2 - 1.99996164 = 0. \tag{3.18}$$

This shows that we have computed the square root of a number very close to 2. You will notice, though, that our computed approximate solution to (3.16) is the *exact solution* of the modified equation (3.18). Thus, we see that at the expense of a slight modification of the problem, we have found an exact solution—not an approximate one.

The quantity on the right hand side of equation (3.17) is called the *residual*. It is the amount by which the computed solution fails to satisfy the original problem (3.16). Since this is the smallest change to the data of the problem that makes the computed solution the exact solution of a modified problem, we see that the residual is in fact the backward error in this case. Thus, we see how the backward error can be computed by *substituting the computed solution back into the original problem*. It is not always possible to compute the backward error this way (see Corless and Fillion, 2013), but it is possible to do so for differential equation problems, which is the case we are concerned with.

There is a significant epistemic benefit from showing that the backward error is small when computing solutions to modeling problems, as we are for material models. By computing the backward error and showing that it is small, we can show that our computed solution is the exact solution to a material model that is only a small modification of our original model. But since error is introduced in measurement of parameters, construction of constitutive equations and in the construction of the theoretical equations of motion, the material model we are solving numerically is only an abstract approximation of the real material. Thus, provided that numerical error does not introduce non-physical perturbations into the problem, a small backward error shows that the numerical solution is the solution of *just as valid* a model as the original one. Stated simply, a small backward error means that if the original model was an accurate portrayal of the physics, then so is the numerical solution.

Due to the computational complexity of partial differential equation problems, particularly in three spatial dimensions, as is required in impact simulation, computing the backward error is not always feasible. When it is possible and it is small, however, it shows that we computed an exact solution to nearby material model. So if our original model is valid, so is the nearby model we solved exactly.

With the two types of numerical error made clear, we can now explain the concepts of convergence and stability in generic terms. In differential equation problems, as in most continuous mathematical problems, numerical solution involves replacing continuous quantities with a discrete mesh. With such a discretization, it can be coarse, using only few points, or fine, using many points. The parameter h is often used to represent the *largest* distance between mesh points. Thus, a mesh becomes increasingly fine the smaller h becomes. The basic idea of a convergent numerical method, then, is that as we make the discretization finer and finer the forward error goes to zero. In this way, the numerical solution *converges* on the exact solution as the mesh size goes to zero ($h \to 0$).

Convergence is a necessary condition for a good numerical method but not a sufficient one. It must also be the case that the numerical method produces a small error in actual computations. This is what the concept of numerical stability pertains to. An algorithm is *backward stable* if it produces a small backward error for any data. This is the basic requirement of stability for a numerical algorithm, and is a natural requirement for modeling problems, as was indicated above.

In modeling problems we often require not only a model that gives behaviour similar to the target system but also that it accurately tracks the target system's behaviour. In this case, we also require a small forward error. Determining this on the basis of a backward stable numerical method applied to a problem requires that the *problem* is well behaved in the following sense. For a small backward error to indicate also a small forward error, it must be the case that small perturbations or changes to the problem do not cause large changes in the solution. This behaviour of the problem is referred to as the *conditioning* of the problem.

If a small change to the problem or model leads to only small changes to the solution, then the problem or model is said to be *well-conditioned*. If, on the other hand, a small change to the problem or model leads to large changes in the solution, then the problem or model is said to be *ill-conditioned*. The conditioning of a problem or model is often measured by a constact called the *condition number* C. A small condition number indicates a well-conditioned problem, and a large condition number indicates an ill-conditioned problem. The condition number satisfies a relation of the form

$$\text{forward error} \lesssim C \cdot \text{backward error},$$

which shows that if the backward error is small and the condition number is small, then the forward error is also small. In this way, then, a backward stable algorithm applied to well-conditioned problem results in a small forward error.

A non-linearity in an equation is generally a source of ill-conditioning of a problem. Since the equations of continuum mechanics are generally non-linear, in the sense that they involve products of the solution and its derivatives, this is a significant concern for impact modeling. Just because a problem is ill-conditioned does not, however, mean that numerical solutions are useless. It does mean that care must be used when interpreting the numerical solution. Consideration of this in detail is beyond our scope, but you may see (Moir, 2013, 2010) or (Corless, 1994b) for more details on this issue in the context of chaotic ordinary differential equations.

3.2 numerical methods for partial differential equations

We now turn to the consideration of different methods of discretizing the differential equations of a dynamical material model. We will focus on mesh-based methods and consider mesh-free methods only briefly. Discretizing partial differential equations using a mesh involves using one method or another to divide a continuous space into finite cells and then solving a discrete version of the equations in terms of these cells, rather than over the entire continuous domain.

Although numerical methods are typically looked at as tools for getting meaningful approximate solutions out of a set of model equations after the modeling process has been completed. There are many technical reasons to consider the use of numerical methods as

a continuation of the modeling process. Considering these reasons are beyond our scope, but we may mention that an interpretation of numerical error as backward error provides one of these reasons. This is beacause backward error can be interpreted as a perturbation of the model equations, and therefore on a par with other perturbations from measurement and modeling errors. In particular, demonstrating a small backward error can usually show that the perturbation introduced by the numerics is smaller than the other sources of error in model equations, therefore showing that the perturbation effected by the numerical computation is less significant than other sources of error in the modeling. For more details on this interpretation of backward error, see (Moir, 2010, 2013; Corless, 1994b).

There is another more heuristic reason to regard numerical methods as part of the modeling process in the case of partial differential equations, which will be useful for our purposes. This has to do with the continuum hypothesis mentioned in the previous section. The continuum hypothesis requires only that a material behave like a continuum down to a scale much smaller than the size of the macroscopic material, not down to infinitesimal scales as is assumed in the construction of the model equations. This is because real materials are not actually continua. From the point of view of the continuum mechanics, the finite parcels of material at small scales become a rough approximation of the infinitesimal fluid parcels used to construct the equations of motion. But since real materials only behave like continua for fluid parcels of some small finite size, from a point of view of describing real physical behaviour, the discretization be regarded as *modeling* this fact.

It must be noted that this is not a technical argument. It is merely to point out that the discretization can resemble in certain respects a more accurate model of real materials than pure contiuua. For the resemblance to be strong it must be the case that the resolution of the mesh is small enough that the cells can reasonably be regarded directly as models of average behaviour over a certain volume. In such a case, and even in general for heuristic purposes, the quantities computed by the numerical method can be regarded as modeling direct *measurements* of material properies of small volumes of material. In this way, then, numerical models can be regarded in a significant sense as models of actual experiments and measurements of material behaviour. Indeed, numerical models are sometimes referred to as "experiments" in the continuum mechanics literature, since they play a similar role to that of physical experiments (Tritton, 1988, 428*ff.*).

Now let us turn to consider some of the details of how different approaches to solving the dynamical model equations numerically.

3.2.1 finite difference methods

The approach of finite difference methods is to discretize the dynamical material model by replacing derivatives, which involve infinitesimal differences, with finite differences between the solution evaluated at different points. This replaces differential equations that we cannot solve with *difference equations* that we can.

To illustrate what is involved in doing this, consider the Navier-Stokes equation (2.7) in

one spatial dimension with the pressure constant and with no stress or body forces, *viz.*,[9]

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = 0. \tag{3.19}$$

In this case, the solution we are looking for is $v(x,t)$. Thus, the solution can be regarded as a *surface* in the two dimensional xt-plane, but it can also be regarded as a spatial curve that changes over time. The finite difference approach discretizes the two dimensional domain by replacing it with a finite set of values of x and t. The simplest way to do this is to make all the x values and t values equidistant. These values define the *nodes* of a square mesh, in the sense that the nodes are the crossing points of vertical and horizontal lines through the nodes, the lines forming a mesh in a quite literal sense. The selection of nodes for the spatial part of a material model breaks up the domain or material into cells or elements (see figure 3.7). In the case of finite elements, which we consider here, the spatial units are called *cells*, and in the case of finite elements, which we consider in the subsequent subsection, the spatial units are called *elements*.[10]

As a simple example, then, replacing equation (3.19) with a simple forward difference scheme results in

$$\frac{v(x_i,t_{i+1}) - v(x_i,t_i)}{t_{i+1} - t_i} + v(x_i,t_i)\frac{v(x_{i+1},t_i) - v(x_i,t_i)}{x_{i+1} - x_i} = 0.$$

Using $h = t_{i+1} - t_i$ to denote the size of a time-step, and assuming that the spatial mesh has the same size $h = x_{i+1} - x_i$, to give a square mesh, this can be rearranged to give a formula that can predict the value of the solution at the next time t_{i+1} based on information available at the current time t_i:[11]

$$v(x_i,t_{i+1}) = v(x_i,t_i)(1 + v(x_i,t_i) - v(x_{i+1},t_i)).$$

Assuming that all the x_i are known at the current time, this provides a way of marching the solution forward in time.

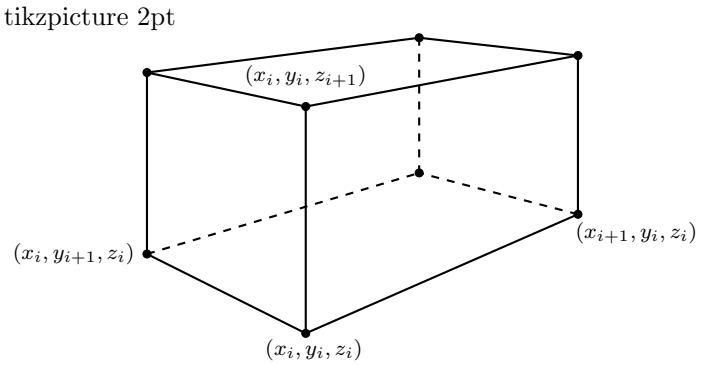
As was mentioned in the previous section, the accuracy improves the finer the mesh. In the case of a square mesh, as we have in our example, the x nodes and t nodes are all equidistant. The smaller we make the mesh size h, the more accurate the solution to the difference equations will be, in the sense of small forward error. The simplest approaches to replacing derivatives with differences are the forward and backward difference methods. The forward approach takes differences between the current and next time or place, the backward between the current and previous. These methods have the property that reducing h by $\frac{1}{2}$, which doubles the number of nodes for both x and t, the forward error also decreases by $\frac{1}{2}$. Thus, they converge on the correct solution as h goes to zero, but they do so at a linear rate.

Typically it is advantageous to have faster convergence, which is accomplished by the central difference method, which evaluates the solution at three points instead of two. This

[9] This special case of the Navier-Stokes equations is called Burgers' Equation.

[10] In the case of finite difference codes, the entire space-time domain is divided into cells. Whereas for finite element codes, at least for hydrocodes, it is common to divide only the space domain into elements and to treat time evolution using finite differences.

[11] Note that it is not to be assumed that such a formula will give accurate results. There is a good deal more than this to ensuring that a finite difference scheme is both convergent and numerically stable. This does, however, give the basic idea of what a finite difference scheme does for this type of equation.



1

Figure 3.7: A depiction of the node labeling for a three dimensional spatial mesh. This defines a *cell* in the case of finite difference, usually Eulerian, codes and an *element* in the case of finite element, usually Lagrangian, codes.

approach has the property that when h is decreased by $\frac{1}{2}$, the forward error is reduced by $\frac{1}{4}$. Thus, the convergence is quadratic rather than linear. This is usually represented by saying that the convergence is $O(h^2)$, read "order h squared". In this notation, the forward and backward difference methods are $O(h)$. For this reason forward and backward differences are called *first-order* methods and central differences is a *second-order* method. Third and higher-order methods are also possible. Higher order methods have the advantage that the forward error decreases much faster as you decrease h, but the computational cost in terms of processing, time and memory allocation becomes larger the higher the order. Thus, there is a trade off between the complexity, time and energy requirements of the computation on the one hand and the accuracy on the other.

The simple case we have considered of a regular grid is a quite special case. It is common to divide up the domain in an irregular way so that fewer nodes are used where the solution is varying slowly and more nodes are used where the solution is varying quickly. Since fast varying solutions can represent a source of ill-conditioning, an attempt is made to increase the accuracy in regions of rapid change to compensate. It is also advantageous to use fewer nodes in areas where the solution is varying slowly because this speeds up the computing time. Without prior knowledge of how the solution behaves there is no general method to predict where instabilities will occur. For this reason an initial low accuracy method can be used to make such predictions, so that a finer irregular mesh can be devised to generate an optimal mesh. This gets into the matter of mesh adaptation, which we will describe below.

advantages and disadvantages

The primary advantage of finite difference methods is their simplicity, both in conceptual terms and in terms of ease of implementation. The manner in which one discretizes the domain is reasonably straightforward in the sense that it simply requires selection of a set of nodes, and the finite difference schemes often give rise to a discrete problem that is fairly simple to solve. In technical terms, they give rise to sparse linear systems, for which there are very fast solution techniques (see, *e.g.*, Gerald and Wheatley, 1994, chapter 7).

One of the main disadvantages, however, is that these advantages generally require rectangular domains or domains that have high symmetry. It is very difficult to deal with irregular domains using finite differences. Having an irregular boundary means that the approach to discretization has to be changed near the boundary, which often makes it difficult to ensure accuracy of the same order as on the rest of the domain. It can also be difficult to locate nodes on the actual boundary (this occurs if the boundary is defined by an equation you need to solve), causing problems with applying the boundary conditions. Because finite element methods do not have the same difficulties, it is for these reasons, and other allied reasons, that finite elements are commonly used for irregular regions.

3.2.2 finite element methods

The approach of finite element methods is to discretize the dynamical material model by dividing up the material into finite sized pieces, the finite elements, and to then find an approximate solution over each element in such a way that the solution is continuous between elements. Intuitively, then, by stitching together the solution over each element we obtain a solution for the entire material.

There are a large number of variants of the finite element method, which vary in terms of how the material is discretized, how the solution is represented over each element and how the solution over each element is calculated. We will barely scratch the surface of this variety in the consideration in outline of what the method involves.

One approach to finite elements, which we will consider, treats time and space differently in the following sense. The material is divided into discrete parts using finite elements, but changes in time are treated with finite differences. Thus, the intuitive picture we should have is that for each instant in time we solve the equations for the material in terms of finite elements, then we step forward in time and solve the equations for the new state of the material in terms of finite elements. Thus, we can ignore changes in time for our purposes and simply consider how the equations are solved for a state of the material.

The **first step** is to divide the material into finitely many pieces. In one dimension this amounts to dividing a line into finite length pieces. Since impact modeling generally requires two or three dimensional material models, we will not consider this case explicitly even though it is the simplest. In two dimensions the material is usually divided up into triangular elements and in three dimensions pyramid shaped elements. We will restrict attention to the two dimensional case, the three dimensional case is similar. The division into elements, then, is accomplished by selecting three *nodes*, points on the material, to define each element, in such a way that all the triangular elements fit together to make up the material.

Now we may focus on a single element, *i.e.*, triangular piece of the material. The **second step** is to choose the kind of surface we will use to approximate the solution over the element. The simplest case is a flat sheet, which is the case we consider. Simple curved sheets are also possible, which allows for an increase accuracy (higher order approximation) at the expense of also increasing complexity.

It is the **third step** that is conceptually difficult, but we will consider a straightforward way to think about it. Because we approximate the solution over an element in terms of a surface, the approximate solution is smooth and defined across the entire surface. This allows us to substitute, the approximate solution back into the partial differential equation in a manner essentially similar to how we substituted our approximate value of $\sqrt{2}$ back into the equation $x^2 - 2 = 0$ in the example considered above. The amount by which the approximate solution fails to satisfy the partial differential equation is called the *residual*, just as before. The strategy for solving the partial differential equation over the element, then, is to *find the (flat) surface that minimizes the residual over the element.* We can think of this as minimizing the backward error over the element. This gives rise to a simple set of equations that we must solve to find the approximate solution over the element.

When we do this for each element the result is a large set of coupled equations that we can solve. They are *coupled* because for adjacent elements, which share an edge of a triangle, the solution must agree along the shared edge. This is what ensures that the solution will be continuous over the entire material. The **fourth step**, then, is to solve this large set of coupled equations, after adjusting the system according to the boundary conditions, which gives us the value of our approximate solution at each of the *nodes* defining the triangular elements. But this is all we need for a solution over the entire material, because fixing the solution at the nodes, fixes the flat sheet that connects them.

Once we have followed these four steps we have a solution for the state of the material at a particular instant of time. We can then carry the material forward in time using whichever finite difference scheme we have chosen for the time derivative.

advantages and disadvantages

This gives us a picture, though extremely simplified, of the finite element method. It should be reasonably clear why the finite element method is well-suited to irregular regions, *viz.*, because the method does not depend on where one places the nodes. The finite difference method has to use a different finite difference scheme for nodes that are placed in different places, *i.e.*, the form of the equations changes for differently distributed nodes. This makes is much easier for the finite element method to control error, since additional nodes may easily be added in places where the state or condition of the material is changing rapidly.

Because finite element methods are complex, they are also complicated to implement. Consideration of the details of the limitations of finite element methods takes us beyond the scope of this document. For more details on numerical methods for hydrocodes, including the advantages and disadvantages of finite element methods, see (Zukas, 2004). For details on hydrocode modeling in the specific case of impact cratering phenomena, see (Osinski and Pierazzo, 2012, chapter 17).

3.3 Lagrangian, Eulerian and Arbitrary Lagrangian-Eulerian

In the previous section we discussed the Eulerian and Lagrangian specifications of the flow, and their associated coordinates. These different coordinate systems have a significant effect on what kinds of phenomena can be treated numerically and on the complexity of the calculations that must be performed.

There are a number of features of the Lagrangian approach that make it natural for material modeling. Recall that Lagrangian coordinates are in the material frame, they travel with the material, as opposed to Eulerian coordinates that are in the space frame, material travels through them. Aside from focusing directly on modeling material dynamics, because Lagrange coordinates travel with the material the equations become simplified as a result of eliminating the need to model material transport. This simplifies the equations significantly since it removes any terms that track material motion. This means that the material derivative $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}\nabla\cdot$ (vector fields) or $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla$ (scalar fields) is replaced simply with the partial derivative $\frac{\partial}{\partial t}$. This removes the advection term in the equations of motion (2.12) and conservation of energy equations (2.11). Since the advective term in (2.12) is nonlinear, and nonlinear terms are notorious for computational difficulties, this is can be a major advantage of Lagrangian, or material, coordinates.

There are also a number of other advantages of material coordinates. Because each cell (finite-differences) or element (finite-elements) of the mesh has a fixed mass, the continuity equation (2.10) is satisfied automatically, reducing the number of equations that need to be solved. In the material frame the interfaces between materials are stationary, rather than moving through space, which makes it much easier to specify and solve both boundary conditions and contact conditions at material interfaces. There are also advantages in terms of

constitutive equations. As mentioned in section 2.1.3, constitutive equations or equations of state can sometimes be path-dependent, *i.e.*, depend on the history of the physical state of the material. This is difficult to account for in Eulerian codes, but easy to handle in Lagrangian codes because the codes compute the material history and store it in memory.

There are, however a number of limiations of Lagrangian codes. In material coordinates, the computation is dominated by contact computations. Because hydrocodes deal with high-energy impacts, the dominant physics occurs on very short time scales, the stresses and wave propagation across material interfaces must be computed very accurately. Conservation of linear and angular momentum must also be conserved where there is slippage between materials. Since the mathematical methods to handle this tend to be very complex, the computational procedures to handle contact-impact have very high computational cost (Zukas, 2004, 128). For this reason, the *contact processor*, *i.e.*, the subroutine that handles contact-impact calculations, tends to dominate the time required to run a simulation.

A major limitation of Lagrangian codes for high-energy impact problems is that in these problems the mesh can become severely distorted, causing the size of the smallest cells or elements to approach zero. This is a serious problem because the conditions for numerical stability for time stepping are determined by the size of the smallest cell or element. As a result, as the size of the smallest cell/element shrinks so does the time step, in order to maintain stability. Thus, large mesh deformations cause the time steps to become so small that the computational cost becomes too high to continue the simulation. Additionally, numerical artefacts can develop that render the calculation physically meaningless (Zukas, 2004, 134). This generally requires stopping the code and "rezoning" to define a new undistorted grid or by "erosion" techniques that remove some particularly distorted elements from the calculation.

For some problems a pure Lagrangian scheme is infeasible. An alternative in such cases is to use an Eulerian approach. Since Euler codes use spatial coordinates they must compute the transport of material. This tends to be handled by dividing each time step into two computational phases. The first phase is actually a Lagrangian time step, which is used to compute how the material evolves. This is followed by a second rezoning phase which involves transforming back to the original undistorted mesh and then computing transport of the material between cells. A key advantage of this approach is that large material distortions can be handled unproblematically and without drastically decreasing the time step size, making them important for high energy impact simulations.

Along with any advantage, however, come certain costs. With Euler codes difficulties arise in terms of how material interfaces are handled and how transport is calculated. Since Euler codes have a fixed mesh with material flowing through it, a single cell can contain portions of multiple materials. As a result, boundaries and interfaces of materials can at best be determined to within a cell width. A major difficulty for Euler codes, then, is to handle boundaries. A number of considerations arise here, including how to compute material fractions in a cell, how to apply consitutive equations and determine yield strength, and the order of transport of materials (see Zukas, 2004, p. 211 *ff.*, for more details on interface handling for Euler codes). Since accurate treatment of interfaces is essential for accurate calculation of material transport, we see that transport calculations are the Eulerian correlate of contact processing for Lagrangian codes. In a similar manner, it is transport processing

that dominates the computation in Eulerian codes.

The manner in which Euler codes need to handle boundaries and material interfaces makes the accurate treatment of interface conditions and surface motions difficult. This can lead to materials with similar properties being treated as bonded, poor treatment of slippage between materials, difficulties with handling materials with different strengths within a single cell, and computational instabilities due to tiny proportions of certain materials in a cell.

Thus, Euler and Lagrange codes are complementary from a modeling point of view since Lagrange codes do interfaces easily and accurately but cannot handle large distortions and Euler codes handle large distortions easily and accurately but have trouble treating interfaces accurately. Consequently, some modern codes attempt to combine the advantages of the two approach, leading to so-called aribitrary Lagrangian-Eulerian (ALE) codes. In simple terms, these codes in general allow variation of:

- the length of Lagrangian computations, from one time-step, producing an Euler-equivalent code, to all time-steps, producing a Lagrangian-equivalent code; and
- the frequency and location of rezoning, from every step and everywhere, producing an Euler-equivalent code, to only when and where Lagrangian time-steps fail due to element collapse, for an Lagrangian-type-equivalent code.

ALE codes allow for automatic rezoning, which allows for greater accuracy (higher order) than user-defined rezones (**?**, 224). Such codes also allow for computational optimzation by localizing intense calculations to parts of materials where there are large distortions. Some codes also allow for Euler treatment of some parts of the system and Lagrangian treatment of others. Since these codes are still continuum mechanics codes, topology changes due to fracture or drop formation of materials can be handled but must be done so manually.

The codes most in use by the impact community are *simplified* arbitrary Lagrangian-Eulerian (SALE) codes. These codes, packages designed to simplify the use of ALE methods to simulate high energy impact events, evolved from the original SALE code developed by Amsden et al. (1980). The package iSALE is a multi-material, multi-rheology shock physics code developed by Collins, Wünnemann, Ivanov and Melosh. This package includes a custom deviatoric stress model designed specifically for impacts in geologic materials and an efficient algorithm for handling porous compaction. See Collins et al. (2004) and Wünnemann et al. (2006) for details concerning the two dimensional version iSALE-2D. The three dimensional version, still in the development process,is described by Elbeshausen et al. (2009) and Elbeshausen and Wünnemann (2011). The three dimensional version is an important development in impact cratering studies because the two dimensional codes can only model vertical impact, but impacts are almost always oblique with 45° being the most typical incident angle. The iSALE code combines the capacities of other SALE codes for handing multiple materials with multiple constitutive equations that are used in the impact community, such as SALES-2, (see Gareth and Melosh, 2002), SALEB, (see, *e.g.*, Ivanov, 2005), and SALE-3MAT (Wünnemann et al., 2006, 517). For a thorough introduction to Lagrangian and Eulerian hydrocode methods more generally, see Zukas (2004) and (Benson, 1992).

3.4 verification and validation

Before moving on to consider some apsects of hyrdocode modeling of impact cratering processes, we must briefly mention two crucial steps for ensuring that the results of a hydrocode simulation will be meaningful: verification; and validation.

Verification involves showing that the code *solves the equations accurately*. A necessary condition for an accurate simulation is that it provide an accurate solution to the equations of the dynamical material model. Thus, verification concerns demonstrating that the code is numerically stable on certain test problems. This is usually assessed in terms of forward stability in the sense of a small forward error, which ensures that the precise behaviour of the simulation accurately reflects the behaviour of the exact solution.[12]

Assuming that the code is numerically stable, the other crucial step of validation is required, which requires showing that the computed solution models the target system accurately. This concerns, *inter alia*, showing that the code recovers experimentally known behaviour. There are two main ways that a hydrocode can be validated. One way is to compare the results of hydrocode models directly with experimental data available for very well understood laboratory experiments. For iSALE codes, this has been done by Pierazzo et al. (2008), Davison et al. (2011) and Miljković et al. (2012). Another is to compare the results of hydrocodes with the results of other hydrocodes that have already been well validated with experiment. A thorough benchmarking validation of this sort was performed by Pierazzo et al. (2008). This work is the result of the first phase of an ongoing collaborative project called the Impact Hydrocode Benchmark and Validation Project[13] organized and conducted within the impact cratering community, which seeks to develop a standard for comparison and validation of impact codes. Another important aim of this project is to better understand the strengths and limitations of different code implementations to prevent incorrect use of codes. The strongest limitations of impact codes arises from the type, extent and accuracy or material models. For more details on this see (Pierazzo et al., 2008) and (Osinski and Pierazzo, 2012, 262-267).

[12] Although not considered much in impact modeling, this could also be assessed in terms of backward stability in the sense of a small backward error or residual. This would ensure that the behaviour in the simulation is the exact behaviour of a very similar dynamical material model, allowing computational error to be interpreted in the same terms as modeling error. Backward stability in this sense is not typically used for solutions of partial differential equations because the interpolation required to compute the residual can be very large for already very expensive computations (Corless and Fillion, 2013). It could, however, be a useful tool for assessing the results of scientific studies that are less constrained by time-sensitive deadlines.

[13] See `http://www.psi.edu/about/staff/betty/Validation.html`.

4 Hydrocode Modeling of Impact Cratering

We conclude with a brief consideration of some more specific aspects of hydrocode modeling of the impact cratering process. Modeling impact cratering phenomena presents a significant challenge for a number of reasons, which include fast, high energy processes, poorly undersood constitutive behaviour of materials at high pressure and energy, and complex system composition and geometry. Thus, modeling an impact event accurately is an enormously complicated affair and is impossible in full detail given current technology. For this reason there is a trade-off between how much of an impact event can be modeled and detailed accuracy of impact processes. Thus, if one wishes to model the entire process of crater formation, one sacrifices a great deal in terms of accurate description of aspects of the different phases of the cratering process. On the other hand, if one wishes simulations highly accurate in detail, then one is limited to modeling only certain parts or aspects of a given phase of the impact process at any one time.

Given the level of understanding of hydrocode modeling we have from the preceding sections, we can have some appreciation for the complexities involved in numerical modeling of high energy wave propagation. Not only do terrestrial and planetary impacts by asteroids and comets involve high energy wave propagration, however, they involve a variety of such processes mixed with various structural dynamics processes. For example, the phenomena in impact cratering include:

- very large stresses, which imply very large strains and strain rates (fast loading and response times), which requires accurate models of the scaling of stress with strain rate in the equations of state that can be well-beyond experimentally accessible conditions;
- conditions easily push beyond the elastic limit of materials, which necessitates accurate constitutive models of plastic deformation as well as structural failure or fracture, which involves different phenomena dominant at different scales;
- very high shock pressures and shock waves occuring in both the impactor and the target, which produces multiple wave systems, *e.g.*, hypervelocity (Faster than sound) shock waves and slower shear and bending waves behind the shock front;
- waves traveling orders of magnitude faster than characteristic times for structural response, which results in highly localized deformation occuring before the effects of far away boundary conditions are felt;
- need to model post failure materials and phase changes, *e.g.*, solid-liquid (impact melt) and solid-solid (shock metamorphism) transitions, which can require computation of latent energies of transfomation;

- multiple materials with complex geometries, which also often have poorly understood constitutive behaviour at high pressures, strains and strain rates.

Because of the complexity of all of these processes, and the large uncertainties involved in constitutive equations and equations of state, modeling error often dominates over computational error, provided that conditions of numerical stablity are met. Thus, particularly in shock conditions, there is no practical advantage and significant increase in computational cost to use second or higher order discretization methods. Thus, linear interpolants are used for spatial discretization in terms of finite elements or finite differences and explicit forward difference schemes are often used for time stepping, which allow direct computation of physical quantities at a given time step in terms of the values at the previous time. Since the energies at which impact phenomena occur are often experimentally inaccessible, a great deal of reliance is placed on models for wave transmission that are based on laboratory experiments with rods and plates (see Zukas, 2004, chapter 7, for details).

4.1 hydrocode mediated inference for impact cratering processes

Due to the experimental inaccessibility and complexity of the behaviour of materials under physical conditions that occur in impact cratering processes, hydrocodes have become an important tool for gaining knowledge of the mechanics and dynamics involved in the formation of impact craters.

The formation process of hypervelocity impact craters, *i.e.*, where the impactor travels faster than the speed of sound in the target, has been divided into three roughly distinct phases: contact and compression; excavation; and modification (Gault et al., 1968; Osinski and Pierazzo, 2012, 3). A further, fourth, stage of chemical and hydrothermal alteration, which occurs over much longer time scales, is sometimes also included and is discussed in (Osinski and Pierazzo, 2012, 4,8) and references are provided there.

The contact and compression phase involves the initial impact between the impactor, typically travelling at 10s of km/s, and the target rock, and the translation of the high kinetic energy of the impactor to the target. The initial impact produces extremely high pressure shock waves (order of 100 GPa) in both the target and the impactor. The majority of the kinetic energy of the impactor is transferred by the shock wave in the target. The dissipation of the shock wave in the impactor typically results in the complete melting or vaporization of the impactor (Osinski and Pierazzo, 2012, 4). In addition to very high pressures this stage also involves very high strains and strain rates of 10^4-10^5 s^{-1} or larger (Zukas, 2004).

An example of the contribution of hydrocode modeling to understanding this part of the process is the modeling melt production and vaporization during the early phases of impact event. This is an area where constitutive behaviour is poorly understood, leading to the importance of hydrocodes. Using a 2D finite difference Eulerian hydrocode, and based on laboratory observations of hypervelocity impact flows, an early analysis by Okeefe and Ahrens (1977) on vertical impacts was able to determine a scaling relation for the ratio of melt-to-vapor production and the energy of the impactor. The scaling behaviour was in agreement with increase in melting with crater diameter observed in terrestrial craters and inferred from lunar craters. This was later validated and extended by Pierazzo et al. (1997), who showed that the region of melt and vapor production is roughly spherical. In the extension to oblique impacts, which requires 3D hydrocodes, Pierazzo and Melosh (2000) determined

the decrease in melt volume as a function of impact angle (20% from 90° to 45°; 50-90% for 30° down to 15°. This work also showed that for almost all angles (≥ 15°) the volume of melt is proportional to the volume of the so-called "transient cavity", which we will now consider.

The excavation phase, which is dominated by the hypervelocity shock wave in the target sequence, involves the opening of the impact crater and the formation of a "transient cavity", which is the temporary cavity formed by the initial movement of target material produced by the shock wave. The shock wave itself develops into two distinct waves. The first forms a roughly hemispherical wave travelling downwards, centred roughly at the depth of penetration of the impactor. The part of this wave that travels upward is reflected at the surface and travels back downward as rarefaction waves that travel downward. It is the combination of these waves that provide the forces that move target material in the formation of the transient cavity. The upper portion of the cavity (excavated zone) consists of melted target rock and material that is ballistically ejected from the crater. The lower portion of the cavity (displacement zone) consists of shocked and translated material, which includes melt-rich material (surface flows) that is emplaced outside of the transient cavity (Osinski and Pierazzo, 2012, 6). At the end of this phase the transient cavity is filled with a mixture of melt and impactites.

A contribution of hydrocode modeling in transient crater growth was made by Wünnemann et al. (2006), who developed a simple model capable of capturing dominant effects of the porosity of materials in the target sequence. They validated their porous compation model by showing that it agrees well with data from a wide variety of experiments, from static compation tests on one end to shock compression experiments from impacts on the other. Using iSALE they then used this porosity compation model to study the effect of porousity and friction internal to materials on the formation of the transient crater. Comparing porous and non-porous targets, they found that both porosity and internal friction play an important limitative role the crater growth.

The final phase of crater modification begins when the motion caused directly by shock waves stops and the transient crater relaxes and, if it is sufficiently large, begins to collapse to form the final crater. For large craters (greater than 2-4 km on Earth depending on target lithology), gravitational forces are strong enough to produce significant motion of the contents of the transient crater, resulting in complex crater formation. For such craters the falling material forms a central uplift, an inward and upward motion of material in the transient cavity. Finally, the initially high walls of the transient crater fall inward and downward, which involve the motion of large blocks of rock (Osinski and Pierazzo, 2012, 8). In addition to rapid motion, this last stage can involve processes that extend over geological time scales.

Hydrocode modeling has also played an important role in increasing understanding of the crater collapse process. As (Pierazzo and Collins, 2004, 13, *ff.*) point out, the initial contributions from hydrocode modeling were negative, showing consistently that if standard strength models of target materials we valid in the impact process, then the formation of central peaks, peak rings or external rings would not happen. Thus, these hydrocode results showed that in order to explain observed features of impact craters, the target materials must be significantly, though temporarily weakend, by shock loading from impact. Although our understanding of the formation of complex craters is still quite incomplete, hydrocode models based on strength weakening mechanisms are now able to account for central peak and peak

ring formation. For example, invoking acoustic fluidization to account for the strength weakening, Collins et al. (2002) showed that the duration of the strength weakening mechanism determines the difference between the formation of central peaks and peak rings.

Other hydrocode studies have involved broader studies of the entire process, modeling basin formation and melt production from impact to final crater structure, such as the detailed doctoral study by Potter (2012) of terrestrial and lunar craters. Such broader studies have also involved modeling crater formation for the large terrestrial craters, *viz.*, Popigai, Chicxhulub, Vredefort and Sudbury, by (Ivanov, 2005, 386 *ff.*). Here a detailed model of crater formation including the formation of the central uplift is developed but not without limitations in accuracy. Ivanov (2005, 404) point out that full crater formation models afford only limited accuracy, and that increasing accuracy requires modeling distinct phases of the process separately. Though they have limited accuracy, such models can also be quite informative when they are broadly successful in capturing known behaviour. In these cases, the deviations between model predictions and observation provide a guide to determining the processes responsible for the deviation, thereby providing a focus for future research by pointing out specific limiations in our understanding.

An interesting final example is that of the hydrocode simulation of the Chicxulub impact event by Pierazzo et al. (1998) and its role in catastrophic climate change. This example provide an interesting case of the power of incorporating borehole data into hydrocode modeling. Using borehole samples from the Chicxulub crater, Pierazzo et al. (1998) formulated a model of the stratigraphy of the target that included the chemical composition of the target rock. This enabled the prediction of the composition and dynamics of the impact plume from the event, in order to test specific hypotheses about the Chicxulub extinction event. This allowed a very detailed model of the impact to be developed and provided a powerful argument that it is the Chicxulub impact that explains the K-Pg extinction event.

We see, therefore, that though not without their limitations, hydrocode models are a key component in the study of impact cratering processes and in large measure because of the theoretical and experimental inaccessibility of the physical conditions that occur at planetary impact pressures and energies. Consequently, hydrocode models will continue to play a significant role in developments in the field in the future. Given that many features of cratering processes require full 3D modeling to model accurately, future increases in computational speed and memory handling that make 3D hydrocode modeling cheaper, simpler and more prevalent are sure to lead to quite significant developments in our understanding than have been provided largely based on 2D codes.

Bibliography

Amsden, A., H. Ruppel, and C. Hirt (1980). Sale: A simplified ale computer program for fluid flow at all speeds. Technical report, Los Alamos Scientific Lab., NM (USA).

Benson, D. J. (1992). Computational methods in lagrangian and eulerian hydrocodes. *Computer methods in Applied mechanics and Engineering 99*(2), 235–394.

Collins, G. S., H. J. Melosh, and B. A. Ivanov (2004). Modeling damage and deformation in impact simulations. *Meteoritics & Planetary Science 39*(2), 217–231.

Collins, G. S., H. J. Melosh, J. V. Morgan, and M. R. Warner (2002). Hydrocode simulations of chicxulub crater collapse and peak-ring formation. *Icarus 157*(1), 24–33.

Corless, R. (1994a). Error backward. In *Chaotic numerics: an International Workshop on the Approximation and Computation of Complicated Dynamical Behavior, Deakin University, Geelong, Australia, July 12-16, 1993*, Volume 172, pp. 31. American Mathematical Society.

Corless, R. (1994b). What good are numerical simulations of chaotic dynamical systems? *Computers & Mathematics with Applications 28*(10-12), 107–121.

Corless, R. M. and N. Fillion (2013). A graduate introduction to numerical methods.

Davison, T. M., G. S. Collins, D. Elbeshausen, K. Wünnemann, and A. Kearsley (2011). Numerical modeling of oblique hypervelocity impacts on strong ductile targets. *Meteoritics & Planetary Science 46*(10), 1510–1524.

Elbeshausen, D. and K. Wünnemann (2011). isale-3d: A three dimensional, multi-material, multi-rheology hydrocode and its applications to large-scale geodynamic processes. In *Proceedings, 11th Hypervelocity Impact Society Symposium*.

Elbeshausen, D., K. Wünnemann, and G. S. Collins (2009). Scaling of oblique impacts in frictional targets: Implications for crater size and formation mechanisms. *Icarus 204*(2), 716–731.

Gareth, S. and H. Melosh (2002). Sales 2: A multi-material extension to sale hydrocode with improved equation of state and constitutive model.

Gault, D. E., V. Oberbeck, and W. Quaide (1968). Impact cratering mechanics and structures.

Gerald, C. F. and P. Wheatley (1994). *Applied Numerical Analysis*. Addison-Wesley.

Ivanov, B. (2005). Numerical modeling of the largest terrestrial meteorite craters. *Solar System Research 39*(5), 381–409.

Miljković, K., G. S. Collins, D. J. Chapman, M. R. Patel, and W. Proud (2012). High-velocity impacts in porous solar system materials. In *AIP Conference Proceedings*, Volume 1426, pp. 871.

Moir, R. H. C. (2010). Reconsidering backward error analysis for ordinary differential equations. M.Sc. Thesis. The University of Western Ontario.

Moir, R. H. C. (2013). Structures in real theory application: A study in feasible epistemology. Ph.D. Thesis. The University of Western Ontario - Electronic Thesis and Dissertation Repository. Paper 1578. url: http://ir.lib.uwo.ca/etd/1578.

Okeefe, J. D. and T. J. Ahrens (1977). Impact-induced energy partitioning, melting, and vaporization on terrestrial planets. In *Lunar and Planetary Science Conference Proceedings*, Volume 8, pp. 3357–3374.

Osinski, G. R. and E. Pierazzo (2012). *Impact cratering: processes and products*. John Wiley & Sons.

Pavliotis, G. A. and A. M. Stuart (2008). *Multiscale methods: averaging and homogenization*, Volume 53. Springer.

Pierazzo, E., N. Artemieva, E. Asphaug, E. Baldwin, J. Cazamias, R. Coker, G. Collins, D. Crawford, T. Davison, D. Elbeshausen, et al. (2008). Validation of numerical codes for impact and explosion cratering: Impacts on strengthless and metal targets. *Meteoritics & Planetary Science 43*(12), 1917–1938.

Pierazzo, E. and G. Collins (2004). A brief introduction to hydrocode modelling of impact cratering. *Impact Studies: Cratering in Marine Environments and on Ice, edited by H. Dypvik, M. Burchell, and P. Claeys*, 323–341.

Pierazzo, E., D. A. Kring, and H. J. Melosh (1998). Hydrocode simulation of the chicxulub impact event and the production of climatically active gases. *Journal of Geophysical Research 103*(E12), 28607–28.

Pierazzo, E. and H. Melosh (2000). Melt production in oblique impacts. *Icarus 145*(1), 252–261.

Pierazzo, E., A. Vickery, and H. Melosh (1997). A reevaluation of impact melt production. *Icarus 127*(2), 408–423.

Potter, R. W. K. (2012). *Numerical modelling of basin-scale impact crater formation*. Ph. D. thesis, Imperial College London.

Tritton, D. J. (1988). Physical fluid dynamics. *Oxford, Clarendon Press, 1988, 536 p. 1*.

Wünnemann, K., G. Collins, and H. Melosh (2006). A strain-based porosity model for use in hydrocode simulations of impacts and implications for transient crater growth in porous targets. *Icarus 180*(2), 514–527.

Zukas, J. (2004). *Introduction to hydrocodes*. Elsevier.