Forces Driving and Resisting Orogeny
GEOL5690, Tectonic History of the Western U.S.

The magnitude and origin of the forces that create mountain belts and deform the crust have long been of interest in geoscience, but a quantitative understanding of these forces is continuing to develop. Because mountains are not accelerating, the net force on any piece of lithosphere is zero. However, the force balance is one between forces driving deformation and those that resist deformation much as the balance of a book on a table is between a force making the book want to fall (gravity) and a force preventing the fall (normal stresses from the table). We will first consider the forces resisting deformation and then consider the forces driving deformation.

Strength of the lithosphere

References:

Brace and Kohlstedt, Limits on lithospheric stress imposed by laboratory experiments, *J. Geophys. Res.*, *85*, 6248-6252, 1980.

Turcotte and Schubert, *Geodynamics*, chapter 7 (and 8 to some degree).

Rheological layering of the lithosphere.

There appear to be two fundamental modes of deformation of rocks: brittle failure and ductile flow. The former is related to earthquakes, which represent such failure; the latter should not generate earthquakes. The boundary between the two regimes is termed the brittle-ductile transition. Although it is often represented as a major discontinuity within the Earth, it is in fact most probably a broad zone within which a number of different mechanisms of rock deformation occur. Somewhere within this zone should be the base of that part of the lithosphere that generates earthquakes (the seismogenic layer).

Our understanding of deformation within the Earth is largely based upon laboratory experiments and, to a lesser degree, observations of naturally deformed rocks. Lab experiments are necessarily over a much shorter time than actual rock deformation and are of a smaller scale; these substantial differences suggest that some caution must be exercised when employing these observations. (One possible example of the lab being too small has been suggested for the frictional behavior of faults during an earthquake. A theoretical mode of motion on a fault would cause the effective frictional resistance of the fault to go to zero during rupture. This mode is observed in experiments with foam rubber blocks but not with rocks. It turns out that to look for this kind of deformation in rocks would require an apparatus with blocks of rock several meters on a side, an experiment not yet conducted).

Brittle upper crust

An examination of rock properties reveals a large variation in rock strength with lithology. This would suggest that lithology would be the most important factor in determining the strength of rocks. However, experiments have revealed that this variation is almost entirely restricted to the effort to break a rock; once broken, the frictional behavior of a fault remains pretty uniform from lithology to lithology. What is more, this behavior depends almost entirely on the effective pressure on the fault, with little dependence on temperature or strain rate. This observation is termed Byerlee's Law and is expressed as a relationship of the shear stress and effective normal stress on a fault. This can be reformed as a relationship between the principal stresses:

$$\begin{array}{ll} \bar{\sigma}_1 \cong 5\bar{\sigma}_3 & \bar{\sigma}_3 < 110 \text{ MPa} \\ \bar{\sigma}_1 \cong 3.1\bar{\sigma}_3 + 210 & \bar{\sigma}_3 > 110 \text{ MPa} \end{array} \quad (1)$$

These stresses are effective stresses, meaning that the pore pressure has been subtracted from them ($\bar{\sigma}_n = \sigma_n - P_p$, where P_p is the pore pressure and n is between 1 and 3).

Eqn. (1) can be rewritten for compression (where σ_3 is the vertical stress, $\rho g z$) as

$$\begin{aligned}\bar{\sigma}_1 - \bar{\sigma}_3 &\cong 4\left(\rho - \lambda\rho_w\right)gz && \left(\rho - \lambda\rho_w\right)gz < 110 \text{ MPa} \\ \bar{\sigma}_1 - \bar{\sigma}_3 &\cong 2.1\left(\rho - \lambda\rho_w\right)gz + 210 && \left(\rho - \lambda\rho_w\right)gz > 110 \text{ MPa}\end{aligned} \qquad (2)$$

where λ is the fraction of hydrostatic pore pressure and ρ_w is the density of the pore fluid. A similar manipulation can be done for the extensional case by setting σ_1 to the weight of the rock above.

This makes pore pressure one large uncertainty in determining the strength of the brittle part of the lithosphere. Observations down to a couple of kilometers depth in continents suggest that a pore pressure between 0 (dry) and hydrostatic ($P_p = \rho_w gz$) are probably about correct for much continental crust. If you recall, there is some evidence in accretionary wedges and possibly fold-and-thrust belts for higher pore pressures in these environments.

Ductile lower lithosphere

At higher pressures, rocks begin to flow ductilely. In this environment, there is no longer a maximum strength per se. Instead, there is a certain stress that will cause flow at a certain rate, usually expressed as a strain rate $\dot{\varepsilon}$. Drop below that stress, and deformation is at a slower rate. Increase the stress, and deformation will go faster. Unlike frictional slip, the relations here depend on the rock type involved, the temperature, and the stress difference σ_1 - σ_3. The general form of this is a power-law flow:

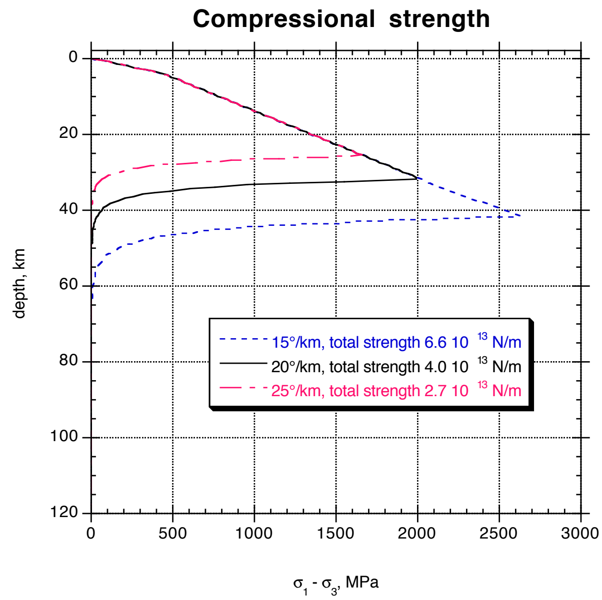
$$\dot{\varepsilon} = C\left(\sigma_1 - \sigma_3\right)^n e^{-E_a/RT} \qquad (3)$$

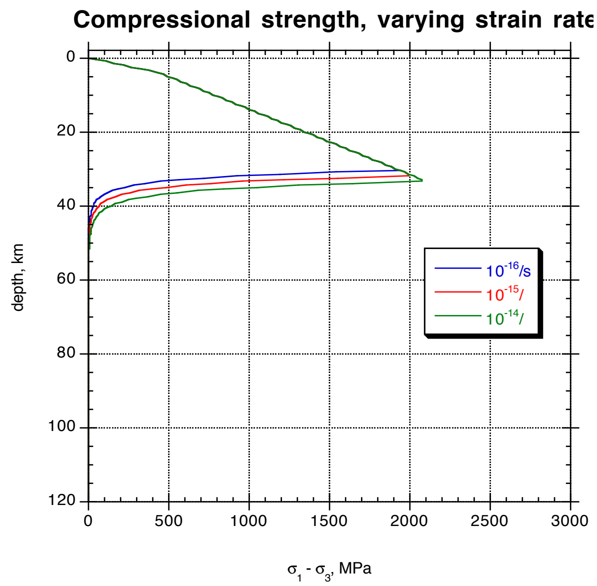
For olivine (from Goetze, 1978), C is about 7 x 10^4, n is 3, E_a (the activation energy) is 0.52 MJ/mol, and R (the ideal (or universal) gas constant) is 8.317 J K^{-1} $mole^{-1}$. T is in Kelvin. (For dry quartzite, E_a is 0.19 MJ/mol, n is 3, and C is about 5 x 10^{-6} (*not* 10^6 as shown in Brace and Kohlstedt).

For purposes of strength of the lithosphere, we may rewrite (3) in terms of a stress difference as a function of temperature:

$$\sigma_1 - \sigma_3 = \left[\frac{\dot{\varepsilon} e^{E_a/RT}}{C}\right]^{1/n} \qquad (4)$$

If we assume a linear change of temperature with depth, $T = A + bz$, then we find a fairly rapid decrease in strength with depth for a given strain rate. (Goetze's work indicates a lower differential stress than this equation at high differential stresses; we will use this form for illustrative purposes, but you'd want to use proper values for any research). We can combine (2) and (4), taking the minimum of the two, to get a feel for the strength of the lithosphere:



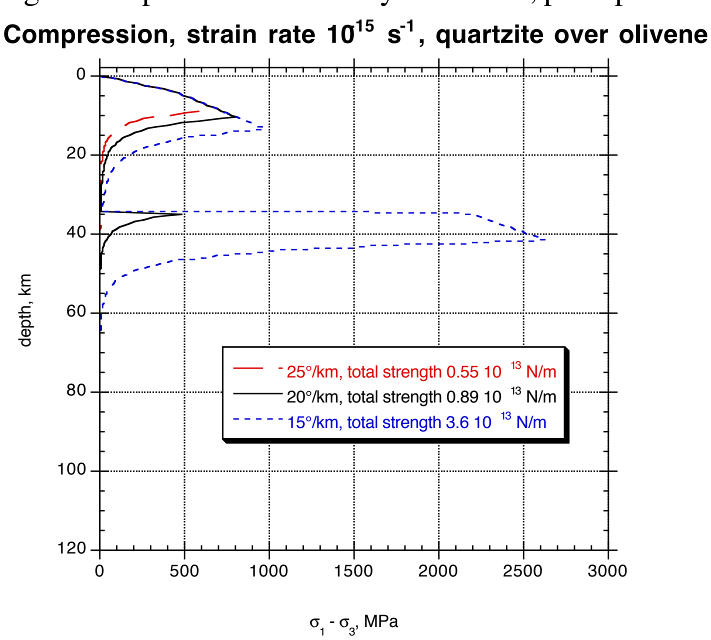


Note that these curves are for a pore pressure of zero (dry rock) in compression.

The upper part of the lithosphere fails as a brittle/elastic medium, the lower part by creep, which fails aseismically. Thus the boundary between the two is often termed the brittle-ductile transition and is often considered the base of the seismogenic part of the crust. Note that this transition moves upward in response to higher temperatures. This variation in the brittle-ductile transition has been observed, at least crudely, around the globe in continental areas.

The figure on the left shows the variation in strength for a strain rate of 10^{-15} s^{-1}; that on the right shows a variation for several strain rates with a fixed temperature increase of 20°C/km. Note that the change in strain rate over two orders of magnitude is fairly small compared to the change caused by temperature variations. The area under each curve is the total strength of the lithosphere, meaning that deviatoric horizontal forces equal to that area would have to be applied to get the entire lithosphere to strain at the rate assumed. The strengths for the three curves on the left are listed in the caption.

The curves above are for a monolithologic lithosphere made entirely of olivine, perhaps a good first approximation for the oceanic lithosphere. We know that the continental lithosphere is stratified, with more quartz-rich rocks above the Moho and ultramafic rocks below. Thus we would like to see the effect of this stratification, if any. Using the flow laws for quartzite above a depth of 35 km and that for olivine below, we can redo the left figure above for a stratified lithosphere (at right).



Because quartzite is quite weak at much lower temperatures than olivine, the net effect is that lithosphere should be considerably weaker with quartz present in the crust than when it is not. Once again we can integrate the curves to get the gross strength of the lithosphere and we find it has decreased by 50-80% for the case above.

It is important to remember that these are extrapolations of laboratory measurements; the constants used, not to mention the stress exponent *n*, are subject to considerable uncertainty. Additionally, the brittle-ductile transition itself probably deforms by neither of the mechanisms assumed above but instead by other modes of failure, making the lithosphere somewhat weaker at these depths. The general effects of temperature in weakening the lithosphere, in strain rate mildly strengthening the lithosphere, and in quartz weakening the lithosphere are probably solid results. Understanding these variations of strength allows us to consider the role of strength and force in continental deformation.

A lingering puzzle remains the absence of earthquakes in the mantle in many continental areas with seismicity extending deep in the crust. The flow and failure laws above would seem to require brittle failure of the mantle beneath crust that is failing through faulting. Two possibilities seem to exist: one is that the upper continental mantle is quite weak, probably because of the presence of water within the olivine (e.g., J. Jackson, *GSA Today*, September, 2002), the other that a non-brittle mechanism comes into play, probably at a lower differential stress than brittle failure. This problem remains unsolved as weak-mantle advocates must explain fairly stiff continental lithosphere while strong-mantle advocates need to address the relative absence of earthquakes in places like the western U.S. (Colorado Plateau, western Sierra Nevada; Wong, I. G., and D. S. Chapman, Deep intraplate earthquakes in the Western United States and their relationship to lithospheric temperatures, *Bulletin of the Seismological Society of America. 80. (3)*, 589-599, 1990).

Forces deforming the lithosphere

Frequently the literature describes continental deformation in terms such as due to continental collision, or from plume emplacement, or post-orogenic collapse, etc. These terms all too frequently hide the fundamental physics causing the deformation. There are no more than five types of forces capable of deforming the lithosphere: horizontal (edge) normal stresses, edge shear stresses, basal normal stresses, basal shear stresses, and internal body forces. We shall consider these in terms of determining the cause of extension in the Basin and Range, somewhat in parallel with the discussion in the *Sonder and Jones*(1999) paper you had available to read.

1) Edge forces

The application of forces to the edge of an orogen became the most widely employed force once plate tectonics was accepted. The idea is that forces applied to a plate by other orogens, spreading centers, and subduction zones are transmitted through the strong plate to the orogen in question. These forces are a combination of a normal stress and a shear stress, depending on the orientation of the plate edge.

Because such forces are applied to the whole orogen, it must be variations in strength that determine the location of any deformation. One such variation is of course the change in rheology that occurs crossing from oceanic (olivine-rich) to continental (quartz-rich) crust. If there are no other variations, we would then find that the edges of continents would deform the most (the edges, in deforming, would not transmit as much stress to the interior). Numerical simulations of this show that for normal sorts of continental rheologies, we would expect deformation to be quite intense at the margin, decreasing to near unobservable some few hundred kilometers inland (less for shear stresses, more for normal stresses). By and large this confirms the most general observation of continental tectonics, namely that it tends to be localized at continental edges.

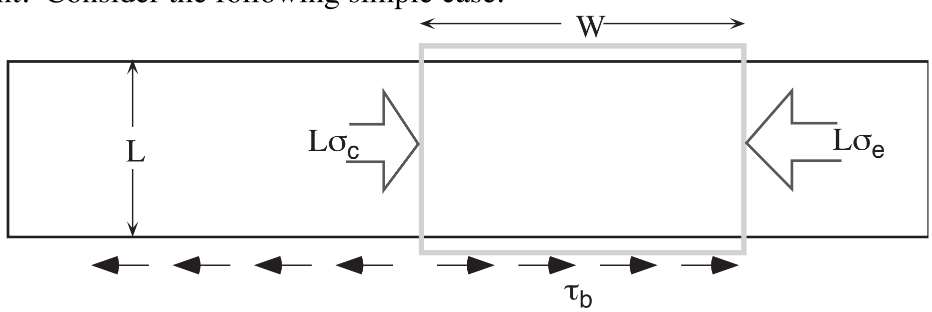
In many cases, there are inconsistencies. Most glaring is the presence of normal faulting in the Altiplano in the Andes and in the Tibetan Plateau; both regions are surrounded by areas undergoing contraction. There is no way for edge forces by themselves to produce such a situation; the sense of deformation should be the same all the way across the orogen. A similar problem might exist for the initiation of Basin and Range faulting in eastern Nevada and northern Utah, with contraction still proceeding at the plate edge to the west and late contraction still ongoing in the Laramide mountain ranges to the east.

More subtle problems can exist. For the Laramide orogeny, the Colorado Plateau remained relatively undeformed. If the Laramide compression was caused by forces on the edge of North America, then why did the Plateau remain undeformed? Once we have assumed that the forces driving deformation come from the edge, the only possibility is that the Plateau was stronger than the area that failed. As we have seen above, the main effects on rheology are temperature and composition. The hypothesis that the Laramide was caused solely by edge forces requires that the Plateau either be more mafic than the area to the northeast, or was colder, or some combination of both. These predictions are testable in the field through the use of geothermometers and studies of the modern composition of the crust in both areas.

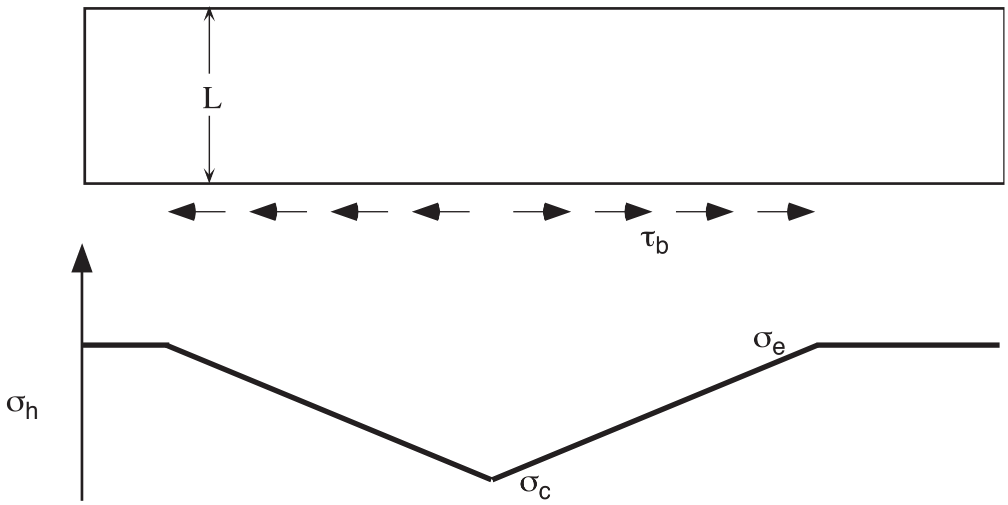
2) Basal shear stresses

The second major class of stresses are those applied at the base of the lithosphere. In this case there is a substantial difference in the behavior of normal and shear stresses, so we shall consider them separately. Shear stresses are most plausibly applied by motion of the lithosphere relative to some deeper level of the mantle; it has been suggested that shear caused by motion of a plate over "fixed" asthenosphere is a possible way of generating a basal shear. More intense shear stresses can be generated by direct contact of two plates on a subhorizontal boundary, such as might exist above a flat subducting slab.

Unlike the edge stresses, the forces on the lithosphere must vary even if the basal shear is constant. Consider the following simple case:



Assume some mean normal stress in the center of the region shown is σ_c and that the basal shear stress is τ_b. To keep the gray outlined box from flying off the view to the right, $L\sigma_c + W\tau_b = L\sigma_e$. As W increases, so must σ_e. (The moments of the shear stresses will dictate how the normal stresses are distributed through the lithosphere, a minor point for our present discussion). Thus if we plot the mean horizontal normal stresses as a function of distance (compression is positive), we find that where the shear stresses diverge, the horizontal normal stress σ_h is at a minimum:

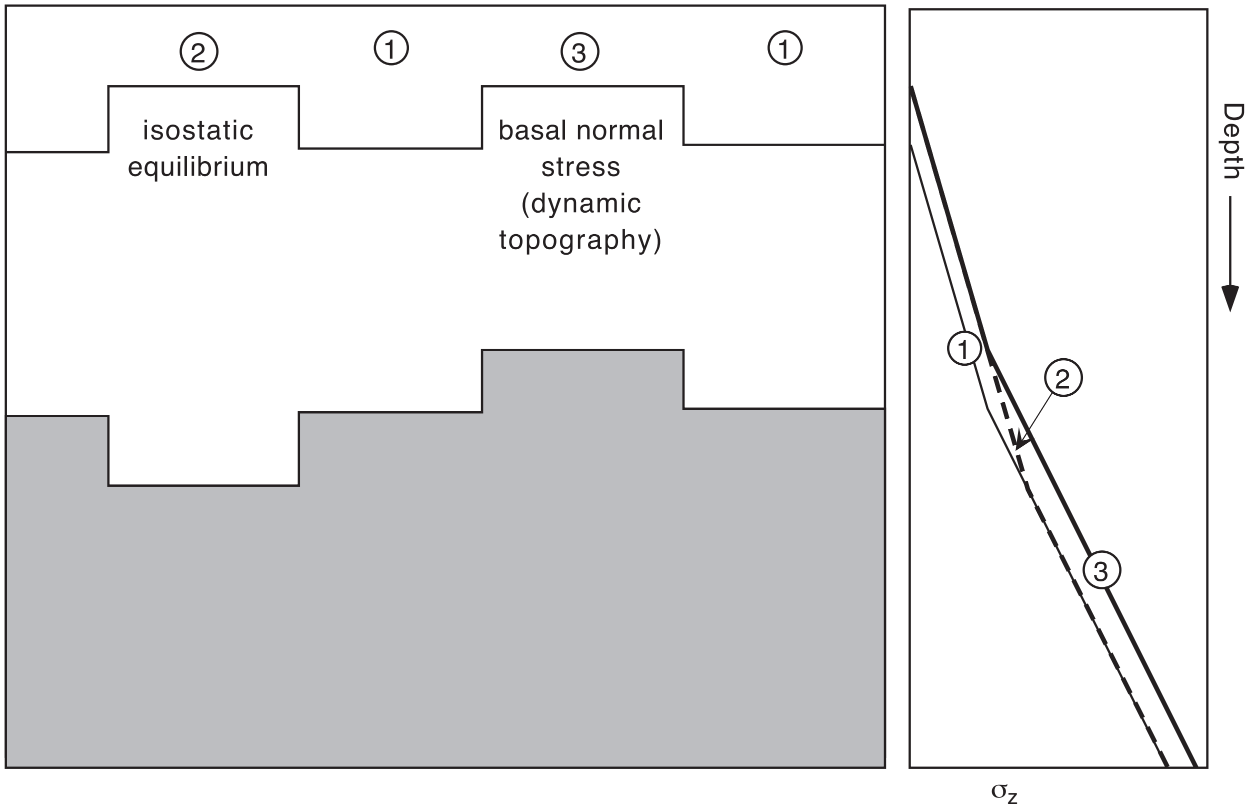


This variation could be enough to create extensional deviatoric stresses in the center of the diverging shear stresses and compressional deviatoric to the sides. Alternatively, if one only considered half of the system (where the shear is the same), the deviatoric normal stresses would increase in the direction of the shear stress. This could be a way of transmitting forces into the interior of an orogen, bypassing the margin.

Basal shear stresses cannot be dismissed very simply; it is unlikely that motion of the deeper mantle exactly mimics the surface tectonics, and any discrepancy should produce a shear stress. For the Laramide, the increase of deformation away from the margin makes basal shear stresses quite attractive. However, observation of variations of stress in the lithosphere, combined with experimental studies of the strength of rocks in the upper mantle and numerical simulations of global tectonics suggest that *usually* the shear is small at the base of the lithosphere.

3) Basal normal stresses and internal body forces

Basal normal stresses do not themselves require a balance from the horizontal normal stresses in the lithosphere; they are usually balanced by the increased (decreased) weight of material above the increased (decreased) basal normal stress (or by shear stresses on vertical planes, which is largely an elastic plate type of response to this force). For an upward directed force, the main result is that the region above is uplifted. Thus throughout the uplifted lithospheric column, the weight of the material above a given depth is more than for the surrounding area (compare columns 1 and 3, below).



On the left side we have a crosssection of two uplifts, one isostatically supported, one supported by basal normal stresses. Mantle is gray, crust is white. Note on the right side of the figure the plot of vertical normal stress vs. depth. Recall that the condition for isostasy is that the vertical normal stress must be constant horizontally at some depth of compensation. Clearly the vertical normal stress for column 3 remains higher than that of columns 1 or 2, which are in isostatic equilibrium. Clearly in the lower lithosphere, where material strength is low (see above), the increase in σ_z would rapidly cause thinning which would in turn cause subsidence of the region above unless material flowed up rapidly from below. For a basal normal stress to remain operative for a geologically significant time, material must be flowing below to maintain the normal stress. For this reason the basal normal stress situation is largely equivalent to so-called "dynamic topography," which is elevation maintained by flow of material in the mantle rather like the waves and holes on a river are maintained by the motion of the water. As you should recall, most topography appears to be in isostatic equilibrium from observations of the free-air gravity anomaly. The small anomalies that do exist, combined with the difficulties of calculating free-air anomalies in continental regions, permit some amount of dynamic topography to be present, but the amount is hotly disputed and no gravity measurements on continents *require* the presence of dynamic topography. Most of the arguments come from theoretical models and not field observations; at present, dynamic topography appears to be limited to elevations of less than about 1 km.

The principal effect on forces of a basal normal stress is the change of the vertical normal stress throughout the lithosphere. This is quite similar to the effect of variations in the density structure of the lithosphere. Compare, for instance, σ_z for columns 2 and 3. They are identical until you reach the compensating mass (here shown as a crustal root). If, say, the region of column 1 had an isotropic stress state (i.e., $\sigma_h = \sigma_z$), then this means that throughout column 3 $\sigma_z > \sigma_h$, which would put column 3 into a deviatoric extensional state of stress. Column 2 is little different, with $\sigma_z > \sigma_h$ down to the base of the crust. The direct measure of the force available to drive deformation is the difference between curves 3 and 1 or 2 and 1; this quantity is frequently referred to as the gravitational potential energy of the lithosphere.

Gravitational potential energy, you might recall from basic physics, for a mass held at a height h is mgh. We can calculate the same quantity through the lithosphere, starting from a compensation depth z_c and going up, we integrate the mass under a unit area as

$$\int_0^{z_c+\varepsilon} \rho(z')gz'\,dz' \tag{5}$$

where we are measuring z' upward from the compensation depth to the surface at elevation ε. (Note that this is actually potential energy per unit area.) This can easily be converted to an integral in depth z below sea level ($z = z_c - z'$):

$$\begin{aligned}\int_0^{z_c+\varepsilon} \rho(z')gz'\,dz' &= -\int_{z_c}^{-\varepsilon} \rho(z)g(z_c - z)\,dz \\ &= gz_c \int_{-\varepsilon}^{z_c} \rho(z)dz - \int_{-\varepsilon}^{z_c} \rho(z)gz\,dz\end{aligned} \tag{6}$$

Now the left hand term on the right hand side is, for areas in isostatic equilibrium, a constant (it is the requirement that the weight above the compensation depth must be constant). We are not interested in the total magnitude of the integral but rather in its deviation from a column that is not deforming, i.e., a column with an isotropic stress state. A column assured of an isotropic stress state is a column of static fluid. In the Earth, this is most closely approximated by a column of asthenosphere, which in turn is pretty close to the column under a mid-ocean ridge. Thus we often end up with something more like

$$\Delta GPE = \int_{H_0}^{z_c} \rho_a gz\,dz - \int_{-\varepsilon}^{z_c} \rho(z)gz\,dz \tag{7}$$

where H_0 is the depth of the top of a column of asthenosphere under air (about 2.4 km from comparison with mid-ocean ridges) and ρ_a is the density of the asthenosphere.
We can also convert this to an integral of the vertical normal stresses by integrating vertical normal stresses by parts, assuming that the vertical normal stress is essentially the weight of the material above it:

$$\begin{aligned}\int_{-\varepsilon}^{z_c} \sigma_z(z)\,dz &= \int_{-\varepsilon}^{z_c}\int_{-\varepsilon}^{z} g\rho(z')\,dz'dz \\ &= \left[zg\int \rho\,dz' \right]_{-\varepsilon}^{z_c} - \int_{-\varepsilon}^{z_c} gz\rho(z)\,dz \\ &= \left[z\sigma_z \right]_{-\varepsilon}^{z_c} - \int_{-\varepsilon}^{z_c} gz\rho(z)\,dz \\ &= z_c \int_{-\varepsilon}^{z_c} g\rho(z)\,dz - \int_{-\varepsilon}^{z_c} gz\rho(z)\,dz = GPE\end{aligned} \tag{8}$$

By vertically integrating the equations of motion, the horizontal equations can be written as

$$\frac{\partial \bar{\tau}_{ij}}{\partial x_j} + \frac{\partial \bar{\tau}_{zz}}{\partial x_i} = \frac{1}{L}\frac{\partial (PE)}{\partial x_i} \tag{9}$$

where $\bar{\tau}$ indicates a vertically averaged deviatoric stress, L is the thickness of the lithosphere, and the subscripts i and j refer only to horizontal coordinates. Thus, stresses are directly related to GPE. In two dimensions (one horizontal and one vertical) the relationship is even simpler; deviatoric stresses (and strain rates if the effective viscosity is constant) are linearly proportional to differences in PE.

From examination of the last figure, we can see that ΔGPE should underestimate the forces available if topography is caused by basal normal stresses and is not compensated by a buoyant mass. Thus we may consider the effect of a basal load to be somewhat greater than that of internal body forces but otherwise equivalent. We will only speak of body forces below but the general characteristics apply to both body forces and basal normal stresses.

The characteristic of body forces is to change the vertical stresses and not the horizontal stresses. Without external (edge) forces, body forces will tend to produce faults paralleling the topography (or, more precisely, contours of ΔGPE); slip on faults in such a system could be quite complex, and some consider the presence of non-parallel slip to be characteristic of orogenic activity dominated by body forces. This is probably a rare condition, as more often there will be far-field constraints either in the form of region horizontal stresses or space restrictions on how the vertical stresses may be relieved through horizontal motion. Unlike any of the previous edge forces, body forces can be quite heterogeneous over distances down to perhaps 100 km or so and so could produce wildly varying structures over similar distances. Hypotheses of body forces predict certain elevations and density structures and can predict a certain variation in deformation different from anything an edge force could produce.

Probably the most obvious application of body forces is in the areas of the Andes and Tibetan Plateau previously mentioned as being inconsistent with edge forces. These areas are high standing and have thick crust; the ΔGPE for these regions is quite high. Lower regions to the sides have lower ΔGPE and thus are dominated by the regional compressional stresses. For the western U.S., some of the differences seen in the northern vs. southern Basin and Range could well reflect the different body forces and the edge forces they interacted with. This is all discussed in *Sonder and Jones*.

A summary of forces in "traditional" models

It is hopefully clear that the forces above may be combined in any ratio. However, it is worth placing them within a more traditional framework for rifting environments.

Passive rift is a rift caused by edge forces acting on a plate. Extension precedes volcanism (which is caused by the decrease in weight as the lithosphere is thinned–so-called decompression melting). Such a system is dominated by edge normal stresses and subject to the predictions above, namely that the deformation be regionally uniform and that deformation was localized by some kind of weakness in the lithosphere.

Active rift is a rift caused by upwelling in the mantle under a region. Upward motion of material in the upwelling causes melt to be generated, producing early volcanism. Tectonism results from the application of basal normal stresses that then drive extension as discussed above. Such a rift should, in addition to the prediction of extension vs. volcanism, overlie an area of the mantle consistent with such upwelling. Rifting could have different orientations in different places but should be centered on topographic swells. Note that basal shear stresses could also be a characteristic of such systems.

Postorogenic collapse is a more informal term indicating that extension is driven by internal body forces, largely from crustal thickening in an earlier orogenic episode. In this case extension should correlate well with areas of high ΔGPE (more or less, areas of thickest crust).

Compressional orogens generally have fewer orogen-scale terms like these, probably because most are well explained by application of far-field (plate) stresses localized near convergent plate boundaries. Nonetheless, similar terms could be envisioned for compressional orogens. "Passive thrusting" might be the standard model of contraction at a contractional plate boundary; "active thrusting" might represent convergence generated above downwelling in the mantle (and might be a mechanism for initiating a subduction zone), while "implosion" might reflect convergence driven by body forces generated by a change in the density structure of the lithosphere and decrease in ΔGPE (e.g., a phase change in the crust, or dynamic thickening of the mantle lithosphere). The Laramide orogeny might possibly be an example of one of the latter two systems.

In all cases, these should be regarded as idealized endmembers. Real world systems, such as the Basin and Range, combine several elements. Recognizing and identifying the effects of each of the force systems described is necessary for properly identifying the cause of tectonism.