Impacts are one of the most important processes in controlling planetary evolution. In addition to shaping the surfaces of the terrestrial planets, they also constituted an important source of heat in the early stages of planetary evolution. Impacts also may have catastrophic consequences on geological, atmospheric and even biological evolution. The importance of impacts had long been overlooked in part because of the tendency of geologic thought to favor slow change rather than short term catastrophic events as significant contributors to evolution. On the Moon, the impact origin of nearly all craters was not established until the Apollo era. A long-held myth was that lunar craters could not be due to impact because they were circular. In fact we shall see from the mechanics of shock wave propagation that this is to be expected in most instances.

To place the impact process in the proper context, we begin with a physical description via the physics of propagation of stress waves and the partitioning of energy between a projectile and target. We then discuss crater morphology as it relates to the mechanics of the impact process and the character and evolution of the impacted substrate. Finally we discuss the use of crater statistics as a means of determining the relative ages of planetary surfaces.

5.1 Impact Mechanics
The clearest exposition of the physics of the impact processes is presented in H.J. Melosh’s fine book *Impact Cratering: A Geological Process*. The following discussion follows the conventions in that text and the reader is encouraged to follow up with that reference for further details.

We pursue a first principles approach beginning with pressure waves in fluids, stress waves in elastic solids, stress waves in which permanent deformation of the target medium occurs, and finally shock waves.

5.1.1 Waves in a Fluid Medium

An impact on a surface compresses the surface and energy is distributed in the medium via waves with energies that are proportional to the impactor speed and energy, and on the material properties of the target medium. An exact treatment can’t be done, but it is tractable to look at an impact into a semi-infinite medium with an initial particle velocity equal to zero. The simplest useful physical description of the process is a one-dimensional pressure or acoustic wave in a strengthless fluid. This wave equation has both temporal and spatial dependences and the solutions describe the propagation of wave trains that travel at constant speed. The expression can be written

\[
\frac{\partial^2 P}{\partial t^2} = c^2 \frac{\partial^2 P}{\partial x^2}
\]

where \( P \) is pressure, \( x \) is the distance along the direction of propagation, \( t \) is time and

\[
c = \sqrt{\frac{K_o}{\rho_o}} \tag{2}
\]

is the wave speed, for which \( K_o \) is the bulk modulus measured at the zero pressure isentrope (= 1/compressibility) and \( \rho_o \) is the uncompressed density. In this arrangement pressure relates to particle velocity \( u_t \) by

\[
P = \rho_o u_t c.
\]

The PDEs that describe the wave propagation are linear, and so pressure waves can be superposed. The stress state in the medium is fully characterized by the pressure and there are no differential stresses.

The energy density in the wave is

\[
E = \frac{1}{2} u_t^2 - P \frac{d\rho}{\rho_o^2}, \tag{4}
\]

where the first term is the kinetic energy per unit mass and the second is the work per unit mass (\( =PdV \) where \( V \) is volume) done in compressing the medium integrated over the wave propagation time. For this simple situation the two contributions are equal.

5.1.2 Stress Waves in an Elastic Medium

---

5–2
Stress waves that propagate through a homogeneous elastic medium are more complicated than pressure waves in a fluid because elastic materials can support differential stresses. There are two types of stress waves in homogeneous elastic media—longitudinal and transverse. **Longitudinal waves** are equivalent to the pressure waves in a fluid. **Transverse waves** have no analog in fluids and arise because solids, in addition to resisting compression, also resist changes in shape (i.e., distortion). Resistance is related to the shear modulus ($\mu$), which is always smaller than the bulk modulus, so the transverse waves propagate more slowly than the longitudinal waves. The transverse waves propagate in a direction perpendicular to that of the longitudinal waves.

Elastic waves are also described by linear PDEs, however, unlike the case of pressure waves in a fluid, the pressure alone does not fully describe the stress state. In one dimension the equations of motion for longitudinal and transverse waves can be written

\[
\frac{\partial^2 u_l}{\partial t^2} = c_l^2 \frac{\partial^2 u_l}{\partial x^2}, \tag{5a}
\]

and

\[
\frac{\partial^2 u_t}{\partial t^2} = c_t^2 \frac{\partial^2 u_t}{\partial x^2}, \tag{5b}
\]

where $u_l$ is the particle velocity in a longitudinal wave that propagates at speed $c_l = [(K_o + 4/3\mu)/\rho_o]^{1/2}$, $u_t$ represents either component of orthogonal transverse particle velocity components that both propagate at speed $c_t = (\mu/\rho_o)^{1/2}$. The propagation velocities depend on the bulk and shear moduli. Longitudinal waves travel faster than pressure waves in a fluid because the elastic materials resistance to distortion augments the resistance to compression, which is controlled by $K_o$. The transverse wave speed depends only on the shear modulus because the transverse wave motion does not cause a change in the volume of the material that it passes through.

Unlike the situation for a fluid, the stress state in a solid must be described by all of the stress components and not just the pressure. The full stress tensor with both normal and shear components is relevant. The stresses can be written

\[
\sigma_l = \rho_o u_l c_l \tag{6a}
\]

and

\[
\sigma_p = \left(\frac{\nu}{1-\nu}\right) \sigma_l, \tag{6b}
\]

where $\sigma_l$ is the longitudinal stress, and $\sigma_p$ is the stress component perpendicular to the direction of propagation. (We use engineering convention where tension is positive). The parameter $\nu$ is **Poisson’s ratio** which indicates the amount of lateral contraction that occurs for a given longitudinal extension. The stress in a transverse wave is pure shear, with the only non-zero components of the stress tensor being in the off-diagonal. The shear stress $\tau$ is

\[
\tau = \rho_o u_t c_t. \tag{7}
\]

Stresses associated with transverse waves are lower than in longitudinal waves because the particle velocities are lower. Like pressure waves in a fluid, the energy density in either
wave type is related to the sum of the mean kinetic and distortional energies integrated over the time of wave propagation. Transverse waves are probably not too important in the cratering process because the strength of these waves is limited by the shear strength of the target. In contrast the longitudinal wave strength has no limit.

5.1.3 The Importance of the Free Surface and Internal Interfaces

Impacts on planetary surfaces and their stress waves will propagate into the subsurface which in many circumstances may contain interfaces between layers with differing mechanical properties. The interaction of impact-generated stress waves with free surfaces and internal interfaces generates new waves, and the interactions can be described exactly in a physical sense for linear waves such as discussed so far. Interactions of waves at interfaces are easy to describe. Layers of different materials may have different material properties, indicating that waves will propagate at different speeds and experience different levels of stress. However, velocities and stresses at the interfaces of materials must be continuous. To conserve energy more waves must be generated.

At a free surface, however, normal and shear stresses cannot be supported. When a wave strikes it, another wave must be generated that maintains the normal stress at the zero level. If a compressional wave impinges on a free surface, it must therefore generate a corresponding tensional wave such that the sum of the longitudinal stresses (Equation (6a)) vanish at the free surface. The compressional and tensional waves overlap for a short period of time resulting in a complex near-surface stress field. But while the stresses vanish at the surface, the velocities do not. Because the waves are linear we may add their velocities during the time during which they overlap. Because the tensional wave is opposite in sign and moves in the opposite direction of the compressional wave, there is a double negative and the wave speed at the free surface equals \(2u_l\). This is known as the velocity doubling rule.

Velocity doubling is responsible for producing spalling of the surface. Spalls are target fragments that break off at the free surface and have a velocity of twice that of the particle velocity in the compressed medium. Spalls do not remove a large amount of material in an impact but they have the capability to eject some material at speeds that exceed a planet’s escape velocity. Hence may samples of planetary surfaces perhaps be ejected into space.

5.1.4 The Hugoniot Elastic Limit

The stress and particle velocities in elastic waves increase as the intensity of the source increases. Eventually the stresses will exceed the strength of the material and plastic or irreversible work will be done in the target medium. Solids can resist almost arbitrarily large compressive stresses, however, their ability to resist stress differences is limited. These stress differences occur because waves generate both longitudinal and transverse stresses. The longitudinal stress is a factor \((1 - \nu)/\nu\) greater than the stress generated by the transverse wave. So as the strength of the wave increases the absolute difference between the stresses also increases. Eventually the stress difference will reach a yield stress in which permanent deformation of the medium will occur. The medium deforms by plastic flow and there is little subsequent increase in the stress difference as the medium cannot support it.
Consider the failure of a typical rock. At (relatively) low stresses (<2 GPa) the shear stress
\[ \tau = -\frac{\sigma_l - \sigma_p}{2} \] (8)
and mean pressure
\[ P = -\frac{\sigma_l + 2\sigma_p}{3} \] (9)
follow a linear relationship defined by a Coulomb friction law. At higher pressures strength is nearly constant and failure is characterized by plastic distortion or ductile flow. The slope of the failure envelope in the Coulomb portion of the curve (30° – 45° for most rocks) rises more steeply than the line representing the stress in a longitudinal elastic wave, which begins at the origin and is \(3(1 - 2\nu)/2(1 - \nu)\), approximately 30° for \(\nu = 0.25\), a typical value for rocks. So failure doesn’t occur at low stresses. But the failure envelope eventually flattens out and the elastic stress trajectory intersects it at \(\tau = Y/2\). The longitudinal stress at this point is
\[ \sigma_l = -\sigma_{HEL} = -\frac{1 - \nu}{(1 - 2\nu)Y} \] (10)
where \(\sigma_{HEL}\) is the Hugoniot Elastic Limit or HEL. Beyond this limit the maximum shear stress \(\tau\) remains constant at \(Y/2\) as \(P\) increases. Both longitudinal and transverse waves propagate in a manner such that \(-(\sigma_l - \sigma_p))/2 always equals \(Y/2\). For very strong waves this stress difference is minuscule compared to the mean stress and can be neglected. The wave can be approximated as a strong pressure wave. The propagation speed of a wave after the HEL has been exceeded drops significantly. For an elastic wave below the HEL both shear and bulk moduli contribute to \(c_l\). Beyond the HEL, however, only the bulk modulus contributes significantly and the speed drops to nearly the bulk wave speed. The bulk modulus increases with pressure, and so the wave speed begins to rise again in high pressure waves. In extremely strong stress waves, called shock waves, the wave propagation speed may exceed \(c_l\).

5.1.5 The Rankine-Hugoniot Equations

Shock waves travel faster than elastic waves in an uncompressed medium and are therefore supersonic. They ‘outrun’ elastic waves and add the energy of the elastic waves to their own. Shock fronts tend to be abrupt, and they are most often represented in a mathematical sense as discontinuities in pressure, particle velocity, density and pressure, which is convenient though not strictly true. The basic equations that describe abrupt shock fronts were first derived by R.P. Hugoniot in 1887 and involve the pressures \(P_o\) and \(P\) in front of and behind a shock front, the particle velocity \(u\) behind the shock front, the shock velocity \(U\), the compressed and uncompressed densities \(\rho\) and \(\rho_o\) and the internal energies or energies per unit mass \(E_o\) and \(E\) on either side of the shock front. The equations derive from the conservation of mass, momentum and energy across the discontinuity.

Consider a block of material through which a shock wave passes. At time \(t\) the block can be divided into a part, with length \(l_s\) through which the shock wave has passed and another, \(l_u\) through which it hasn’t. At a later time \(t'\) the shock wave has propagated a
distance \( U(t' - t) \) and the particle velocity has progressed \( u(t' - t) \). The lengths of the unshocked \((l'_u)\) and shocked \((l'_s)\) regions are

\[
l'_u = l_u - U(t' - t) \quad (11a)
\]

and

\[
l'_s = l_s + U(t' - t) - u(t' - t). \quad (11b)
\]

The mass contained in the unshocked part of the block at time \( t \) is the product of its volume and density \( \rho_o l_u A \), where \( A \) is area, while the mass in the shocked section is similarly \( \rho l_s A \). Mass conservation dictates that we equate these quantities:

\[
\rho l_s A + \rho_o l_u A = \rho l'_s A + \rho_o l'_u A \quad (12)
\]

By canceling \( A \) and substituting (11) yields

\[
\rho(U - u)(t' - t) - \rho_o U(t' - t) = 0 \quad (13)
\]

which, with further canceling, yields the first Hugoniot equation:

\[
\rho(U - u) = \rho_o U \quad (14)
\]

which relates the shock and particle velocities to the densities in the medium before and after shock wave passage.

The second equation comes from momentum conservation. Because the pressure \( P \) on the shocked end of the block is greater than the pressure \( P_o \) on the unshocked end there is a net force \( F = (P - P_o)A \) that acts in the direction of shock wave propagation. The momentum in the material at time \( t \) is \( \rho l_s u_p A \), which differs from that in the material at time \( t' \) \( (\rho l'_s u_p A) \) at the applied force over the time interval \( F(t' - t) \). The momentum balance requires

\[
\rho l'_s u A - \rho l_s u = (P - P_o)A(t' - t). \quad (15)
\]

By canceling and substituting (11b) we obtain

\[
\rho(U - u)u = (P - P_o) \quad (16)
\]

Using (14) we arrange to get the second Hugoniot equation

\[
P - P_o = \rho_o U u, \quad (17)
\]

which relates velocities to the pressures ahead of and behind the shock wave.

The final expression that we seek derives from conservation of energy and describes the internal energy of shocked material. As for momentum, the energy in the system changes between times \( t \) and \( t' \) because the passage of the shock wave does work on the system. Because the end of the unshocked block has a zero displacement the increase in energy during the time interval is

\[
P A u(t' - t) \quad (18)
\]
where $PA$ is the force on the shocked end of the block and $u(t' - t)$ is distance over which the force acts. Contributions to energy come from the internal energies on either side of the shock front and the kinetic energy in the shocked part of the block. At time $t$:

$$E_{tot}(t) = \rho_o l_u E_o A + \rho l_s E A + \frac{1}{2} \rho l_s u^2 A,$$

(19)

and at time $t'$

$$E_{tot}(t') = \rho_o' l'_u E_o A + \rho l'_s E A + \frac{1}{2} \rho l'_s u^2 A.
$$

(20)

We equate (19) and (20), adding the contribution from (18) to find

$$E_{tot}(t') - E_{tot}(t) = PAu(t' - t).$$

(21)

We now substitute (19) and (20) into (21), and then further substitute (11). After various cancellations and rearrangement we obtain

$$-\rho_o E_o U + \rho E(U - u) + \frac{1}{2} \rho u^2 (U - u) = Pu.$$  

(22)

We next substitute $\rho_o U$ from (14) to get

$$\rho_o U (E - E_o) + \frac{1}{2} \rho_o u^2 U = Pu.$$  

(23)

From the first two Hugoniot equations [(14) and (17)] we may derive convenient expressions for the shock and particle velocities:

$$U = \frac{1}{\rho_o} \sqrt{\frac{(P - P_o)}{(V - V_o)}}$$  

(24)

and

$$u = \sqrt{(P - P_o)(V - V_o)},$$  

(25)

where $V_o = 1/\rho_o$ and $V = 1/\rho$ are the uncompressed and compressed specific volumes. Substituting (24) and (25) and re-arranging gives

$$E - E_o = \frac{1}{2} (P + P_o)(V_o - V),$$  

(26)

which is the third and final Hugoniot equation.

The Hugoniot equations are explicitly valid for fluids and also for elastic solids when $P$ is replaced by $-\sigma_l$.

5.1.6 Equation of State

In the Rankine-Hugoniot equations there are four unknowns ($P$, $\rho$, $u$ and $U$) but only three equations, given these expressions the problem is underdetermined. Another
equation is needed to completely specify conditions on either side of the shock front. The **equation of state** relates scalar pressure, specific volume, internal energy, and takes into account complexities of the atomic, molecular and crystalline properties of the material, i.e. the thermodynamic properties of the system. The equation of state has the form \( P = P(V, E) \). The equation of state cannot be derived from first principles, but rather is approximated from experiments, though most estimates are for materials at speeds of \( \leq 5 \text{ km s}^{-1} \), the lower end of the range of interest. (There has been limited analytical work on the equation of state of hydrogen, the simplest element, at high pressures, due to its relevance to the internal structures of giant planets. This work and its experimental verification will be discussed in Chapter 9.)

It has been shown that for particle velocities \(<10 \text{ km s}^{-1}\), the equation of state can often be fortuitously expressed algebraically in terms of the shock and particle velocities, i.e., in the form

\[
U = a + bu
\]

where \( a \) is a constant usually close to the isentropic bulk sound speed at zero pressure and

\[
b = \frac{1 + \Gamma}{2}
\]

where \( \Gamma \) is the **Gruneisen parameter**

\[
\Gamma = \frac{\alpha K_v}{\rho_0 C_v}
\]

where \( \alpha \) is the volume coefficient of thermal expansion and \( C_v \) is the specific heat at constant volume. Such a parameterization of the equation of state is the basis for the widely used **Tillotson equation of state** valid at pressures \(<100 \text{ GPa}\). Interestingly, there is no physical reason why the simple linear approximation works so well.

The shock wave equation of state can be represented in a \( U - u \) plot, or in a \( P - V \) diagram. The latter is more common, but the representations are equivalent and can be converted to each other using the Hugoniot equations. \( P - V \) curves represent the most common presentation in the thermodynamic sense but \( U - u \) shows the quantities that are actually measured. Equation of state curves are very frequently misunderstood. They do not represent a continuum of states but rather loci of a series of individual shock events. A Hugoniot may more correctly be described as representing all possible end states. Hugoniot curves on a \( P-V \) path are not thermodynamic paths because shock compression is not a thermodynamically reversible process. Physical properties change rapidly across a shock front so there are no intermediate states. The Hugoniot equations conserve mass, momentum and energy but not entropy. The actual path that a material passes through is a straight line in \( P - V \) space called the **Rayleigh line**.

Not all energy behind the expanding shock is available to drive further expansion of the wave since some is consumed in heating, melting or vaporizing the material behind the shock front. To account for this loss it is assumed that when the shock front expands the total energy is decremented by the amount of heat that will eventually be irreversibly deposited. This approximation is valid at low pressure if the release adiabat approximates...
the Hugoniot and if the Hugoniot is described by a linear shock particle velocity relation. After release from high pressure the decompressed material is seldom returned to its initial state. The high pressure induced by the shock wave is ‘unloaded’ by propagation of the rarefaction wave, which radiates from free surfaces into the shocked material. Unlike elastic wave reflection, where the tensional wave propagates away from the free surface at the same velocity as the incident pulse, the rarefaction wave from a shock event generally moves faster than the shock wave. The speed is proportional to the slope of the adiabatic release curve on the $P − V$ diagram

$$c_r = \sqrt{\frac{dP}{ds}} \quad (230)$$

at constant entropy. The area between the release adiabat and the hypotenuse of the $P − V$ plot corresponds to the amount of irreversible work done by the shock wave. This corresponds to the amount of heating or waste heat produced in the impact and allows an estimation of the amount of melting and vaporization.

5.1.7 Comminution

The energy expended in fracturing and crushing the target can be estimated based on calculations of the new surface area of fragmented material. Experiments have shown that basalt targets follow a simple comminution law

$$\frac{m}{M} = \left( \frac{l}{L} \right)^\alpha \quad (31)$$

where $m$ is integrated mass of the fragments, $M$ is the total mass of ejected material with a size equal to or smaller than $l$, $L$ is the size of the largest fragment, and $\alpha$ is an empirical constant ($0.3 < \alpha < 0.6$). Most of the area is taken up by the finest particles.

**Breccias** are broken and fractured rocks that are commonly found in the floors of impact craters.

5.1.8 Ejecta

Studies from explosion cratering have shown that the amount of material ejected from a crater does not scale in a simple way with crater size. This is because ejecta scales with gravity while crater size is most sensitive to target strength. In practice, the amount of ejected material is found by integrating cumulative mass given measurements of ejecta velocities.

Most crater ejecta falls back onto the surface of the planet, as very little achieves escape velocity. Ejecta follow nearly ballistic trajectories. The horizontal distance $R_b$ followed by ejecta when $R_b << R_{planet}$ is

$$R_b = \frac{v_e^2}{g} \sin2\phi, \quad (32)$$
where $v_e$ is the ejection velocity, $\phi$ is the ejection angle, and $g$ is planetary surface gravity. The time of flight of ejecta is

$$T_f = \frac{2v_e}{g} \sin(\phi).$$

(33)

Experiments show that particles nearest the location of the impact are ejected at the highest velocities because they achieve the highest pressures. The smallest particles are ejected at the highest angles. Ejection velocity falls off quickly as a function of ejection angle.

The size of ejected material generally follows a power law. The most common empirical representation is a cumulative power law

$$N_c(m) = Cm^{-b},$$

(34)

where $N_c$ is the cumulative number of fragments with mass greater than $m$, and $b$ and $C$ are constants.

The ejecta blankets of many craters on Mars have a distinctive morphology unlike seen on any of the other planets. Craters up to diameters from about 3 to 15 km are surrounded by a blanket of material that extends about one crater radius beyond the rim and terminates with a convex scarp. Larger craters of this type display petal-like lobes in their continuous ejecta blankets. Craters with these kinds of ejecta blankets are referred to as rampart craters, and the surrounding material are referred to as fluidized ejecta blankets (FEBs). The FEBs look like material that has flowed on the surface instead of being ballistically emplaced. The blankets appear to have behaved in a fluidized manner. These structures are in fact believed to represent evidence that a fluid was present in the target material at the time of the impact. The likely fluid is water. J. Garvin used topography from the Mars Global Surveyor mission to demonstrate that all craters in the Martian northern lowlands greater than about 3 km in diameter display FEBs. This is evidence that water or ice was pervasive in the shallow subsurface of Mars during its early history.

5.1.9 Cratering Efficiency

**Cratering efficiency** is defined as the effectiveness of the impact process in displacing target material to produce a surface depression due to excavation and subsequent modification of the crater cavity. Modification is evaluated with respect to the transient cavity, a largely theoretical estimate of the original, unmodified crater cavity. The ratio of the volume of ejecta (as a proxy for excavated volume) to that of the final fresh crater cavity is a measure of this efficiency. Efficiency can also be defined in terms of energy, using the modeled kinetic energy of the impactor versus the final crater dimensions in terms of geometry.

5.1.10 Energy Partitioning

Studies of the partitioning of energy during impact events are of particular interest to planetary scientists because studies of the thermal states of planets require estimates of the amount of energy that is available for heating. Experimental results from analysis of
Table 5.1
Energy Partitioning in Hypervelocity Impact

<table>
<thead>
<tr>
<th>Energy expended for</th>
<th>Percent Projectile Kinetic Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irreversible and waste heat</td>
<td></td>
</tr>
<tr>
<td>- Projectile</td>
<td>4-12</td>
</tr>
<tr>
<td>- Target</td>
<td>19-23</td>
</tr>
<tr>
<td>Comminution</td>
<td>10-24</td>
</tr>
<tr>
<td>Ejecta Throwout</td>
<td>43-53</td>
</tr>
<tr>
<td>Etc.</td>
<td></td>
</tr>
<tr>
<td>- Elastic waves</td>
<td>&lt;1</td>
</tr>
<tr>
<td>- Radiant energy</td>
<td>negligible</td>
</tr>
</tbody>
</table>

Estimates from Gault and Heitowitz [1963]

aluminum projectiles into basalt targets give the following breakdown of the partitioning of energy in an impact:

Energy partitioning can also be estimated from numerical experiments. Such analyses are approximate at best but, unlike experiments, provide insight on the change in energy partitioning with time during the course of the impact event. The Hugoniot Equations can be used to estimate the fraction \( f \) of a projectile’s kinetic energy that is manifest in the target as kinetic \( (KE) \) and internal \( (IE) \) energy:

\[
f(KE) = f(IE) = \frac{(u/u_t)}{(1 + u/u_t)}, \quad (35)
\]

where \( u_t \) is the particle velocity in the target. Note that the internal and kinetic energy of the target are equal and written

\[
KE = IE = \frac{1}{2} \rho_o t_c U_t u_t^2, \quad (36)
\]

where the subscript \( t \) refers to the target and \( t_c = L/U \) where \( L \) is the length of the projectile and \( t_c \) is the time it takes the shock wave to reach the rear surface of the projectile.

In practice, the partitioning of energy changes with time.

5.1.11 Estimating Shock Pressure

The Hugoniot is also useful in estimating the shock pressure during an impact event in a simple graphical way. At the point of contact shock waves propagate through both the target and projectile. At the contact of these media pressures and velocities must match. The technique of **impedance matching** is used to estimate the impact conditions. Impedance matching derives from the continuity of pressure and particle velocity across the boundary between the projectile and target.
To perform impedance matching, the Hugoniot must be plotted in terms of pressure and particle velocity. Usually Hugoniots are given in terms of pressure-volume \((P - V)\) or shock and particle velocity \((U - u)\), but in either case the conversion can be simply accomplished using the Rankine-Hugoniot equations. The Hugoniot of the target is plotted normally. The Hugoniot of the projectile must be plotted ‘backwards’, starting at the particle velocity equal to the impact velocity. The intersection of the Hugoniots yields an estimate of the pressure behind the shock waves in both the target and projectile. From (3), the peak pressure associated with a projectile (subscript \(p\)) hitting a target (subscript \(t\)) can be written

\[
\rho_o P(u_i - u_p)U_p = \rho_o u_t U_t
\]  

where \(u_i\) is the impact velocity. Note that in addition, the particle velocity of the target and projectile can be read off from the \(x\)-axis corresponding to the intersection point.

In terrestrial impact craters, impact pressures and temperatures can be constrained from field work. Low P/T minerals that are shocked will achieve higher P/T phases. For example, at increasingly higher pressures there will be a transition from quartz to coesite to stishovite to amorphous silica glass. Plagioclase feldspar undergoes a transition to the higher pressure phase maskelynite.

5.2 Stages of the Impact Process

The cratering process is a rapid process, but one which contains a distinct sequence of events. As first proposed by D.E. Gault, impact events are divided into the contact and compression, excavation, and modification stages.

5.2.1 Contact and Compression Stage

The first state of impact occurs when the projectile comes into contact with the target surface. The fast-moving projectile pushes target material out of its path, compressing it and accelerating it to a significant fraction of the impact velocity. Concurrently, the target’s resistance to penetration decelerates the projectile. Material in the contact zone of the target and projectile are strongly compressed. Shock pressures developed during hypervelocity impacts generally reach several hundreds of GPa and thus far exceed the Hugoniot Elastic Limit of the target and impactor. The contact and compression state ends after the projectile has been unloaded from high pressure by the rarefaction wave generated when the shock wave impinges on the trailing end of the projectile. This is the shortest stage of the cratering process, with a duration of contact

\[
\tau = \frac{L}{v_i}
\]  

where \(L\) is the projectile diameter and \(v_i\) is the impact velocity. Contact and compression lasts for less than a second for all but the largest impacts. In planetary-scale impact events most of the projectile is vaporized.

The highest velocity material is ejected when the projectile first impinges on the target. The non-planar impingement of the projectile on the target generates localized zones of high pressure near the edges of the projectile. Oblique contact angles are responsible for
increased peak pressures and particle velocities. From the highest pressure zones jets of highly shocked material are "sprayed out", at a velocity that may be a few times impact velocity. This process is referred to as jetting. Jetting occurs very rapidly and usually finishes within about half the time of the contact stage (Equation (38)). Jets include minor amounts of material in vertical impacts but may be more important in low angle impacts.

5.2.2 Excavation Stage

Excavation of the target initiates as the contact/compression stage ends. During this time an approximately hemispheric shock wave propagates through the target material, weakening it severely. Because the shock wave is nearly hemispherical craters are usually circular and do not so deviate unless impact angles are less than 10° or so. The shock wave and rarefaction wave set target material in motion, initiating excavation that forms the crater. Decompression by the rarefaction wave converts much of the internal energy gained in compression into mechanical work. The energy converted is equal to the area between the release adiabat and the lines parallel to the axes on the $P - V$ plot.

The peak pressure of the shock wave decays in an exponential fashion as

$$P(r) = P_o \left( \frac{r_o}{r} \right)^b$$

where $1.5 \leq b \leq 3$, $P_o$ is the initial peak pressure, $r_o$ is the length of the isobaric core, and $r$ is the penetration distance.

The crater that produced in the excavation event is many times larger than the projectile, unlike low velocity impacts. The crater is formed by a combination of compression and ejection of material from the crater. The ejecta blanket covers the surrounding terrain. The targets material strength and gravity become important. Excavation produces a bowl-shaped transient cavity that will later collapse under gravity. The shape of the crater cavity is determined by the depth attained by material that is displaced downward by compression beneath the crater. The excavation stage of a crater can take seconds to minutes to reach completion. Near the end of the excavation stage the target’s material strength and gravity become important in determining crater shape.

5.2.3 Modification Processes

During the ensuing modification stage, loose material slides down the steep interiors of the walls of small craters and pools on the crater floor. Larger craters collapse more dramatically – slump terraces form on the walls and central peaks rise in the interior. At progressively larger diameters peak rings and rings will develop. At much longer time scales, diameters isostatic rebound may follow collapse, eventually flattening out craters.

On the long term, planetary surfaces, even those made of rock, behave to some extent like a fluid. All substances will flow or creep on a characteristic time scale. The creep rate is strongly temperature dependent. The process of long-term flow of topography is referred to as viscous relaxation, and is a consequence of the release of potential energy of topography due to the force of gravity; topography will seek to flatten out to minimize the energy in the system.
Viscous relaxation of impact craters represents a means of probing the time-integrated thermal evolution of the shallow planetary subsurface. It is possible to quantify in a simple way the process of relaxation. For a viscoelastic material the strain rate ($\dot{\epsilon}$) is a function of the differential stress associated with topography ($\sigma$) as

$$\dot{\epsilon} = A\sigma^n e^{Q/RT},$$  
(40)

where $A$ is the pre-exponential frequency factor, $n$ is an experimentally-derived constant that equals unity for linear or Newtonian flow is greater than one for nonlinear or non-Newtonian flow, $Q$ is the activation energy, $R$ is the gas constant, and $T$ is temperature.

In a viscoelastic material the resistance to flow is commonly expressed by the Maxwell time, $t_M$, which can be expressed in terms of the strain rate as

$$t_M = \frac{\epsilon}{\dot{\epsilon}} = \frac{\eta}{\mu},$$  
(41)

where $\epsilon$ is the elastic strain, $\eta$ is the viscosity and $\mu$ is the shear modulus of the substrate. The viscosity is a resistance to shear and is a ratio of the shear stress over strain rate as

$$\eta = \frac{\tau}{2\dot{\epsilon}}.$$  
(42)

Because even solid rock behaves as a fluid over geological time scales (hundreds of millions to billions of years), crater topography will relax away over time, depending on the viscosity of the substrate. The rate at which a crater with depth $d$ and diameter $D$ will change can be simply estimated to first order, as noted by Melosh [1989]. Stress is a force per unit area so we take the negative buoyancy force associated with the crater cavity $\pi/8\rho gdD^2$ and divide by the area of the crater cavity $\pi/2D^2$. This gives the shear stress ($\tau$) beneath the crater

$$\tau \approx \frac{1}{4} \rho gd,$$  
(43)

where $\rho$ is substrate density and $g$ is gravitational acceleration. From Equations (42) and (43), we find the strain rate of material that flows inward towards the crater to be

$$\dot{\epsilon} = \frac{\rho gd}{8\eta},$$  
(44)

The rate of uplift of the floor is approximately the crater diameter times the strain rate, so

$$-\frac{dd}{dt} = D\dot{\epsilon} = d(t)\frac{\rho gd}{8\eta}.$$  
(45)

Solving (44) as a function of time gives the temporal change of crater depth due to viscous relaxation of crater topography

$$d(t) = d(0)e^{-\frac{\dot{\epsilon}}{t_r}},$$  
(46)

where $d(0)$ is the initial depth, and $t_r$ is the relaxation time of topography for a crater in a uniform viscous medium

$$t_r \approx \frac{8\eta}{\rho gd}.$$  
(47)
Equation (47) indicates that craters relax faster for higher gravity or density. In addition, craters relax faster for lower viscosities and large diameters. So large craters relax faster than small ones. Extending this, the long-wavelength characteristics of craters (e.g. diameters) relax faster than short-wavelength characteristics (e.g. rims). This is true if the substrate viscosity is either uniform or decreases with depth (due to an increase of temperature with depth in the interior). To first order we know that is what the viscosity of a shallow planetary interior is like because some craters on the surfaces of the icy satellites exhibit only short-wavelength structure such as rims; the broader crater cavities have relaxed away. But sometimes even the rim of a crater on an icy satellite decays away and what is left is a quasi-circular albedo marking that is in essence the “ghost” of a relaxed crater. These ghost craters are referred to as palimpsests.

5.3 Crater Types

5.3.1 Simple Craters

Simple craters are the smallest hypervelocity impact structures. They are bowl-shaped in planform and have sharp rims, and over-turned stratigraphy in the ejecta blanket. The morphology of these structures is controlled mainly by the strength of the substrate. Simple craters have the largest depth/diameter ratio \(d = D_{\text{r}}\), where \(d \approx 0.2 - 0.33D\).

5.3.2 Complex Craters

As impact energy increases, the target loses strength and there is a collapse of the walls and an uplift of the crater interior. Complex craters are consequently formed in a regime where gravity is the dominant factor. Uplift is also associated with unloading associated with the rarefaction wave. Complex craters are characterized by terraced walls that are the surface manifestation of subsurface faults, and they also show central peaks that contain material brought up from deep beneath the crater. Complex craters have smaller \(d/D\) ratios because the increased importance of gravity collapse results in more uplift of floor. In these craters \(d \approx 0.1D\).

The transition from simple to complex craters depends on gravity according to

\[
D \propto g^{-1}.
\]

The transition from simple to complex craters occurs at a higher diameter on planets with higher gravity. For example, the transition occurs at a diameter of 3 km on the Earth and about 18 km on the Moon.

5.3.3 Basins

The next morphological step up in the energy scale corresponds to basins, which transition from structures that have a central peak and rings (aka peak-ring basins) to no central peak and multiple rings (aka multi-ring basins). In the transition to a peak ring basin the central peak collapses to form a small ring that increases with increasing impact kinetic energy. In a multi-ring basin the number of rings scales with the impact
energy and the mechanical properties of the near-surface layer into which the impact occurred.

Basins have even smaller $d/D$ ratios than complex craters due to more central uplift. There have been arguments about whether basin penetration depth scales proportionally or non-proportionally with respect to diameter in comparison to smaller craters. The majority of researchers believe that the scaling is non-proportional with increasing basin size, because if it was proportional then we should be sampling well into the mantle of the largest impacts on some planets and there is no geochemical evidence to suggest that this is the case. As for simple-complex craters, the diameter of the complex crater to basin transition also depends on gravity, but the morphology of large multi-ring basins cannot be attributed to gravity alone.

5.3.4 Crater Size

How big will a crater be for a given impact? The answer depends on the material properties of the target, the velocity and angle of impact, whether the impact occurs in the strength or gravity regime, and other things. A simplified scaling relationship for crater radius $R$ can be written

$$R \left( \frac{\rho_t}{m_p} \right)^{1/3} = f \left( \frac{Y}{\rho_p U_p^2} \frac{g r_p}{U_p^2} \right),$$

where $\rho_t$ and $Y$ are the density and yield stress of the target, $m_p$, $U_p$, $\rho_p$ and $r_p$ are the mass, velocity, density and radius of the impactor, and $g$ is gravitational acceleration.

5.4 Crater Statistics and Relative Ages of Planetary Surfaces

5.4.1 Impactor Populations

Craters can be used as relative age markers by counting their numbers and size distributions on a planetary surface. To use craters in the dating of surfaces one must consider the rate of crater production and obliteration. Estimates of crater production require an estimate of the flux of impacting bodies and the effect of an atmosphere, if relevant, in ablating incoming projectiles. Obliteration represents the processes that remove craters from surfaces and may include erosion, lava flows, overlapping impacts or ejecta from nearby impacts, subduction, and viscous relaxation. To assess the crater population, the principal piece of information is the measured number of craters as a function of diameter over planetary surfaces over all ages.

5.4.2 Crater Counting

By counting the number of craters on a planetary surface it is possible to derive an estimate of relative age. That is, one can tell how old one surface is compared to another on the same planet. The determination of an absolute age would require knowledge of the flux of impactors (asteroids, meteoroids and comets), which are generally poorly known.
The principal piece of information in crater counting is the number of craters per unit area as a function of diameter. **Cumulative frequency distributions** are usually employed, where \( N_{\text{cum}} \) is the number of craters per unit area equal to or greater than a given diameter. The number of craters of a given diameter that would be expected is a product of the flux rate and surface area. It has been shown in practice that the cumulative crater distributions closely approximates a power law function of diameter:

\[
N_{\text{cum}} = \alpha D^{-\beta}
\]  

(50)

where \( \beta=1.8 \) for post-maria craters on the Moon. Equation (50) indicates that the best constraints on relative age come from small craters, which are more abundant that large craters.

The fact that \( \beta \approx 2 \) is interesting because when \( \beta=2 \), \( \alpha \) is dimensionless because \( N_{\text{cum}} \) is the number of craters per unit area. This relationship corresponds to a situation where craters follow a fractal distribution and are therefore scale-invariant. A population of craters with a power law near two might arise from a formation process in which there is no length scale, or perhaps a series of independent processes that are so complex that no single length scale dominates.

The integral of the crater production rate over the age of the surface is a special theoretical crater population known as the production population. A production population is the size-frequency distribution of all the craters (excluding secondary impacts – those that occur as ‘offshoots’ of an impact) that have ever formed since craters began accumulating on the surface. The population is theoretical in the sense that it ignores all obliteration processes, and is hence never observed, though crater populations on lightly cratered surfaces may approximate a production population. The production population is a useful concept for the study of the evolution of crater populations. Such studies usually begin with an assumed or inferred production population and then postulate a model of crater obliteration that results in a predicted crater population that matches the observations.

If cratering continues the case will arise that the number of craters cannot increase despite the fact that impacts are still occurring. This is because there are so many craters on the surface that is in a state of saturation, and every new impact will remove existing craters. A surface that has been saturated with craters is referred to as an equilibrium surface. On equilibrium surfaces only a lower limit of the relative age can be obtained.

On Earth there are very few impact craters because there are so many processes that remove craters such as erosion and subduction. The best studied planet from the point of view of crater populations is the Moon. Crater counting has shown that the rate of cratering has declined significantly over the past 4.5 BY. The Moon is the one planetary body where absolute ages of surfaces have been derived in part from crater counts. This is because we have absolute ages of some areas from lunar samples, and this permits the relative age curve to be ‘anchored’. Lunar crater counts indicate an exponential drop in the flux at about 4 BY. Highland terrains have ages greater than 4 BY and the maria mostly have ages that span 3.8-4.0 BY. There are very few surfaces younger than 3 BY, and these are associated with localized volcanic events.

**5.5 Case Studies**
5.5.1 Meteor Crater

Meteor Crater, first recognized as an impact structure by Eugene Shoemaker, formed 50,000 years ago by the passage of a 40-50 m iron projectile through Earth’s atmosphere with an impact velocity of 11 km s$^{-1}$. The crater is 1.2 km (0.8 mi.) across and 200 m deep. The kinetic energy released is estimated to have been 15-20 megatons (1000 Hiroshimas or Nagasakis, which were about 15 kilotons). Because of its excellent preservation state and easy accessibility which enabled many years of detailed analysis, data from Meteor Crater has provided the basis for much of our current understanding of planetary impacts.

5.5.2 Tunguska

The Tunguska impact event occurred in 1908 in a remote region of Siberia. The impact produced an area of devastation 1000 km$^2$. While there was much evidence of massive destruction, no crater was produced. After many years of study, made difficult because of inaccessibility to the site, it was determined that the event was probably a stony meteor 80 m across that disintegrated in the atmosphere shortly before impact. The impact velocity was estimated to be about 22 km s$^{-1}$.

5.5.3 Cretaceous-Tertiary Event

In 1980 there was a paper published that changed the way that many people thought about the evolution of the Earth. Before this paper it was a strongly held belief among geologists that the Earth evolved through a series of gradual changes. This included the formation of landforms and the evolution of life. For example, mass extinctions were attributed to causes like gradual changes in climate. The paper by Walter Alvarez and colleagues in Science magazine challenged this way of thinking.

Geologist Walter Alvarez of Berkeley had been analyzing clays in a thin layer at the Cretaceous-Tertiary (K-T) boundary in an outcrop in Gubbio, Italy. Chemical analysis of sediments in the immediate vicinity of the boundary layer showed them to be enriched in the element iridium by a factor of $\approx 10$ over what would be expected. The only rocks that contain iridium in that concentration are iron meteorites. Alvarez and colleagues (which included his father Luis, the Nobel Prize-winning physicist) boldly hypothesized that the K-T extinction, which wiped out nearly half the species on Earth, was a consequence of the impact of an asteroid $\approx 10$ km in diameter. Others argued that elements like iridium were also enriched in volcanic rocks (though none were present in the concentration observed at the K-T boundary) and they alternatively hypothesized massive volcanism at the time of the boundary. (Interestingly, the massive Deccan traps in India have the same age as the K-T.)

Alvarez’s paper inspired many workers to closely analyze K-T boundary sediments all over the Earth, including sediment cores from the ocean. The results were astounding. Iridium and other similar minerals were everywhere enriched in these layers. Results indicated that the K-T in many areas also displayed a distinctive ash layer. indicated that there were wildfires large enough in scale to preserve a global signature. Ash was apparently carried up in the atmosphere and distributed over the Earth. It was hypothesized that so much ash was thrown into the atmosphere that it blocked sunlight for several months to
a year. The estimate is that the atmosphere at that time contained \( \approx 100 \) times as much dust lofted by any volcano in the past 2 centuries.) Calculations based on ash abundance and distribution suggested that the dusty atmosphere implied from the cores could have depressed global temperatures from a few degrees to up to 10\(^\circ\), which would sever the food chain in many places. If this occurred, it was argued, many life forms, including the dinosaurs, could have succumbed.

Further evidence came to light in the form of shocked quartz grains, that were only known to be produced at impact pressures. The propagation of a shock wave through certain minerals produces changes in crystal structure that are observable in microscopic thin sections and are diagnostic not only of the fact that shock has occurred but also of the magnitude. This evidence pointed strongly toward an impact, though the actual crater had yet to be found. The problem was that impacts are not well preserved on Earth. Odds were that the impact, if that was indeed the explanation for the iridium, carbon, and shock features, occurred in the ocean, where the signature would be difficult to detect. There was even a finite probability that the crater had subducted back into the mantle. On the continents erosion, sedimentation and vegetation makes identification of impact structures a challenge. There are only about 140 craters on Earth, and no known crater close to 65 MY in age that was big enough to have produced the kind of "event" implied by the K-T evidence. Then in 1992 the smoking gun was found. Scientists doing field work in Haiti in the Caribbean looking at K-T sediments and found abundant glass spherules representing melted sediments. This indicated that this area was close to 'ground zero'. So the search was on. After some effort scientists obtained marine gravity maps and well cores in the Gulf of Mexico. These data had existed some time but were mostly proprietary because the measurements were made by oil companies that were prospecting. Alan Hildebrand, then a graduate student at the University of Arizona, found evidence in seafloor gravity of a buried circular structure just off the Yucatan Peninsula of Mexico. Core samples verified that indeed the structure was produced in an impact event. The crater is called \textbf{Chicxulub}, and with a diameter of 180 km it is the largest known terrestrial impact crater. Energy scaling calculations suggest that the crater was produced by an \( \approx 10 \)-km diameter projectile. Its age was determined to be 64.98 MY, which placed it precisely at the K-T boundary.

There is now no doubt of a catastrophic impact at the end of the Cretaceous. But whether or not the impact was directly responsible for the mass extinction is still hotly debated. In any case, the Chixulub event provided the impetus to better understand the potential links between impacts and biological processes.

5.5.5 South Pole-Aitken Basin

South Pole-Aitken is a 2200-km diameter impact basin on the far side of the Moon. It could have been produced by a 215-km diameter projectile with an impact velocity of 20 km s\(^{-1}\) or a 900 km diameter projectile at 2.4 km s\(^{-1}\) (lunar escape velocity). Despite an age of more than 4.1 BY, the basin retains over 8 km of topographic relief. Spectral analysis of the basin floor reveals materials with mafic composition that indicate that the impact penetrated into the deep upper crust (but not well into the mantle as would be required by proportional scaling of \( d/D \) ratios). The azimuthal heterogeneity of the
structure suggests a slow, oblique impact.
Problems

5-1. The Galilean satellite Callisto is in synchronous rotation about Jupiter and therefore keeps the same hemispheres at the apex (leading hemisphere) and antapex (trailing hemisphere) of orbital motion. The weighted mean impact velocities at the apex and antapex of Callisto are estimated to be $v_{i-\text{apex}}=23 \text{ km s}^{-1}$ and $v_{i-\text{antapex}}=13 \text{ km s}^{-1}$. Assume the impact of a pure $H_2O$ ice comet into a pure $H_2O$ ice surface. Use the impedance matching technique and the attached Hugoniot curve in Figure ?? to estimate the ratio of the average peak pressures for impacts into the leading and trailing hemisphere. Also estimate the particle velocities of the projectile and target at the time of initial impact.

5-2. Figure ?? plots amount of melt and vapor produced by basaltic impactors of various sizes and impact velocities into a granitic target (i.e. terrestrial continental crust). The plot indicates that for given impactor and target types and fixed impact kinetic energy, slow large bodies produce more melt than fast small bodies. However, the Rankine-Hugoniot equations indicate that the greater the impact velocity the greater the peak pressure and therefore the greater amount of waste heat trapped in the area between the Rayleigh line and the release adiabat (or Hugoniot). Why, then, does a small fast body produce less melt than an energetically equivalent large slow body?

5-3. Altimetry and radar imaging data from the Magellan spacecraft indicates that the topography of impact craters on Venus is virtually unrelaxed. (a) Assume that the Venus lithosphere can be approximated by a halfspace of uniform viscosity. Also assumed that most 10-km diameter craters have shallowed by approximately 10 same size. Given a surface density of 300 kg m$^{-3}$ and the gravitational acceleration of the Venus surface of 8.87 m s$^{-2}$, estimate the viscosity of lithosphere and the relaxation times of for craters with ages of 0.01, .1, 1 and 4 GA. Calculate the viscosity in units of Pa-s (kg m$^{-1}$ s$^{-1}$) and relaxation times in years. (b) A slightly more realistic structure for the Venusian lithosphere may be a halfspace in which viscosity decreases exponentially with depth. Assume that the depth of 10-km diameter craters is much less than the scale depth of the lithosphere. Would the relaxation times determined in (a) increase, decrease or stay approximately the same? What about for a 100-km diameter crater whose depth is comparable to the scale depth of the lithosphere?

References

