

Why subduction zones are curved

L. Mahadevan,¹ R. Bendick,² and Haiyi Liang^{1,3}

Received 8 April 2010; revised 29 July 2010; accepted 10 August 2010; published 11 November 2010.

[1] We give an explanation for the polarity, localization, shape, size, and initiation of subduction zones on Earth. By considering a soft, thin, curved lithospheric cap with either elastic or viscous rheology supported by a thick, nearly incompressible mantle, we find two different characteristic subduction geometries arise depending on boundary conditions: (1) plate boundaries where subduction results primarily from the gravitational body force (free subduction) have characteristic plate lengths and form arc-shaped dimpled segments resulting from the competition between bending and stretching in edge buckling modes of thin spherical shells, and (2) subduction zones due to localized applied loads that push one slab of thin, positively buoyant lithosphere beneath an overriding plate (forced subduction) form localized straight segments, consistent with the deformation of indented spherical shells. Both types of subduction are nonlinear subcritical instabilities, so small perturbations in the mechanical properties of the lithosphere have pronounced effects on subduction initiation and evolution. Yet in both cases, geometric relationships determined by the shape of the Earth itself play the most critical role in controlling the basic morphology and characteristic length scales of subduction zones. **Citation:** Mahadevan, L., R. Bendick, and H. Liang (2010), Why subduction zones are curved, *Tectonics*, 29, TC6002, doi:10.1029/2010TC002720.

1. Introduction

[2] On Earth, where tectonics dominate the dynamics of the lithosphere, oceanic lithosphere is formed at ocean ridges and subsequently cools [Turcotte and Ahern, 1977]. The thickness and stiffness of oceanic tectonic plates therefore vary in time and space, depending on both the cooling rate of the lithosphere and the rate of plate extrusion (spreading rate) at the ridges. However, the curvature of these plates is nearly constant and equal to the inverse of the Earth's radius (~6400 km) everywhere except along subduction zones, where lithosphere bends into the Earth's interior. These zones are long (≤ 5000 km along strike) but narrow (< 200 km

normal to strike) features where a localized bending zone assumes either a characteristic arcuate or straight geometry, often with multiple segments separated by narrow syntaxial cusps (Figure 1).

[3] Despite their importance to global tectonics, there is no consistent explanation for the shape or scaling of these features. Frank's [1968] purely geometric analogy between the shape of arcuate of subduction zones and the bending strip of the dimple on a partly inverted thin spherical shell (such as a ping-pong ball) implies a direct relationship between slab dip and the radius of the subduction arc: the greater the radius of curvature of the arc, the steeper the dip of the downgoing slab. Although this analogy is commonly invoked as the explanation for subduction zone curvature [e.g., Turcotte and Schubert, 2002], observations of our planet do not show such a relationship between radius and dip [Tovish and Schubert, 1978; Bevis, 1986].

[4] Instead, four mechanical conditions control the shape and scaling of subduction zones on Earth: (1) the subducting lithosphere is an incomplete spherical cap with a small but finite thickness (rather than either a complete spherical shell or a flat plate); (2) the shell thickness and cap buoyancy vary due to the cooling of the crust after its extrusion from oceanic ridges; (3) the underlying mantle is thick and nearly incompressible; and (4) subduction can be initiated either by negative buoyancy of oceanic lithosphere (free subduction) or by an applied localized force due to an overriding lithospheric plate (upper plate forcing) [Gurnis et al., 2004].

[5] Previous analytical and numerical studies of subduction mainly use flat sheets that bend into a cylindrical sheet at the trench boundary [e.g., Ribe, 1992, 2001; Schellart, 2004; Morra et al., 2006] omitting both the stiffening due to double curvature and the development of characteristic instabilities at the free or forced edge of doubly curved caps. This geometric stiffening is seen for example, when we fold a slice of pizza azimuthally before munching on it, so that it does not bend radially away from one's mouth. In order to excite the observed geometry of terrestrial subduction, flat plate studies instead focus on mechanical heterogeneity of the slab or interaction of the slab with mantle flow [Morra et al., 2006; Ribe, 2010; Stegman et al., 2010; Funiciello et al., 2003]. The former possibility emphasizes the spatial collocation of syntaxial cusps with ocean islands or other types of buoyant thickened oceanic crust [e.g., Vogt, 1973; Hsui and Youngquist, 1985] or special along-strike variations in buoyancy of the lithosphere [e.g., Hsui and Youngquist, 1985; Stegman et al., 2006]. The latter emphasizes toroidal mantle flow around the edges of foundering slabs [Schellart, 2004; Schellart et al., 2007]. None of these models explain why subduction zones are either straight or arc shaped but do not form other shapes (such as the tightly folded end-member for toroidal-flow models), nor do they account for either the

¹School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA.

²Department of Geosciences, University of Montana, Missoula, Montana, USA.

³Now at Department of Modern Mechanics, University of Science and Technology of China, Hefei, China.

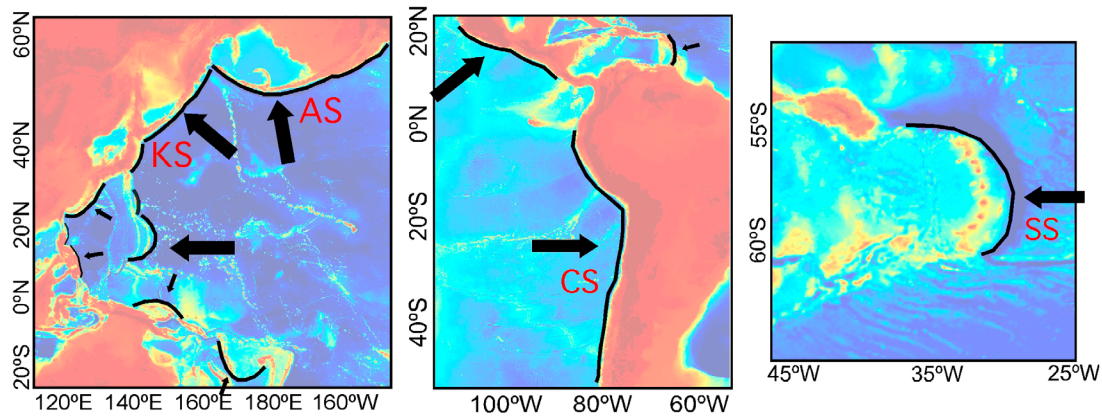


Figure 1. Some examples of terrestrial subduction zones. Global bathymetry and topography of the Pacific margin illustrates multiple arc-shaped subduction segments, separated by syntaxial cusps. Subduction zone trenches are delineated with black lines, and downgoing caps are labeled with arrows showing the convergence direction. Slabs are labeled in all views as AS for Aleutian Slab, KS for Kurile Slab, CS for Chile Slab, and SS for Scotia Slab.

development of multiple adjacent arc segments or the size of oceanic plates. Where geometry has been partially considered for subduction zones, it demonstrably changes the stress field in 2-D subduction of a spherical cap [Tanimoto, 1997, 1998] and generates a dimpled edge in analog 3-D models [Yamaoka *et al.*, 1986]. While adjacent arc segments, such as those along the western Pacific margin, have commonly been modeled as separate individual slabs, compilations of 3-D seismic data (Slab 1.0) clearly show that adjacent slabs are continuous throughout their extent and pass through a tight fold at syntaxes, for example as in the Kurile Slab (KS) in Figure 2, rather than a slab discontinuity.

[6] Here we show that using simple physical notions in a minimal but essential way to account for the double curvature of lithospheric caps, spatially varying buoyancy, and two different subduction mechanisms allows us to qualitatively explain all the observed large-scale morphological features of subduction zones and oceanic plates.

2. Deformation of Curved Elastic Shells

[7] Double curvature of any thin shell leads to geometric stiffening relative to flat plates of the same thickness and rheology because doubly curved objects must be stretched to be deformed even slightly. Thus, while a flat sheet will bend cylindrically as soon as its edge is pulled downward by a body force, a doubly curved cap will resist large deformations until it is sufficiently negatively buoyant or has a sufficient applied load [Landau and Lifshitz, 1986; da Silveira *et al.*, 2000], and then it does so abruptly via a localized nonlinear subcritical instability. Although the rheology of the mantle and the lithosphere are complex functions of space and time, the stability conditions of curved tectonic plates can be illuminated by considering shell-like deformations for the simple limiting cases of elastic subduction and viscous subduction.

[8] One of the simplest solutions for the deformation of a doubly curved elastic cap is a circular indentation around a

point load applied radially [Landau and Lifshitz, 1986]. While small loads lead to a gentle flattening of the sphere, initiation of a dimpled indentation occurs only when the applied load crosses a threshold. This implies, for example, that an intact eggshell can support much larger normal loads than a flat sheet of the same thickness and stiffness because of this geometric contribution. In the case of point loading, the circular geometry of an indentation is determined by localization of deformation along a “bending strip,” the ring around the load where the radii of curvature of the cap change. In the center of the indentation, the shape, and hence, radii of curvature, of the dimple are the same as the original shell, but with inverted polarity. The localization of both bending and stretching within a narrow ring minimizes the energetic costs of deformation [Landau and Lifshitz, 1986]. Equivalently, theory based on force balance leads to the same result.

[9] The same considerations apply to more complicated cases of deflection of an incomplete cap due to a body force, such as gravity or a combination of an applied load and a body force. However, when an incomplete cap such as the lithosphere bends farther into the interior of a sphere it must further change its shape because of simple spatial constraints and does so by faceting, rippling or folding, breaking the initial small-circle radial symmetry of the bending strip in favor of polyhedral shapes [Vaziri and Mahadevan, 2008]. This effect that can be easily observed with any toy ball (auxiliary material); pressing gently on the ball’s surface in the radial direction initially generates a symmetrical circular dimple which then loses stability on further indentation into a set of periodic dimples.¹ While the detailed size of the bending zones, and their growth rates is a function of the rheology of the material of the shell, their existence is a consequence of the geometry of thin shells. For ease and familiarity of exposition, we start by exploring these phe-

¹Auxiliary materials are available in the HTML. doi:10.1029/2010TC002720.

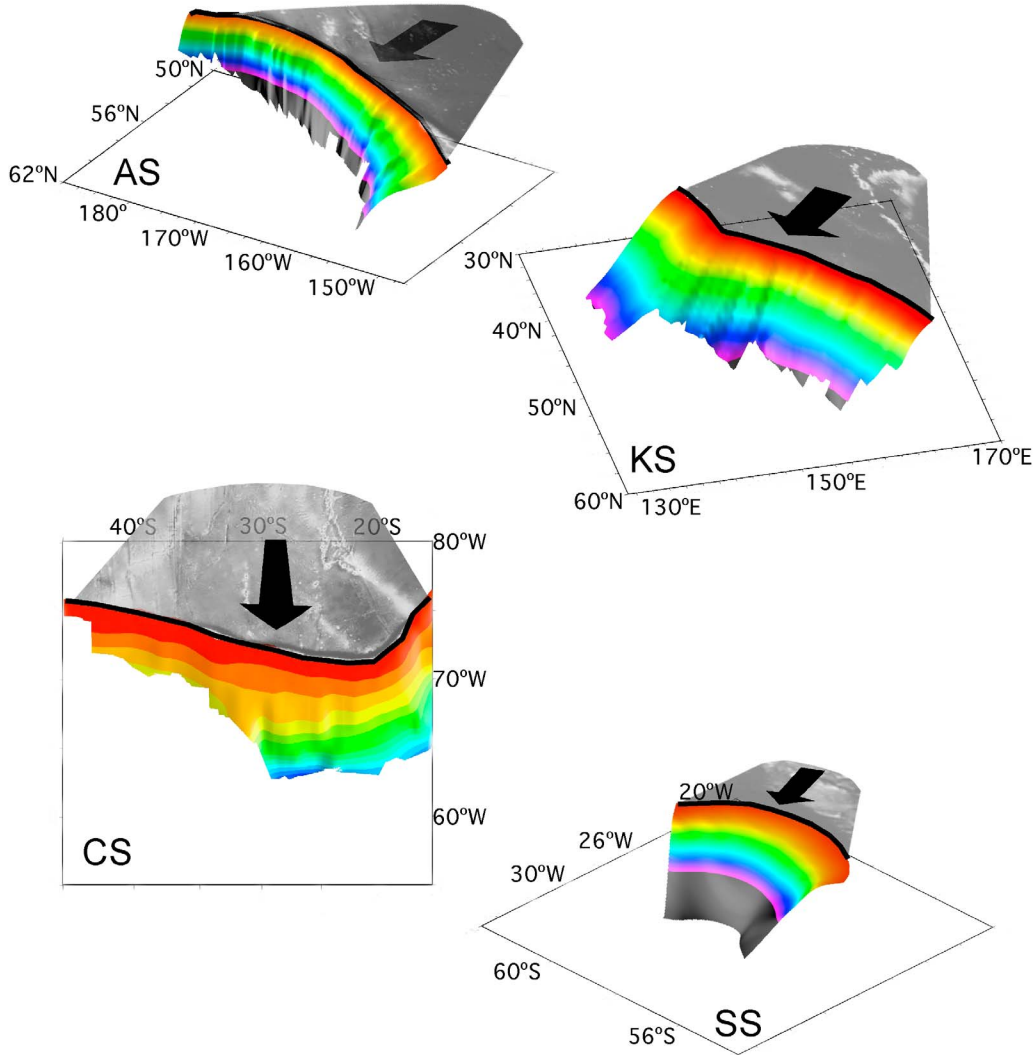


Figure 2. Perspective views of example slabs at depth from Slab 1.0 [Hayes *et al.*, 2009]. The incoming ocean floor bathymetry is shown in grayscale, with the convergence direction provided by the black arrow. As in Figure 1, the trace of the subduction trench is marked by a black line determined by the maximum bathymetric depth. The overriding plate is omitted in these views.

nomena by considering the limit of an elastic material and then use the Stokes-Rayleigh analogy [Rayleigh, 1878] to consider the viscous case before a brief discussion of the more general case.

3. Elastic Scaling

[10] Four different measurable features of subduction zones have characteristic scales related to geometry: the distance between the spreading ridge and the subduction zone (plate length L_p), the width of the arc forebulge (bending strip width, W), the amplitude of the arc forebulge (bending strip amplitude, A), and the distance between arc syntaxes (segment length, L_s) (Figure 3). (Table 1 provides a complete list of parameters and terminology).

[11] The plate length, L , (Figure 3) where subduction is initiated by a body force such as negative buoyancy (free

subduction), must always be between an upper bound, L_{\max} , at which the negative buoyancy of the cap is sufficient to overcome the stiffness of the cap due to both rheology and double curvature and a lower bound, L_{\min} , at which the stress due to negative buoyancy overwhelms the plate causing it to deflect substantially. Of course, L_{\min} must also be greater than the minimum distance from an oceanic ridge required for the lithosphere to have negative buoyancy with respect to the sublithospheric mantle, which itself depends on the ratio of spreading and cooling rates (the Peclet number). Between L_{\max} and L_{\min} , curved, negatively buoyant caps are subcritically unstable, and so they may subduct, but need not. Because of the nonlinearity of the instability which is strongly dependent on the presence of any heterogeneities in material properties, geometry or forcing, cap length is related to the product of cap age and spreading rate (Figure 5), but not strictly proportional to either, as initiation of bending

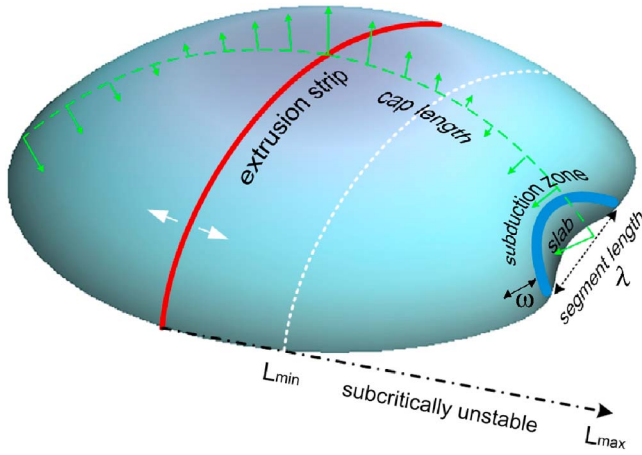


Figure 3. For a thin layer extruded at a strip, L_{\min} is the distance at which a flat sheet of homogeneous rheology would bend inward under a body force and L_{\max} is the distance at which all spherical caps of the same rheology would bend under the same body force. Between these distances, an incomplete spherical cap is subcritically unstable, so may bend but need not, depending on small cap perturbations. Bending that does occur is localized to a bending strip or subduction zone, of width λ and amplitude α , where stretching necessarily also occurs. The portion of the cap that has already passed through the bending strip is the slab. One subduction arc has segment length σ . Green vectors represent the latitude variation of a body force, either difference in pressure or buoyancy, used in the numerical solutions for the dimpling analysis. The extrusion strip represents a spreading ridge, where new lithospheric cap is created.

is also influenced by the relationship among buoyancy, temperature, time (the onset of negative buoyancy requiring the aging of lithosphere of perhaps as little as 10 Ma [Cloos, 1993] to as much as ~50 Ma [Oxburgh and Parmentier, 1977]), and small variations in initial lithospheric thickness. This sensitivity to small perturbations is consistent with the observation that seamount chains often bound subduction arcs; that is, that seamounts and syntaxes are often aligned [Vogt, 1973].

[12] When the cap's free edge crosses the buoyancy threshold, it starts to sink, resisted by but dragging along the positively buoyant part of the cap. To find L_{\min} , we equate gravitational and elastic forces (per unit length) on a naturally flat plate, giving a minimum characteristic plate length of $L_{\min} \sim \left(\frac{Eh^2}{\Delta\rho g}\right)^{1/3}$, above which a flat sheet deflects appreciably as a response to gravity acting on its excess density $\Delta\rho$. Substituting reasonable values for the parameters of $E \sim 0.7 \times 10^{11}$ Pa, $h \sim 25$ km, and $\Delta\rho \sim 0.3 \times 10^3$ kg yields $L \sim 2000$ km for the case of constant-thickness, constant-density lithosphere. For $L < L_{\min}$, subduction may only occur if forced (by a localized boundary load), which, as discussed below, leads to a different, faceted, boundary shape. Indeed, many important characteristics of subduction zones differ between free and forced boundaries [Molnar and Atwater, 1978; England and Wortel, 1980]. There are

currently no arcuate subduction zones bounding plates much smaller than L_{\min} , although some positively buoyant lithosphere could presumably be dragged into the mantle by a large negatively buoyant slab [England and Wortel, 1980], shortening the ridge-trench separation.

[13] We may also calculate L_{\max} for the free subduction mode, by which all plates must subduct even accounting for the initial curvature of the plate. The total mechanical energy of a spherical cap of radius R (the radius of the Earth), size L (distance from the spreading ridge), and thickness h , deflected into the Earth by an amount δ is $U \sim \frac{Eh^3}{R^3}\delta^2LR + Eh\frac{\delta^4}{L^2}LR + \Delta\rho ghLR\delta$, where the first two terms correspond to the strain energy associated with bending and stretching of the cap, and the last corresponds to perturbation to the gravitational potential energy. For large deflections, where δ is not negligible compared to R , such as in fully developed subduction, minimizing this energy with respect to L this leads to the expression $L_{\max} \sim \left[\frac{ER^6}{Eh^2 + \Delta\rho gR^3}\right]^{1/4}$ [Mahadevan et al., 2007]. For $L > L_{\max}$ the third term in the energy equation dominates so that even doubly curved caps must deflect. For Earth, however, $L_{\max} \sim 5 \times 10^4$ km is larger than the Earth's circumference. A comparison of L_{\min} and L_{\max} illustrates the energy contributions neglected in flat plate approximations, namely the bending and stretching energy. Alternatively, the stiffness for small deflections of a flat sheet may be parameterized as $k = \frac{Eh^3}{L^4}$, while for a curved sheet $k = \frac{Eh\delta}{R^2}$.

[14] Once subduction bending has initiated, the double curvature of the cap is also crucial in determining the length of individual bending segments (L) and the width of subduction forebulges (W). In the linear approximation these values are equal. In terrestrial subduction zones, it appears that nonlinear strain localization shortens the width of forebulges after initiation, reducing widths by an order of magnitude, while arc segment lengths do not undergo nonlinear strain and generally retain the linear length scale.

[15] The characteristic scaling, expressed in the width of the bending strip (W), amplitude of the bending strip (A), and segment length (L_s) are shown schematically in Figure 3 and controlled by a trade-off between deformation of the mantle below the lithosphere at long wavelengths and the short-wavelength deformations of curved lithospheric caps, as the slab changes shape to fit into the Earth's interior. Specifically, penetration of a doubly curved cap of lithosphere into the interior of the Earth requires deformation of both the lithosphere and the sublithospheric mantle. The lithosphere must bend at the trench itself, and then must fold to fit into the interior of the Earth, as there is less and less room available the further it penetrates into the planet. The sublithospheric mantle must also deform to conform to the shape of the lithosphere at the forebulge and trench as well as accommodate slab penetration. We do not include the last effect, as it applies similarly to flat and curved plate models and has been thoroughly addressed in detail elsewhere [e.g., Conrad and Lithgow-Bertelloni, 2004]. The mechanics of coupled lithosphere and mantle means that, if the bending strip width and amplitude are large the force required to deform the sublithospheric mantle is very high even though the deformation of the lithosphere is small; if the bending

Table 1. Parameters Used in This Work and Comparison of Their Descriptions in Soft Materials Studies and Tectonics Studies

Term	Soft Materials	Earth Sciences	Example or Description
L	cap length	ridge-trench distance	distance from East Pacific Rise to Tonga Trench
ω	bending strip width	forebulge width	width of the forebulge at the Aleutian Trench
α	bending strip amplitude	forebulge height	height of the forebulge at the Aleutian Trench
λ_l	wavelength of folding instability	subduction segment length	distance from the St. Elias syntaxis to the Kamchatka syntaxis
L_{\max}	upper bound on cap length	maximum possible oceanic plate size	$\sim 10^4$ km (see text)
L_{\min}	lower bound on cap length for body force case	minimum free subduction plate size	~ 2000 km (see text)
E	Young's modulus	Young's modulus	10^{11} Pa for basalt
h	shell thickness	effective lithospheric thickness	25 km for oceanic lithosphere
$\Delta\rho$	excess density	excess density	difference in density between cold basalt and asthenosphere
g	gravitational acceleration	gravitational acceleration	10 m/s^2
R	cap radius	tectonic plate radius	radius of the Earth
δ	cap deflection in the radial direction	depth of slab	depth of Benioff zone or slab position from tomography
U	mechanical energy of spherical cap	mechanical energy of subduction	energy of subduction
h_l	shell thickness	lithospheric thickness	25 km for oceanic lithosphere
E_l	shell Young's modulus	lithospheric Young's modulus	10^{11} Pa for basalt
H_m	foundation thickness	mantle thickness	600–700 km to density discontinuity
E_m	foundation Young's modulus	mantle Young's modulus	much less than Young's modulus of lithosphere
K	foundation stiffness	mantle stiffness	resistance of the mantle to deformation
k	cap or shell stiffness	lithospheric stiffness	resistance of the lithosphere to deformation
B	shell bending stiffness	curved lithosphere bending stiffness	resistance of the lithosphere to deformation due to the cost of bending a curved shell
S	shell stretching stiffness	curved lithosphere stretching stiffness	resistance of the lithosphere to deformation due to the cost of stretching a curved shell
μ	viscosity	viscosity	effective viscosity of the lithosphere
p	volumetric body force	body force	gravity
w	radial displacement	displacement	depth of Benioff zone or slab position from tomography
ϕ	Airy stress function	stress function	stress from negative buoyancy
k	wave number	wave number	1/subduction segment length
n	number of dimples or facets	number of subduction zone segments	number of distinct arcs along the northwestern Pacific margin

strip width and amplitude are very small the force required to deform the lithosphere is very high even though the deformation of the sublithospheric mantle becomes negligible. For example, accommodating subduction of the Pacific plate through a very large, smooth bulge would have the effect of changing the radius of the Earth in one hemisphere, requiring huge amounts of energy to deform the sublithospheric mantle; accommodating Pacific subduction through a very narrow kink requires comparably large forces to deform the stiff oceanic lithosphere. Therefore, the trade-off strongly favors intermediate length scales for subduction zone features.

[16] Specifically, the deformation of an elastic lithospheric cap of thickness h_l (and Young's modulus E_l) on a mantle of thickness H_m (and modulus E_m) can be initially treated as a doubly curved thin shell supported by a thick elastic foundation because the thickness of the sublithospheric mantle is much greater than the thickness of the lithosphere. The scaling can be estimated using our energetic argument, because the total mechanical energy per unit area of the lithospheric cap is given by the sum of the elastic energy stored in the deformed mantle "foundation" with a local stiffness $K \sim E_m/H_m$ (here we do not consider the more general nonlocal model of mantle deformation for simplicity), and the stretching and bending energy to deform the lithosphere with bending stiffness $B \sim E_l h_l^3$, and stretching stiffness $S = E_l h_l$, so that: $U = KA^2 + S \frac{A^2}{R^2} + B \frac{A^2}{\omega^2}$ where α is the amplitude of the subduction forebulge, and ω is its width. For an inextensible cap, $W \sim A$, so the gravitational

and stretching terms trade-off against the bending term over a range of wavelengths, with minimum energy at the intermediate scale of the critical wavelength $W \sim (E_l h_l^3 / (K + S/R^2))^{1/4}$. We see that for a shell of sufficiently large radius $R \gg (S/K)^{1/2}$, the bending strip width $W \sim (E_l h_l^3 H_m / E_m)^{1/4}$ is independent of R , consistent with the well known result for the buckling of a flat plate on an elastic foundation, whereas for a shell of sufficiently small radius $R \ll (S/K)^{1/2}$, the width $A \sim (R h_l)^{1/2}$, is the well known result for the buckling of a full spherical shell with different internal and external pressure. The segment length of the buckles in a spherical shell, λ , also reflects the balance between mantle deformation and lithospheric bending and stretching in exactly the same way as the width of the bending strip (forebulge) so that the fundamental scaling is the same, $L_p \sim (E_l h_l^3 H_m / E_m)^{1/4}$.

4. Viscous Scaling

[17] Our subduction analysis may be extended to the case where the lithosphere and underlying mantle deform as viscous bodies by using the mathematical similarity between the equations for the equilibrium of a linear elastic solid and those for creeping flow of a viscous fluid, first enunciated by Stokes and Rayleigh for bulk flows, and exploited by solving problems in one domain using known results in the other [Rayleigh, 1878]. This Stokes-Rayleigh analogy should be familiar to geophysicists through a comparison of Hooke's Law, $\sigma_{ij} = c_{ijkl} \epsilon_{kl}$, describing equilibrium in elastic

geologic materials, with the equation for bulk creeping flow of Newtonian fluids, $\sigma_{ij} = \mu \dot{\epsilon}_{kl}$. A related use of the Stokes-Rayleigh analogy is used for numerical models of subduction in the work by *OzBench et al.* [2008], demonstrating that viscoelastic, viscoelastoplastic, and purely viscous formulations all allow for similar emergent subduction dynamics. Hence, assuming Newtonian viscosity, we may use the scaling relations derived above by substituting strain rate for strain and viscosity for the elastic shear modulus. Because we only considering the limits on scaling of subduction zones along with the generalized constraints on their geometry, we believe the approximation of Newtonian viscosity is appropriate, even though the real constitutive law for the coupled lithosphere-mantle system is certainly more complicated.

[18] An asymptotic procedure similar to one used to derive analogous equations for naturally flat thin viscous sheets [Howell, 1996] may be used to derive the equations for viscous doubly curved shells, but here we simply use the Stokes-Rayleigh analogy to write down a set of differential equations for the mechanics of thin shallow spherical caps of radius R , thickness h made of an incompressible liquid of viscosity μ with a volumetric body force such as gravity or a pressure p as

$$\frac{1}{3}\mu h^3 \nabla^4 w_{,t} + p + \frac{1}{R} \nabla^2 \phi = w_{,xx} \phi_{,yy} + w_{,yy} \phi_{,xx} - 2w_{,xy} \phi_{,xy};$$

$$\nabla^4 \phi - \frac{\mu h}{R} \nabla^2 w_{,t} = -\mu h (w_{,xx} w_{,yy} - w_{,xy}^2)_{,t}$$

where $w(x,y,t)$ is the radial displacement field of the shell relative to its natural spherical shape, and $\phi(x,y)$ is the Airy stress function whose second derivatives correspond to the various components of the in-plane stress tensor. These equations are analogous to the Donnell–Foppl–von Karman equations [Donnell, 1975], which may be obtained by replacing the shear viscosity by the shear modulus, and the radial velocity field by the radial displacement field, and govern the dimpling instability of a viscous or elastic shell loaded by a body force and are complete when boundary conditions are specified.

[19] As an example of a stability calculation in the viscous limit, we consider a complete spherical shell in a uniformly shrinking state that is prestressed by a uniform pressure so that $\phi(x,y) = -\frac{1}{4}pR(x^2 + y^2) + \Phi(x,y)$, with $w(x,y,t) = -\frac{pR^2 t}{4\mu h} + W(x,y,t)$. A linearized analysis about this isotropic state may be carried out by looking for spatially periodic solutions of the form $W \sim \exp[i(k_x \bar{x} + k_y \bar{y})] \exp(\sigma t)$; $\Phi \sim \exp[i(k_x \bar{x} + k_y \bar{y})]$. This yields the following relation for the growth rate of dimples in the shell: $\sigma = -\frac{pR}{2} \left[\mu h \left(\frac{h^2 k^2}{3} + \frac{1}{k^2 R^2} \right) \right]^{-1}$ where $k^2 = k_x^2 + k_y^2$. We note that for negatively buoyant shells ($p = -\Delta \rho g h$) all modes grow, so this is not strictly an instability in that there is a no critical threshold pressure above which the shell is unstable (unlike for the elastic case where there is a critical pressure). However, there is a maximum growth rate with wave number $k = 1/\sqrt{Rh}$, so with a characteristic wavelength $\lambda = \sqrt{Rh}$. As expected, this is mathematically analogous to the calculation carried out for the elastic case where the same wavelength corresponds to the energetically favored mode of buckling at some

critical value of the pressure [Donnell, 1975; Koiter, 1969]. Since there is no preferred direction for the Eigen functions, for an elastic shell, these modes will tile the surface; this leads naturally to the triad condition $\vec{k}_1 + \vec{k}_2 + \vec{k}_3 = 0$ for wave vectors that is the basis for the almost hexagonal array seen in buckled elastic shells [Hutchinson, 1967].

[20] For a viscous shell, and especially for a slightly heterogeneous incomplete viscous cap, the tiling condition no longer applies; instead, the pattern of buckles depends crucially on heterogeneities in the initial thickness of the shell and deviations from sphericity. Hence for the Earth, small variations in lithospheric thickness, as might occur beneath seamounts or fracture zones, should affect the arrangement of dimpled arc segments [Vogt, 1973]. We note that just as for an elastic shell, there is a range of critical differential pressures at which dimples grow (owing to the subcritical nature of the instability), so a range of wave numbers are simultaneously unstable for any given pressure as shown above.

5. Edge Stability

[21] Sections 3 and 4 give scaling relations for complete spherical shells. However, subduction on Earth involves a fundamental discontinuity in the lithosphere; the tip of subducting slabs is a free edge. This edge changes the dynamics of subduction. Because free edges are the easiest part of a spherical cap to bend, they can deform most easily, and thus one would expect the dimples required by the subduction space constraint to appear at free edges first.

[22] The length and shape of individual subduction segments along a free edge can be calculated in numerical simulations of a layered cap loaded by a body force that is positive near the center of the cap and gradually changes to becoming negative at its free edge (to simulate changing buoyancy) (Figure 3); the shell forms concave dimples along its free edge with the same morphology of those seen in arcuate subduction zones; each arc is concave in the direction of the subducting slab and arcs are separated by folded syntaxes.

[23] We model the mantle-lithosphere composite cap using ABAQUS, a commercial finite element package, as a stiff elastic skin on a soft but thick elastic foundation. The material parameters used are $E_m = 1.6$ MPa, $E_l = 10^3$ MPa, and Poisson's ratio of 0.3 for both lithosphere and underlying mantle. We recognize that the sublithospheric mantle behaves as a Maxwell solid, so our choice of Young's modulus in this simulation is merely intended to represent qualitatively the relative stiffness of the two layers. Therefore, in this model, the sublithospheric mantle is softer than the lithosphere but is of much greater thickness, so that the lithosphere forms a thin stiff skin, as in the complete spheres used for our derivations in sections 3 and 4. The radius of the spherical composite cap is 6400 km, and the inner cap surface (at the core-mantle boundary) has a clamped (no displacement) boundary condition. The lithosphere is loaded by a smoothly varying normal pressure reflecting the change in cap buoyancy from the cap center where it is positive to its edge where it is negative (Figure 3). The other surfaces of the cap are traction free.

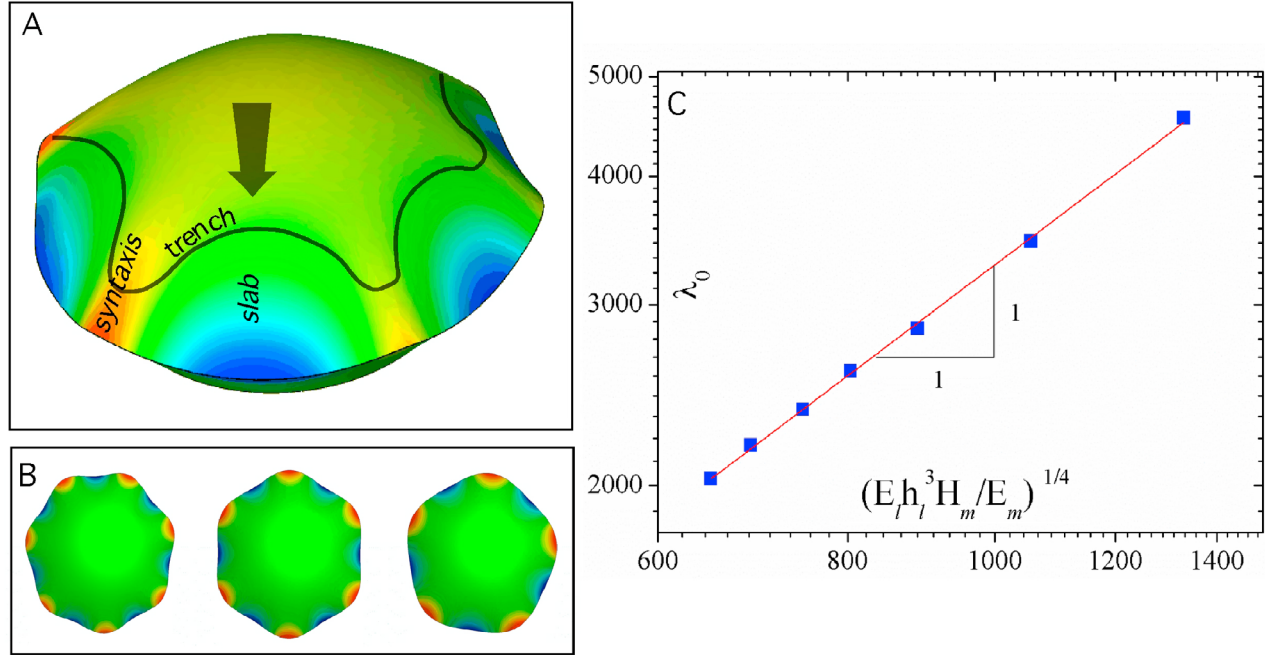


Figure 4. (a) A numerical simulation of the edge instability in a thin elastic cap over a thick elastic foundation which is loaded by a radially varying body force. (b) Top view of the primary buckling modes for different thicknesses of the mantle, i.e., $H_m = 500, 1000$, and 2000 km (from left to right), shows that the wavelength increases with H_m , but the deformations remain localized to the edge of the cap. The colors in the images simply give the dimensionless displacement in the radial direction. (c) The scaling relation for the wavelength of the dimples as a function of the natural length scale arises from the competition between lithosphere bending and mantle deformation (see text for details). All lengths are shown in kilometers. The simulations were carried out with the following parameter values: $E_m = 1.6$ MPa, $E_l = 10^3$ MPa, $h_l = 25$ km.

[24] As the vertical force is increased, the free edge starts to subduct, as shown in Figure 4. In Animation S1, we show the first 20 linear buckling modes of the lithosphere-mantle system. In particular, we see that although the first three modes are all edge modes that are strongly localized to the periphery of the lithospheric cap, the higher modes switch from those that are localized to the edge to those that involve the entire cap. This subtle interplay between the local and global modes coupled with the nonlinear onset of subduction suggests that a detailed analysis is required to fully account for the role of heterogeneities in geometrical and material properties.

[25] In Animation S2, we show the buckling behavior of the elastic model, now focusing only on the first localized mode of cap deformation. The deformation remains localized to the edge of the cap, over a scale comparable to the wavelength predicted by linear analysis of an elastically supported plate or shell.

[26] A set of elastic analog models (auxiliary material), scaling limits for both elastic and viscous shells, and numerical simulations of elastic caps all demonstrate the same result: the general morphology of subduction is profoundly influenced by the curvature of the Earth, and the specific characteristic scales in subduction zones depend on that curvature, the elastic or viscous moduli of the lithosphere and mantle, and the thickness of the lithosphere. In

Figure 4, we show the linearized edge buckling modes of an elastic shell against thicknesses and moduli of the lithosphere and sublithospheric mantle with the radius of the Earth. This plot was generated by measuring subduction buckle wavelength in simulations in which the radius of the Earth and the thicknesses of the sublithospheric mantle and the lithosphere were varied, as well as the ratio of the moduli of the two layers over a range consistent with observed variance of these parameters on Earth. The wavelength of the buckles parallel to the cap edge is $\lambda \sim (E_l h_l^3 H_m / E_m)^{1/4}$. We can use this result to put bounds on the extent of localization of buckling to the cap edge, which is a function of the relative buoyancy of the edge and the Poisson ratio of the material of the shell [Tostvik and Smirnov, 2001]. Inserting typical parameter values ($\mu_h = 10^{22} - 10^{23}$ Pa s, $h_l = 25$ km, $H_m = 600 - 700$ km, and $\mu_m = 10^{19} - 10^{20}$ Pa s) [Billen and Hirth, 2005; Billen et al., 2003] especially from the systematic assessment of viscosity ratio and lithospheric thickness in yields $\lambda \sim 500$ km, which is qualitatively consistent with terrestrial observations [Smith and Sandwell, 1997]. Caps with long edges, such as the Pacific Rim from Alaska to the Marianas Trench, are divided into $n \approx L/\sigma \approx 1 - 5$ arcs separated by narrow syntaxes (Figure 1). Thus, the curvature of the Earth results in limits on the length (and width) of subduction arcs, leading to the segmentation seen on the very long subducting edges of the Pacific plate,

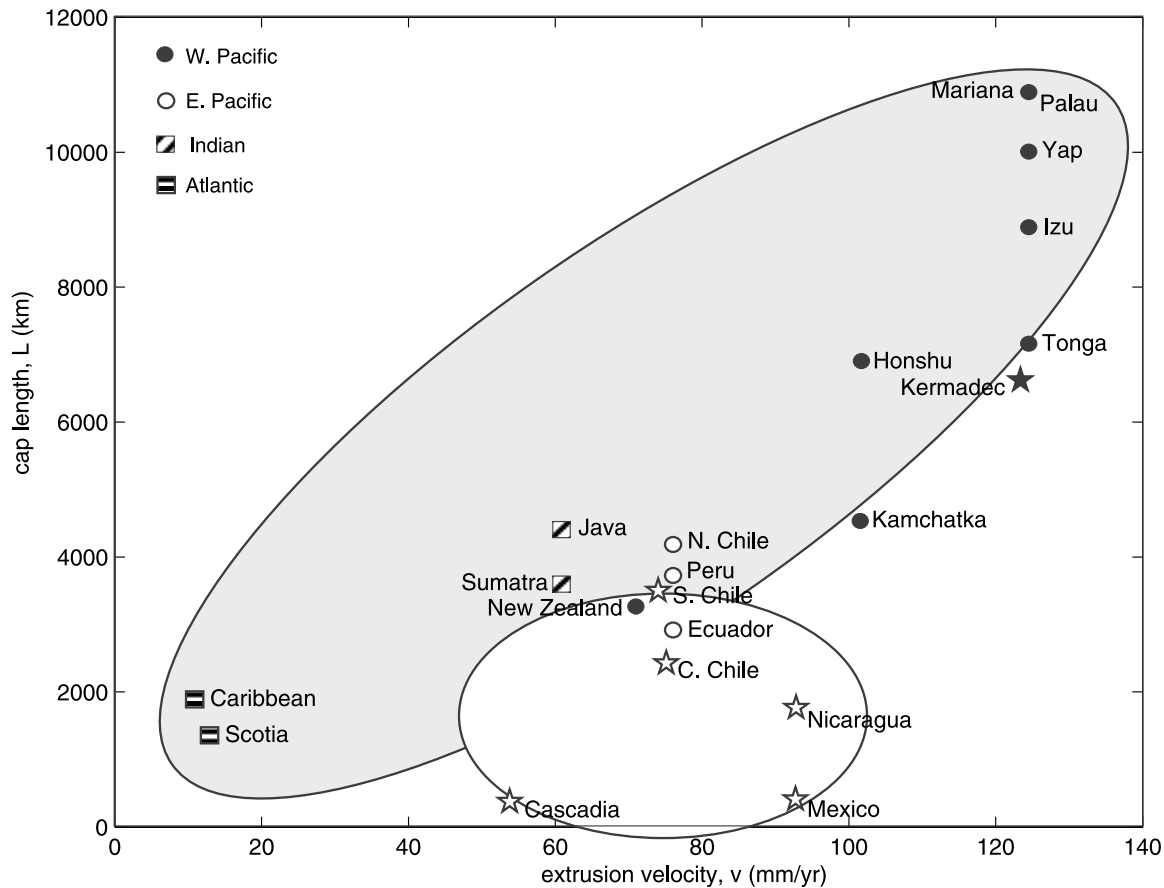


Figure 5. Subduction of negatively buoyant lithosphere (within gray ellipse) has the predicted arcuate morphology (circles and squares) and correlation between plate length and extrusion velocity. Subduction of positively buoyant, thin lithosphere (white ellipse) has the predicted straight morphology (stars) and lacks a correlation between plate length and velocity.

especially, but in subduction zones everywhere. The boundary conditions on subduction determine the morphology of these segments, specifically their curvature where negative buoyancy is causal.

[27] All subduction on Earth with this curved form occurs in oceanic lithosphere greater than 50 Ma, consistent with the predicted significance of negative buoyancy [England and Wortel, 1980], including the entire western Pacific, Java-Sumatra, Scotia, the Caribbean, and probably the Himalayan and the Hellenic arcs (Figure 5). Furthermore, the relatively steep mean dip of slabs at arcuate zones of $49^\circ \pm 18^\circ$ (1σ) is consistent with the requirement of a body force provided by negative buoyancy as the cause of plate bending. Taking this result in an inverse sense, subduction zone morphology can be used to identify where initial plate bending is excited by negative buoyancy.

6. Forced Subduction

[28] Forced subduction, the other mode of bending the curved lithosphere, is initiated by the emplacement of a point or end load on a cap that lacks negative buoyancy but whose rheological stiffness is low, such as where it is young

and thin. In this mode, the space constraints and the generalized scaling relations developed for the body force case above are still the same, but the morphology of the edge instability is different. A forced spherical cap bends with straight-sided polygonal facets.

[29] This subduction mode occurs only when the plates are thin, and therefore young and positively or neutrally buoyant. Subduction under this condition is not subject to bounds on plate length (L) perpendicular to the subduction zone, because it depends only on the location of the load. The characteristic bending zone width, amplitude, and segment length (hence number of facets) in this case are the same as those in the dimpling case because they are still determined by the trade-off between volume change of the Earth and the bending energy of the curved lithospheric cap. Again, facets cannot occur at large wavelengths because of internal incompressibility and they cannot occur at very small wavelengths because displacement of the lithosphere becomes difficult.

[30] All straight (and no arcuate) subduction on Earth occurs in oceanic lithosphere younger than 50 Ma old, such as Central America, Cascadia, and Chile, thus confirming that faceted (i.e., forced) boundaries occur in positively or

neutrally buoyant but thin lithosphere (Figure 5) loaded by horizontal compression or an overriding plate. The relatively shallow dips of slabs at straight zones of $37^\circ \pm 16^\circ$ (1σ) are consistent with the correlation of thin caps with positive buoyancy. Furthermore, all modern occurrences of flat slab subduction, strongly suggestive of positive or neutral buoyancy, occur at boundaries with the faceted shape.

[31] In light of the Stokes-Rayleigh analogy, we expect that a doubly curved viscous cap with a nearly incompressible foundation subject to a radial load also leads to the development of concave, arcuate dimples along the free edge (Figure 4).

7. Discussion

[32] By correctly accounting for double curvature of tectonic plates in the mechanics of subduction and limiting rheological behaviors of the convecting mantle or of the lithosphere, several important characteristics emerge and match observations of real subduction zones. (1) Both arcuate and straight subduction zones exist; arcuate zones occur in negatively buoyant, hence steeply dipping, lithosphere, and straight zones in positively or neutrally buoyant, hence shallowly dipping, lithosphere (Figure 5). (2) There are no diffuse subduction zones; localized deformation of the bending strip in subduction zones is expressed in the topography as a narrow fore-arc bulge and trench pair. (3) Long subduction zones are divided into multiple segments whose scale is limited by properties of the lithosphere but also influenced by small heterogeneities in the subducting plate. (4) Arc polarity is set by cap edges, always concave away from spreading ridges and toward slabs. (5) The distance between ridges and subduction zones depends on spreading rate in free subduction, but not in forced subduction (Figure 5).

[33] Some authors [e.g., *Gurnis et al.*, 2004] suggest a genetic relation between forced and free subduction, in which the former may evolve into the latter. We have not explored how such a situation would affect subduction zone morphology, but it does appear that the type of edge instability, dimpled (curved) or faceted (straight sided) may be determined at or near the time of subduction initiation and then persists of the duration of the plate boundary. This morphological persistence implies that we may discuss the shape and scaling of subduction segments at initiation and in steady state interchangeably. Both forms of subduction instability are themselves nonlinear, hence sensitive to small variations in cooling rates, composition, plate thickness, and local rheology, consistent with the spatial coincidence between aseismic ridges or seamount chains and syntaxes [e.g., *Vogt*, 1973; *Hsui and Youngquist*, 1985] and with the substantial variation in oceanic plate length globally.

[34] As evident in Figure 5, the most significant exception to these scaling and morphology rules is the Tonga-Kermadec subduction zone. This boundary is straight, or even convex trenchward with a tightly folded northern tip despite the great age and negative buoyancy of incoming oceanic lithosphere. We note that this system is actually two subduction segments separated by a poorly developed syntaxis, so may reflect nonlinear deformation processes of the

cap edge that we do not consider in our solutions. Seismic observations [e.g., *Fischer et al.*, 1991; *Shiraishi et al.*, 2008] are consistent with major differences between the properties and geometry of the northern and southern parts of the system, and *Gurnis et al.* [2004] note that the current subduction zone represents reactivation of an ancestral one. Because of the subcriticality of deformations of the edge of curved caps, we cannot exactly predict subduction zone shape or segment arrangement without knowing the complete initial conditions for the plate, an impossibility for real terrestrial subduction.

[35] Although the target of this study is the morphology of terrestrial subduction zones, our scaling relations and stability analyses for both spherical shells and spherical caps hold for curved objects over a wide range of scales. Analog experiments on complete and incomplete shells made of a variety of materials with diameters from 20 to 1500 mm and thicknesses from 1 to 5 mm reproduce our calculated scaling and morphology when loaded by either a point force or a body force (see auxiliary material).

8. Conclusions

[36] Subduction of oceanic lithosphere can initiate under two very different circumstances: foundering of old, cold, thick lithosphere whose negative buoyancy overcomes plate stiffness, or young, hot, thin lithosphere whose stiffness is insufficient to support an applied load. Examples of the former case, free subduction, have steeper slab dips than examples of the latter case, forced subduction, to 90% confidence. In both cases, the geometry of the curved lithosphere is important to the scaling, initiation, and morphology of subduction because both entail changes in cap curvature at subduction zones, which are limited by the geometry and mechanics of doubly curved shells. Hence the shape, spatial scale, and initiation of subduction boundaries are related by a tradeoff between the energy costs due to bending and to stretching of the lithosphere-mantle system as it changes shape.

[37] Numerical solutions and analytic end-member analyses for curved cap edge geometry demonstrate that dynamic approximations of subduction on any planet in two dimensions or in three dimensions using flat sheets omit important mechanical effects. Complete characterization of subduction zones and other plate deflections (such as about ocean islands) requires inclusion of geometric effects in both elastic and viscous approximations. This requirement is especially important where either the state of stress or the morphology of the deflection is important, such as in calculating coupling for seismic moment storage, considering along-arc strain in a downgoing plate, analyzing the shape of accretionary wedges, determining the position of convergent mountain ranges and volcanic arcs, or considering along-arc variations in strain. Correctly accounting for shape is also important when using subduction morphology to invert for elastic moduli or effective viscosity of lithospheric plates, because some of the competence of the plate to support a load (whether a body or boundary force) is related not to its rheology but simply to its geometry.

References

- Bevis, M. (1986), The curvature of Wadati-Benioff zones and the torsional rigidity of subducting plates, *Nature*, 323, 52–53, doi:10.1038/323052a0.
- Billen, M. I., and G. Hirth (2005), Newtonian versus non-Newtonian upper mantle viscosity: Implications for subduction initiation, *Geophys. Res. Lett.*, 32, L19304, doi:10.1029/2005GL023457.
- Billen, M. I., M. Gurnis, and M. Simons (2003), Multi-scale dynamics of the Tonga-Kermadec subduction zone, *Geophys. J. Int.*, 153, 359–388, doi:10.1046/j.1365-246X.2003.01915.x.
- Cloos, M. (1993), Subduction of oceanic plateaus, continental margins, island arcs, spreading ridges, and seamounts, *Geol. Soc. Am. Bull.*, 105, 715–737, doi:10.1130/0016-7606(1993)105<0715:LBACOS>2.3.CO;2.
- Conrad, C. P., and C. Lithgow-Bertelloni (2004), The temporal evolution of plate driving forces: Importance of “slab suction” versus “slab pull” during the Cenozoic, *J. Geophys. Res.*, 109, B10407, doi:10.1029/2004JB002991.
- da Silveira, R., S. Chaïeb, and L. Mahadevan (2000), Rippling instability of a collapsing bubble, *Science*, 287, 1468–1471, doi:10.1126/science.287.5457.1468.
- Donnell, L. H. (1975), *Beams, Plates and Shells*, McGraw Hill, New York.
- England, P., and R. Wortel (1980), Some consequences of the subduction of young slabs, *Earth Planet. Sci. Lett.*, 47, 403–415.
- Fischer, K. M., K. C. Creager, and T. H. Jordan (1991), Mapping the Tonga slab, *J. Geophys. Res.*, 96, 14,403–14,427, doi:10.1029/90JB02703.
- Frank, F. (1968), Curvature of island arcs, *Nature*, 220, 363, doi:10.1038/220363a0.
- Funiciello, F., C. Faccenna, D. Giardini, and K. Regenauer-Lieb (2003), Dynamics of retreating slabs: 2. Insights from three-dimensional laboratory experiments, *J. Geophys. Res.*, 108(B4), 2207, doi:10.1029/2001JB000896.
- Gurnis, M., C. Hall, and L. Lavier (2004), Evolving force balance during incipient subduction, *Geochim. Geophys. Res.*, 5, Q07001, doi:10.1029/2003GC000681.
- Hayes, G. P., D. J. Wald, and K. Keranen (2009), Advancing techniques to constrain the geometry of the seismic rupture plane on subduction interfaces a priori: Higher-order functional fits, *Geochim. Geophys. Res.*, 10, Q09006, doi:10.1029/2009GC002633.
- Howell, P. (1996), Models for slender thin viscous sheets, *Eur. J. Appl. Math.*, 7, 321–343, doi:10.1017/S0956792500002400.
- Hsui, A., and S. Youngquist (1985), A dynamic model of the curvature of the Mariana trench, *Nature*, 318, 455–457, doi:10.1038/318455a0.
- Hutchinson, J. W. (1967), Imperfection sensitivity of externally pressurized spherical shells, *J. Appl. Mech.*, 34, 49–55.
- Koiter, W. (1969), Nonlinear buckling problem of a complete spherical shell under uniform external pressure 1–4, *Proc. K. Ned. Akad. Wet., Ser. B Phys. Sci.*, 72, 40–123.
- Landau, L., and E. Lifshitz (1986), *Theory of Elasticity*, Butterworth Heinemann, Oxford, U. K.
- Mahadevan, L., A. Vaziri, and M. Das (2007), Persistence of a pinch in a pipe, *Europhys. Lett.*, 77, 40003, doi:10.1209/0295-5075/77/40003.
- Molnar, P., and T. Atwater (1978), Interarc spreading and cordilleran tectonics as alternates related to the age of subducted oceanic lithosphere, *Earth Planet. Sci. Lett.*, 41, 330–340, doi:10.1016/0012-821X(78)90187-5.
- Morra, G., K. Regenauer-Lieb, and D. Giardini (2006), Curvature of oceanic arcs, *Geology*, 34, 877–880, doi:10.1130/G22462.1.
- Oxburgh, E., and E. Parmentier (1977), Compositional and density stratification in oceanic lithosphere, *J. Geol. Soc.*, 133, 343–355, doi:10.1144/gsjgs.133.4.0343.
- OzBench, M., et al. (2008), A model comparison study of large-scale mantle-lithosphere dynamics driven by subduction, *Phys. Earth Planet. Inter.*, 171, 224–234, doi:10.1016/j.pepi.2008.08.011.
- Rayleigh, J. (1878), *The Theory of Sound*, vol. 2, Macmillan, London.
- Ribe, N. (1992), The dynamics of thin shells with variable viscosity and the origin of toroidal flow in the mantle, *Geophys. J. Int.*, 110, 537–552, doi:10.1111/j.1365-246X.1992.tb02090.x.
- Ribe, N. (2001), Bending and stretching of thin viscous sheets, *J. Fluid Mech.*, 433, 135–160.
- Ribe, N. (2010), Bending mechanics and mode selection in free subduction: A thin-sheet analysis, *Geophys. J. Int.*, 180, 559–576, doi:10.1111/j.1365-246X.2009.04460.x.
- Schellart, W. P. (2004), Kinematics of subduction and subduction-induced flow in the upper mantle, *J. Geophys. Res.*, 109, B07401, doi:10.1029/2004JB002970.
- Schellart, W. P., J. Freeman, D. R. Stegman, L. Moresi, and D. May (2007), Evolution and diversity of subduction zones controlled by slab width, *Nature*, 446, 308–311, doi:10.1038/nature05615.
- Shiraishi, R., E. Ohtani, K. Kanagawa, A. Shimojuku, and D. Zhao (2008), Crystallographic preferred orientation of akimotoite and seismic anisotropy of Tonga slab, *Nature*, 455, 657–660, doi:10.1038/nature07301.
- Smith, W. H. F., and D. T. Sandwell (1997), Global sea floor topography from satellite altimetry and ship depth soundings, *Science*, 277, 1956–1962.
- Stegman, D. R., J. Freeman, W. P. Schellart, L. Moresi, and D. May (2006), Influence of trench width on subduction hinge retreat rates in 3-D models of slab rollback, *Geochim. Geophys. Res.*, 7, Q03012, doi:10.1029/2005GC001056.
- Stegman, D. R., R. Farrington, F. A. Capitanio, and W. P. Schellart (2010), A regime diagram for subduction styles from 3-D numerical models of free subduction, *Tectonophysics*, 483, 29–45, doi:10.1016/j.tecto.2009.08.041.
- Tanimoto, T. (1997), Bending of spherical lithosphere—Axisymmetric case, *Geophys. J. Int.*, 129, 305–310, doi:10.1111/j.1365-246X.1997.tb01583.x.
- Tanimoto, T. (1998), State of stress within a bending spherical shell and its implications for subducting lithosphere, *Geophys. J. Int.*, 134, 199–206, doi:10.1046/j.1365-246x.1998.00554.x.
- Tostvik, T., and A. Smirnov (2001), *Asymptotic Methods in the Buckling Theory of Elastic Shells*, World Sci., Singapore.
- Tovish, A., and G. Schubert (1978), Island arc curvature, velocity of convergence and angle of subduction, *Geophys. Res. Lett.*, 5, 329–332, doi:10.1029/GL0051005p00329.
- Turcotte, D. L., and J. L. Ahern (1977), On the thermal and subsidence history of sedimentary basins, *J. Geophys. Res.*, 82, 3762–3766, doi:10.1029/JB082i026p03762.
- Turcotte, D. L., and G. Schubert (2002), *Geodynamics*, 2nd ed., Cambridge Univ. Press, Cambridge, U. K.
- Vaziri, A., and L. Mahadevan (2008), Localized and extended deformations of elastic shells, *Proc. Natl. Acad. Sci. U. S. A.*, 105, 7913–7918.
- Vogt, P. (1973), Subduction and aseismic ridges, *Nature*, 241, 189–191, doi:10.1038/241189a0.
- Yamaoka, K., Y. Fukao, and M. Kumazawa (1986), Spherical shell tectonics: Effects of sphericity and inextensibility on the geometry of the descending lithosphere, *Rev. Geophys.*, 24, 27–53, doi:10.1029/RG024i001p00027.

R. Bendick, Department of Geosciences, University of Montana, Missoula, MT 59812, USA. (bendick@mso.umt.edu)

H. Liang, Department of Modern Mechanics, University of Science and Technology of China, Hefei 230026, China. (hyliang@ustc.edu.cn)

L. Mahadevan, School of Engineering and Applied Sciences, Harvard University, Pierce Hall, 29 Oxford St., Cambridge, MA 02138, USA. (lm@seas.harvard.edu)