

# Mantle Convection with Plates and Mobile, Faulted Plate Margins

Shijie Zhong\* and Michael Gurnis

A finite-element formulation of faults has been incorporated into time-dependent models of mantle convection with realistic rheology, continents, and phase changes. Realistic tectonic plates naturally form with self-consistent coupling between plate and mantle dynamics. After the initiation of subduction, trenches rapidly roll back with subducted slabs temporarily laid out along the base of the transition zone. After the slabs have penetrated into the lower mantle, the velocity of trench migration decreases markedly. The inhibition of slab penetration into the lower mantle by the 670-kilometer phase change is greatly reduced in these models as compared to models without tectonic plates.

The surface expression of mantle convection is defined by tectonic plates, but modeling plates and mantle convection self-consistently has proven difficult because of the extreme variations in rheology associated with plates, especially of plate margins. Plate kinematics are time-dependent, which introduces a level of complexity that has yet to be adequately modeled. Recent computational studies have emphasized the role played by the internal thermodynamics of the mantle in modulating the time-dependent characteristics of convection. The seismic discontinuity at a depth of 670 km, probably a solid-solid phase transition, may cause slabs to periodically accumulate within the mantle transition zone before leading to a huge, potentially global, sinking of cold material into the lower mantle (1).

It has been speculated that these events may influence the geological and geochemical evolution of the planet (1, 2). Unfortunately, the surface expression of these recent convection models is unlike the pattern of flow displayed by plate tectonics. Three features of convection are well established (3): (i) the flow across a plate is uniform (strain is concentrated at plate margins); (ii) the heat flow and bathymetry decay as  $\text{age}^{-1/2}$  and  $\text{age}^{1/2}$ , respectively; and (iii) the converging margin has asymmetrical convergence (only one plate subducts) and is defined at the surface by an oceanic trench. If a mantle convection model satisfies these characteristics, then the inferences made from the model regarding the deep interior and the influences on geological and geochemical observations are much more certain.

In addition to these instantaneous features of plate tectonics, plate kinematics are time-dependent and are characterized by

changes in the size of plates and by the migration of the margins between plates. One of the most fundamental aspects of time dependence is the oceanward migration of plates overriding subduction zones. After the breakup of Pangea, the North and South American plates moved westward in a hot spot reference frame with new ocean floor generated in the Atlantic (4); this motion led to an increase in the size of both of these plates and to the migration of trenches on the western margins of North America and South America (4). Moreover, from the correlation of observables on a regional scale, it has been suggested that the migration of converging margins influences the tectonics in back-arc regions (5), the dip angle of subducted slabs (6), and slab penetration through the transition zone (7). If we want to understand these fundamental, time-dependent features of mantle dynamics, we need to formulate models of the mantle that include plates with mobile margins. The positions and characteristics of plate margins must change in response to mantle flow.

Earlier studies have shown that the primary influence of plates is to organize mantle flow into a large-scale pattern commonly referred to as plate-scale mantle flow (8). Most studies have used either a material property method (9) or an imposed plate velocity method (8) to generate plates with a predetermined geometry. However, Gurnis and Hager (10) formulated two-dimensional Cartesian convection models in which plate margins could move freely with respect to the center of mass of the entire system in response to internal and plate dynamics. These models with periodic boundary conditions showed that the subducting plate generally decreases in size as the diverging margin moves toward the converging margin and that migration of the converging margin leads to a shallow subduction dip angle (10). The margins were weak zones defined entirely by the

viscosity necessary to decouple plates from adjacent plates, but such a method is unrealistic because it imposes weak zones with a lateral dimension of about the thickness of the plates, 100 km (11).

Recent developments in modeling faults (12, 13) now allow us to use significantly more realistic plate margins in convection models. Dynamic models of subducted slabs (12) that incorporate faults as converging margins suggest that oceanic trenches and back-arc basins dynamically originate from compensation of dense slabs. Such models can be used to explain the observed relation among trench depth and subduction dip angles and slab age (12).

In this article, we describe the results of models that incorporate plates with mobile converging margins and realistic rheology, continents, and phase changes in a cylindrical geometry. The converging margin is simulated by a fault whose position is determined dynamically by mantle flow (12, 13). We discuss how plates with a mobile converging margin interact dynamically with convection, subduction, and a phase change. These results have important implications for the observed past plate kinematics of North and South America, the dynamics of subduction zones, and the mechanism of slab penetration through the transition zones and the 670-km mantle discontinuity.

## Model Setup

The mobile plate margin is created with a fault algorithm (Table 1). We used the finite-element method with constrained elements and Lagrange multipliers to solve convection models with a fault (12); the mobility of the fault emerges from a mixed Eulerian and Lagrangian formulation (13). The models with an endothermic phase change at a depth of 670 km are computed in a two-dimensional cylindrical geometry (14) (Fig. 1). A fault extending down to the bottom of a plate (15) with an initial dip angle of  $30^\circ$  is used to represent a converging plate margin. Overriding the fault is a continent with a constant length; the remaining surface is cooling oceanic lithosphere. Initially, the interior is isothermal ( $1600^\circ\text{C}$ ), and an initial oceanic plate is included (10). Within the continent there is no horizontal gradient in temperature. At the diverging margin of the overriding plate (that is at  $\theta = \pi$  in Fig. 1), a high temperature is set initially to mimic continental

The authors are with the Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA.

\*To whom correspondence should be addressed.

rifting. The continent initially covers the whole region from the converging margin to the weak margin at  $\theta = \pi$ .

The continent, oceanic plates, and mantle have different rheologies. For regions below a depth of 410 km, we assume that the predominant deformation mechanism is diffusion creep (that is, Newtonian flow), whereas for shallower regions, excluding the continent, we assume that dislocation creep (that is, non-Newtonian flow) predominates (16) (Fig. 1). Viscosity is expressed by a generalized rheological law with a set of constants ( $\mu_0$ ,  $c_1$ ,  $c_2$ ,  $n$ ,  $\tau_t$ ) (Table 1). For most cases, these constants for the lower mantle, the transition zone, the upper mantle, and lithosphere are chosen such that the resulting average effective viscosity (Table 1) is similar to that inferred from geoid studies (17), in which the viscosity is high in the lower mantle but low in the transition zone. The nonsubducting continents are simulated with a constant high viscosity (Table 1). On divergent margins, a smaller  $\mu_0$  or  $\tau_t$  was used to account for the weakening by partial melting.

The boundary conditions are isothermal

boundaries, free slip on both the bottom and the top boundaries, and reflecting on the two side walls. With the reflecting boundary conditions, the reference frame is explicitly the side walls. Positions of the continent are updated for the assumption that the length of the continent is constant. This is a reasonable assumption because strain within the continent is small as a result of high viscosity. The region vacated by the continent is occupied by new ocean floor having a temperature- and strain rate-dependent rheology identical to that of the subducting plate. A mesh with 300 by 70 elements was used for all the cases, and the mesh was refined near the fault, phase change, and plates (13).

## Results

We address two questions: Do trench migration and the nature of deep subduction strongly influence each other, and are the plate kinematics predicted from our convection models compatible with the plate kinematics observed for the Americas? On

the basis of many calculations with a wide range of different model parameters, we have found strong support for an affirmative answer to both questions. The results from case 1 with a phase change (Clapeyron slope of  $-3.5 \text{ MPa K}^{-1}$  and a density increase of 8.5%, Table 2) are representative. The buoyancy introduced with the initial cooling lithosphere and the high temperature in the continental margin give rise to subduction of the oceanic plate and continental drift early in the time history of the model. Continental drift results in an increase in the size of the overriding continental plates with new ocean floor added at the spreading center and oceanward migration of the converging margin (that is, trench migration). Oceanward trench migration leads to a shallow dipping subducted slab, which has a small propensity to penetrate through the 670-km endothermic phase change.

*Time evolution of plates and slabs.* When the newly descending slab encounters the phase change at 670 km, plate motion and slab motion are slightly retarded, as shown by the small decrease in plate velocities at 55 million years (My) (Fig. 2A). Because the slab cannot penetrate the phase change

**Table 1.** Physical and geometrical model parameters.

Physical parameters	
Thickness of the fluid layer, $D$	$2.9 \times 10^6 \text{ m}$
Outer radius, $R_0$	$6.37 \times 10^6 \text{ m}$
Temperature contrast, $T_b - T_0$	1600 K
Reference density, $\rho_0$	$3.3 \times 10^3 \text{ kg m}^{-3}$
Thermal conductivity	$3.0 \text{ W m}^{-1} \text{ K}^{-1}$
Thermal diffusivity, $\kappa$	$10^{-6} \text{ m}^2 \text{ s}^{-1}$
Acceleration of gravity, $g$	$10 \text{ m s}^{-2}$
Thermal expansion, $\alpha$	$2.0 \times 10^{-5} \text{ K}^{-1}$
Average effective viscosity	
in the transition zone (410 to 670 km)	$\sim 4.0 \times 10^{21} \text{ Pa}\cdot\text{s}$
in the upper mantle (120 to 410 km)	$\sim 4.0 \times 10^{20} \text{ Pa}\cdot\text{s}$
in the oceanic lithosphere* (0 to 120 km)	$\sim 2.0 \times 10^{23} \text{ Pa}\cdot\text{s}$
Viscosity of the continent† (0 to 120 km)	$2.0 \times 10^{25} \text{ Pa}\cdot\text{s}$
Equilibrium depth for the endothermic phase change	$6.7 \times 10^5 \text{ m}$
Equilibrium temperature for the endothermic phase change	1400 K
Density jumps across the phase-change boundaries	8.5%
Phase transition width, $d$	$3.2 \times 10^4 \text{ m}$

### Rheological law to express viscosity

For either Newtonian or non-Newtonian rheology, the effective viscosity is given by

$$\mu_{\text{eff}} = \frac{\mu}{1 + \left(\frac{2\mu\dot{\epsilon}}{\tau_t}\right)^{(n-1)/n}}$$

where  $\tau_t$  is transition stress;  $\dot{\epsilon}$  is strain rate;  $n = 1$  for purely Newtonian flow, and  $n \neq 1$  for non-Newtonian flow;  $\mu$ , the Newtonian part of viscosity, may be expressed as

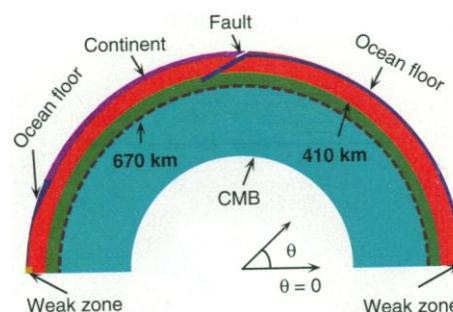
$$\mu = \mu_0 \exp\left(\frac{c_1}{T + c_2} - \frac{c_1}{1 + c_2}\right)$$

where  $c_1$  and  $c_2$  are constants;  $\mu_0$  is the viscosity at nondimensional temperature  $T = 1$ .

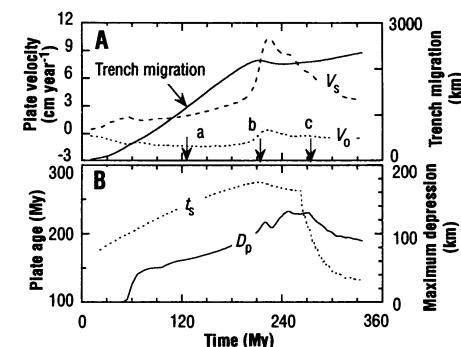
Viscosity in a region is determined by a set of constants ( $\mu_0$ ,  $c_1$ ,  $c_2$ ,  $n$ ,  $\tau_t$ ).

For case 1 ( $\mu_0$ ,  $c_1$ ,  $c_2$ ,  $n$ ,  $\tau_t$ )‡ for the lower mantle: (2 $\eta_r$ , 3.0595, 0.75647, 1,  $\tau_t$ )  
 for the transition zone: (0.4 $\eta_r$ , 3.0595, 0.75647, 1,  $\tau_t$ )  
 for the upper mantle§: (0.02 $\eta_r$ , 1.9187, -1.177, 3,  $\tau_t$ )  
 for the oceanic lithosphere: ( $\eta_r$ , 1.9187, -1.177, 3,  $\tau_t$ )

\*For oceanic plates, the cutoff for the maximum effective viscosity is  $4.0 \times 10^{23} \text{ Pa}\cdot\text{s}$ . †Continents may have higher viscosity than oceanic plates as a result of the loss of volatiles. ‡ $\tau_t$  is 2.5 bar;  $\eta_r$  is the reference viscosity,  $2.0 \times 10^{22} \text{ Pa}\cdot\text{s}$ . For cases 4 and 5,  $\mu_0$ ,  $c_1$ , and  $c_2$  are different (see text). §The value of  $\mu_0$  for cold slab within the upper mantle is equal to that for the oceanic plate,  $\eta_r$ .



**Fig. 1.** Model setup. The dashed line represents the 670-km phase change. The arc length of the model box for most cases is  $\pi$ , except for case 6 for which the arc length is  $\pi - 1$ . CMB, core-mantle boundary.



**Fig. 2.** (A) Time history of overriding ( $V_o$ ) and subducting ( $V_s$ ) plate velocities and trench migration for case 1. (B) Time history of maximum phase-change depression ( $D_p$ ) and age of subducting plate before subduction ( $t_s$ ) for case 1.

initially, it bends in the transition zone slightly (Fig. 3A; Fig. 4A for the zoomed-in thermal structure of the slab). During this period, the subducting and overriding plates converge with relatively constant velocities (Fig. 2A). As a result, the overriding plate increases its size at the expense of the subducting plate at a constant rate. By about 210 My (time b in Fig. 2A), the length of the overriding plate has increased by more than 2000 km (Figs. 2A and 3B). Quite suddenly, the slab sinks into the lower mantle (Fig. 3B). Once the cold material starts sinking into the lower mantle, the velocity of the subducting plate increases rapidly, while the velocity of the overriding plate decreases substantially. The decrease in velocity of the overriding plate slows trench migration and growth of the overriding plate. The velocity of the subducting plate increases to a maximum of about  $10.5 \text{ cm year}^{-1}$  when the slab reaches the middle of the lower mantle (Table 2) and then gradually decreases to a relatively constant velocity of about  $4.0 \text{ cm year}^{-1}$  (Fig. 2A). After cold material sinks into the lower mantle, the slab dip angle becomes increasingly steep (Fig. 3C). Once the slab penetrates into the lower mantle, the phase change fails to influence the history of slab penetration—the slab penetrates steadily into the lower mantle, and the final con-

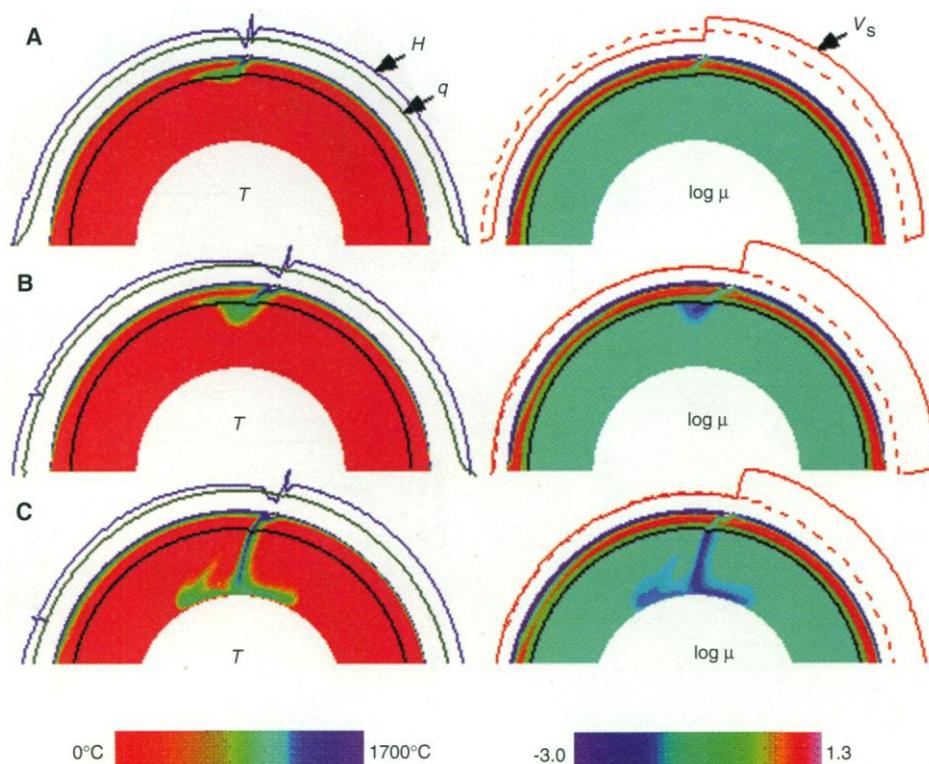
figuration of slab and plates is quite stable (Fig. 3C). Even though the model has only been integrated for 100 My after slab penetration, the stability of the final configuration of slab and plates indicates that slabs penetrate only transiently but that the time dependence of penetration induced by the phase changes may not be as strong as that suggested by isoviscous convection models (1). This steep dip is typical for the earlier models of Gurnis and Hager (10) with reflecting boundary conditions. More realistic periodic boundary conditions should lead to further mobility of the converging margin and perhaps a shallower dip angle (14).

*Surface velocity, topography, and heat flux.* Surface features in this model are consistent with observations. Surface velocity profiles demonstrate that plate-like behavior (that is, significant strain exists only near plate margins) is achieved for both overriding and subducting plates (Fig. 3). The overriding plate consists of both a continental and an oceanic component, but the uniform surface velocity indicates that they form a single coherent plate (Fig. 3A). Asymmetric subduction is effectively achieved with the high continental viscosity and the shallow dipping thrust fault. The continental motion is primarily horizontal, whereas the subducting plate near the fault has significant vertical and horizontal velocities (13).

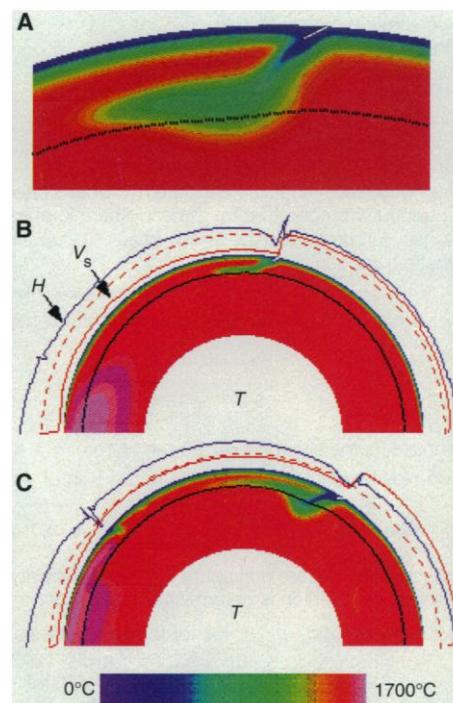
This contrasts with the symmetric convergence displayed by all traditional convection models (18).

The topography of ocean floor and oceanic heat flux decrease with distance from spreading centers and with the age of plates for the subducting and new oceanic plates (Fig. 3) (19). An oceanic trench with width and depth consistent with those from instantaneous flow models (12) forms on the oceanic plate near the converging margin. The cold slab causes a significant depression of topography on the continent near the converging margin, and this depression may extend a few thousand kilometers landward depending on the geometry of slabs (20) (Fig. 3, B and C). This depressed topography over the continent may be overestimated because the high continental viscosity results in a strong coupling between continental plate and the mantle. Heat flux out of the continent is smaller than that out of the ocean floors (21) (for Fig. 3B,  $24 \text{ mW m}^{-2}$  for the continent and  $61 \text{ mW m}^{-2}$  for ocean floors). The time-averaged surface heat flux is  $45 \text{ mW m}^{-2}$  (Table 2) (21).

*Temperature and viscosity.* During the early stages of subduction, slabs may be deflected above the 670-km phase change with a shallow dip angle (Figs. 3A and 4A), and the slabs can extend landward by more than 1000 km (Figs. 3, A and B, and 4A). Slabs that penetrate into the lower mantle with a steep dip may extend down to and spread over the core-mantle boundary (Fig. 3C).



**Fig. 3.** Temperature ( $T$ ), effective viscosity ( $\mu$ ), and surface distribution of dynamic topography ( $H$ ), plate velocity ( $V_s$ ), and heat flux ( $q$ ) for case 1 at three instants in time (time markings in Fig. 2A): (A) at time a; (B) at time b; and (C) at time c. The black lines in the snapshots of  $T$  and  $\mu$  represent the 670-km phase change. The dashed lines near  $V_s$  are the zero-velocity lines. Profiles of  $H$ ,  $V_s$ , and  $q$  have different scales in (A), (B), and (C).



**Fig. 4.** (A) The zoomed-in thermal structure of the slab in Fig. 3A. Temperature ( $T$ ), dynamic topography ( $H$ ), and surface velocity ( $V_s$ ) for case 8 (B) before and (C) after the start of slab penetration.

The viscosity structure is characterized by a low-viscosity layer with variable thickness within the upper mantle, a high-viscosity slab, and a low-viscosity converging margin (Fig. 3). For the given rheological law (Table 1), the viscosity contrast between slabs and ambient mantle is about 50 in the upper mantle and about 3 within the transition zone and lower mantle. The viscosity in the weak converging margin is about 1/1000 of that within the interior of oceanic plates. The low-viscosity layer beneath the continent is at a greater depth than under the ocean floor when the continental motion is small (Fig. 3, B and C). The small continental motion gives rise to a small strain and, through the non-Newtonian rheology, to a large viscosity under the continent.

An older subducting lithosphere or a larger subducting velocity gives rise to a lower temperature in the slab and depresses the depth at which the phase change marking the lower mantle occurs. The maximum depression can be as large as 130 km (Fig. 2B), and this is about twice the depression beneath the Izu-Bonin subduction zone (22). But the average depression over the slab from the model, which is 60 km, is consistent with the observed value (22).

*Effects of different parameters.* We computed models with different phase-change strength (cases 2 and 3), viscosity structure (cases 4 and 5), and initial thermal buoyancy (cases 6, 7, and 8) (Table 2). For all the cases, we observed results similar to those for case 1, which include the formation of a coherent overriding plate with both continental and oceanic components, the temporal evolution of plates, and the connection between trench migration, slab dip angles, and slab penetration.

For case 2, without a phase change, the slab enters the high-viscosity lower mantle without any delay (Table 2), quite similar to the results of earlier work by Gurnis and Hager (10). Not surprisingly, when the Cla-

peyron slope is increased to  $-5.0 \text{ MPa K}^{-1}$  (case 3), the slab is held above the phase change for a longer time before sinking into lower mantle (Table 2). Because the overriding plate continues converging toward oceanic plate until the slab penetrates through 670 km, the overriding plate travels a greater distance compared with cases with smaller Clapeyron slopes.

In the model without a viscosity contrast between the transition zone and the lower mantle (case 4), the slab penetrates into the lower mantle earlier and the converging margin travels a shorter distance, compared with case 1. A smaller lower mantle viscosity decreases the resistance to slab penetration. In case 5, in which the viscosity of slabs within the transition zone is increased by a factor of 10 compared with case 1, it takes longer for the slab to sink into the lower mantle. This seems to be inconsistent with studies (23, 24) that are based on a fixed converging margin in which higher viscosity slabs penetrated more easily. For case 6 with an initially short and thus less negative buoyant oceanic plate, it takes longer for the slab to penetrate into the lower mantle (24).

*Continental motion and deep thermal buoyancy.* A plausible hypothesis (25, 26) for the breakup of Pangea is that hotter than average lower mantle underneath Pangea caused the breakup. The area now vacated by Pangea contains most of the world's hot spots (25, 27) and has low seismic velocity (28). Thermal anomalies within the lower mantle may also be responsible for the rapid drift rate of Laurentia and Baltica during the Late Precambrian and Early Paleozoic (29). Considering that subduction beneath a continent pulls it oceanward (30), it seems prudent to ask which process drives the oceanward continental motion more efficiently, a slab or a lower mantle anomaly.

We therefore included additional lower mantle thermal anomalies (31) under the continent as an initial condition in cases 7

and 8, compared with case 1. The maximum magnitudes (31) of the thermal anomalies for cases 7 and 8 are 25 and 100 K (Table 2), respectively. For both cases, the additional thermal anomaly enhances the oceanward continental motion. Especially for case 8, continental motion is more than twice as large as that for case 1 (Fig. 4B and Table 2). The velocity of the continental plate may be larger than that of the oceanic plate before the slab penetrates into the lower mantle (Fig. 4B), and the subduction is caused primarily by the rollback of the continent over the oceanic plate. This result suggests that a moderately strong thermal anomaly in the lower mantle can produce significant effects on the continental motion and subduction under continents.

For both cases 7 and 8, the continental motion causes slabs under the continent to have shallow dips, which produce large-scale surface depressions on the continent (Fig. 4, B and C). The continental motion diminishes when slab penetrates into the lower mantle (Fig. 4C), similar to that in case 1 (Fig. 3). However, compared with case 1, the converging margin travels a larger distance, and it takes longer for the slab to penetrate into the lower mantle (Table 2). For case 8, when the slab starts to penetrate through into the lower mantle (Fig. 4C), the cumulative trench migration is about twice the migration in case 1, and the resulting new oceanic floor is so large that it can no longer form a coherent plate with the continent. As a result, subduction initiates at the passive margin of the continent (Fig. 4C). As the new ocean floor increases in length, the buoyancy associated with the ocean floor increases. The increased buoyancy tends to enhance the motion of the new ocean floor, but the continental motion decreases as slab penetration occurs. The difference in motion between the continent and the new ocean floor results in subduction at the passive margin.

## Inferences on the Dynamics of Plates and Subduction

The incorporation of mobile faults into convection models results in the formation of realistic plates. We have observed that the temporal evolution of plate size and kinematics are similar to those of the North American and South American plates since the breakup of Pangea. The models demonstrate that the motion of overriding plates and hence the position of the converging margins strongly influence the morphology of slabs and slab penetration through an endothermic phase change at the 670-km depth.

The plates in the models are realistic in terms of the resulting piecewise constant surface velocity with only significant strain

**Table 2.** Model parameters and statistics of models;  $\Delta T$  is the magnitude of the lower mantle thermal anomaly;  $L_0$  is the initial length of oceanic lithosphere;  $\gamma$  is the Clapeyron slope;  $\eta_{lm}$  is the average lower mantle viscosity;  $t_s$  is time that a slab takes to reach about the 1200-km depth;  $M_{last}$  is the cumulative trench migration at the end of each model run.  $V_{omax}$  and  $V_{smax}$  are the maximum overriding and subducting plate velocity, respectively;  $q_{mean}$  is the average heat flux on the surface. The numbers in parentheses are the standard deviations.

Case	$\Delta T$ (K)	$L_0$ (km)	$\gamma$ (MPa $\text{K}^{-1}$ )	$\eta_{lm}$ (Pa·s)	$t_s$ (my)	$M_{last}$ (km)	$V_{smax}$ (cm $\text{year}^{-1}$ )	$V_{omax}$ (cm $\text{year}^{-1}$ )	$q_{mean}$ ( $\text{mW m}^{-2}$ )
1	0	10829	-3.5	$2 \times 10^{22}$	203	2353	10.5	1.4	45 (4.3)
2	0	10829	0	$2 \times 10^{22}$	80	1238	6.7	1.5	47 (2.7)
3	0	10829	-5.0	$2 \times 10^{22}$	307	3252	16.0	1.5	43 (4.7)
4	0	10829	-3.5	$4 \times 10^{21}$	126	1928	11.5	2.2	47 (2.7)
5	0	10829	-3.5	$2 \times 10^{22}$	264	3314	14.5	2.1	45 (3.3)
6	0	4459	-3.5	$2 \times 10^{22}$	231	1274	6.0	1.5	46 (4.7)
7	25	10829	-3.5	$2 \times 10^{22}$	220	2496	9.4	1.9	44 (2.7)
8	100	10829	-3.5	$2 \times 10^{22}$	237	4631	7.4	3.8	48 (2.6)

at plate margins and the decrease in topography and heat flux with distance from the spreading centers (Fig. 3). A single coherent overriding plate forms even though it consists of both continental and oceanic components (Figs. 3 and 4B). The oceanward motion of the overriding continental plate leads to the migration of converging margins and to the growth of continental plates as new ocean floor is generated at spreading centers (Fig. 3, A and B). If the length of new ocean floor is sufficiently large, subduction may initiate at the passive continental margin (Fig. 4C). The oceanward continental motion is driven by the initially set, hot continental margin, the subducting slabs, and the lower mantle thermal anomaly under the continent in those models where it is included. The lower mantle thermal anomaly has a significant influence on continental motion (case 8).

Seismic tomography studies reveal that subducted slabs may either be deflected by the 670-km discontinuity with a relatively shallow dip angle, as under central Japan (32, 33), or penetrate into the lower mantle with a steep dip, as beneath the Mariana trench (34). These two kinds of slab structure are the same as those from our models (Figs. 3 and 4). Grand (35) has recently shown that the subducted slab under western North America can be clearly traced to the core-mantle boundary, with a structure nearly identical to that seen in Fig. 3C. The cold slabs spreading over the core-mantle boundary may produce the large-scale structure in the lowermost mantle, as observed with the use of core-reflected shear waves (36).

Our models strongly suggest an intimate relation between slab penetration, slab dip, and migration of converging margins. A subducted slab with a steep dip angle is more likely to be associated with a relatively stationary trench, and the subducted slab is more likely to extend into the lower mantle (Fig. 3C). On the other hand, a subducted slab with a shallow dip is more likely to be associated with a rapid retrograde trench migration, and the subducted slab is less likely to have penetrated into the lower mantle (Figs. 3A and 4B). This relation between slab morphology and trench migration is similar to the correspondence between slab morphology revealed through seismic tomography studies and plate reconstruction (7). Seismic tomography of western Pacific subduction zones indicate that the steeply dipping Mariana slab may extend into the lower mantle, whereas the relatively shallow-dipping Izu-Bonin and central Japan slabs may be deflected above the 670-km discontinuity (34). Plate reconstruction suggests that, in the last 45 My, the Mariana trench has not had significant retrograde migration whereas the Izu-Bonin trench and the Japan trench have migrated

oceanward significantly (37). Laboratory studies (38) also suggested that subducted slabs with shallower dips have a smaller propensity of penetration into a lower layer with both an increased density and viscosity; phase changes cannot be included in laboratory models.

The shallow-dipping slabs under the continents are caused by oceanward continental motion or trench migration (6, 10) (Figs. 3A and 4B). This result is consistent with the observation of a shallow-dipping slab under the Andes and supports the hypothesis of slabs with shallow dips under North America during the Cretaceous (39). Our models show that slabs have shallower dips and extend farther landward when the oceanward continental motion (that is, trench migration velocity) is greater (Figs. 3A and 4B). With changes in the dip angles and the length of slabs under continents, variations in continental motion may have significant effects on continental tectonics that are related to subduction, including topography (20), subsidence, and the evolution of major basins on active continental margins (39). However, exceptions to the simple relation between slab dip angles and trench migration velocity may be found in some other subduction zones. For example, the Solomon trench currently is migrating oceanward at a fast rate, but the slab dip underneath the Solomon trench is steep (40). Cumulative trench migration in the past may be more important than present values in influencing slab dip.

## REFERENCES AND NOTES

1. A strong and periodic mantle overturn has been observed in isoviscous flow models with no plates [S. Honda *et al.*, *Science* **259**, 1308 (1993); P. J. Tackley *et al.*, *Nature* **361**, 699 (1993); S. A. Weinstein, *Geophys. Res. Lett.* **20**, 101 (1993)]. This process is quite similar to that first discussed by U. R. Christensen and D. A. Yuen [*J. Geophys. Res.* **90**, 10291 (1985)].
2. W. R. Peltier and L. P. Solheim, *Eos* (spring suppl.) **74**, 47 (1993); M. Stein and A. W. Hofmann, *Nature* **372**, 63 (1994).
3. G. F. Davies and M. Richards, *J. Geol.* **100**, 151 (1992).
4. J. W. Morgan, *Tectonophysics* **94**, 123 (1983).
5. P. Molnar and T. Atwater, *Earth Planet. Sci. Lett.* **41**, 330 (1978).
6. T. A. Cross and R. H. Pilger, *Geol. Soc. Am. Bull.* **93**, 545 (1982).
7. R. D. van der Hilst and T. Seno, *Earth Planet. Sci. Lett.* **120**, 395 (1993).
8. G. F. Davies, *J. Geophys. Res.* **93**, 10451 (1988).
9. M. Gurnis, *Geophys. Res. Lett.* **16**, 179 (1989).
10. \_\_\_\_\_ and B. H. Hager, *Nature* **335**, 317 (1988).
11. Weak zones of such a size may lead to unrealistic time-dependent variations in slab structure seen in the models with mobile weak zones (10) and in fixed-margin models.
12. S. Zhong and M. Gurnis, *J. Geophys. Res.* **99**, 15683 (1994).
13. \_\_\_\_\_, in preparation.
14. A mantle convection formulation in a cylindrical geometry [S. Zhong and M. Gurnis, *J. Geophys. Res.* **98**, 12219 (1993)] is used because of the natural curvature and periodic boundary conditions in this formulation. We are unable to exploit the latter feature in a full cylinder at this time because

of computational limitations.

15. The fault parameterization used in the convection formulation is based on the ability to match present-day dependence of oceanic trench depth versus plate age, angle of decoupling zone, and slab dip angle (12). These relations can best be fit if the fault or decoupling zone completely penetrates the thickness of the lithosphere. Using the formulation in (12), we find that as the depth at which the fault extends decreases, (i) surface trench topography progressively becomes broader, resembling the topography obtained above a slab sinking in a simply viscous media and (ii) the horizontal surface velocity becomes less plate-like.
16. Similar rheological structure of the mantle was proposed by S.-i. Karato and P. Wu [*Science* **260**, 771 (1993)] and was used in mantle convection studies [A. P. van den Berg *et al.*, *Geophys. J. Int.* **115**, 62 (1993)].
17. B. H. Hager, *Eos* (fall suppl.) **71**, 1567 (1990).
18. In traditional convection models [for example, Christensen and Yuen (7) and Tackley *et al.* (1)], subduction is symmetric, that is, surface fluids on both sides of a converging margin sink into the deep mantle, and this is inconsistent with the observation that subduction is one-sided.
19. The model heat flux decays with lithospheric age with a function close to that from the half-space model. For Fig. 3C, the function that best fits the heat flux and age is  $444.3(\text{age})^{-0.47}$  [the function for the half-space model is  $290.9(\text{age})^{-0.5}$ ]. The model sea-floor elevation decreases with age but with a smaller slope than that from the half-space model and the observed values [J. G. Scalter *et al.*, *Rev. Geophys.* **18**, 269 (1980)]. For Fig. 3C, the maximum difference in topography between the half-space model and our model is about 1 km.
20. M. Gurnis, *Geology* **21**, 29 (1993).
21. The ratio of heat flux on continents to that on ocean floors, which ranges from 0.3 to 0.5, is consistent with the observed ratio of reduced heat flux on continents to heat flux on ocean floors [H. Pollack, *Annu. Rev. Earth Planet. Sci.* **10**, 459 (1982)]. The average heat flux is about half of the observed value [H. Pollack *et al.*, *Rev. Geophys.* **31**, 267 (1993)]. The small heat flux may be due to the small superadiabatic temperature contrast in the models.
22. C. W. Wicks Jr., and M. A. Richards, *Science* **261**, 1424 (1993).
23. U. R. Christensen and D. A. Yuen, *J. Geophys. Res.* **89**, 4389 (1984).
24. S. Zhong and M. Gurnis, *ibid.* **99**, 15903 (1994).
25. D. L. Anderson, *Nature* **297**, 391 (1982).
26. M. Gurnis, *ibid.* **332**, 695 (1988).
27. C. G. Chase, *ibid.* **292**, 464 (1979).
28. B. H. Hager and R. W. Clayton, in *Mantle Convection*, W. R. Peltier, Ed. (Gordon and Breach, New York, 1989), pp. 657-763.
29. M. Gurnis and T. H. Torsvik, *Geology* **22**, 1023 (1994).
30. W. M. Elsasser, *J. Geophys. Res.* **76**, 1101 (1971).
31. The functional form of the thermal anomaly  $T_a$  is given by
 
$$T_a = \Delta T \cdot \cos \left[ \frac{\pi}{2} \left( 3 + \frac{\theta - \theta_{mc}}{\pi - \theta_{mc}} \right) \right] \sin \left[ \pi \left( \frac{r - r_i}{r_{lm} - r_i} \right) \right]$$
 for  $\theta_{mc} < \theta < \pi$  and  $r_i < r < r_{lm}$ , where  $\Delta T$  is the maximum anomaly,  $r_{lm}$  is the radial coordinate at the top of the lower mantle,  $r_i$  is the radial coordinate at the bottom of the lower mantle, and  $\theta_{mc}$  is the azimuthal coordinate at the middle of the continent. A 100-K increase in temperature in the lower mantle can be caused by continental insulation within 400 My, assuming that all the heat blanketed by a continent is used to heat the lower mantle beneath the continent. This estimate of temperature is close to that of Gurnis and Torsvik (29).
32. H. Zhou, D. L. Anderson, R. W. Clayton, *J. Geophys. Res.* **95**, 6799 (1990).
33. T. Fukao, M. Obayashi, H. Inoue, M. Nishii, *ibid.* **97**, 4809 (1992).
34. R. D. van der Hilst, E. R. Engdahl, W. Spakman, G. Nolet, *Nature* **353**, 37 (1991).
35. S. P. Grand, *J. Geophys. Res.* **99**, 11591 (1994).

36. M. E. Wyssession, L. Bartko, J. B. Wilson, *ibid.*, p. 13667.  
 37. T. Seno and S. Maruyama, *Tectonophysics* **160**, 23 (1989).  
 38. C. Kincaid and P. Olson, *J. Geophys. Res.* **92**, 13832 (1987).  
 39. T. A. Cross and R. H. Pilger, *Nature* **274**, 653 (1978).  
 40. R. D. Jarrard, *Rev. Geophys.* **24**, 217 (1986).  
 41. This work is based in part on one chapter of S.Z.'s

thesis (University of Michigan, 1994). (S.Z. thanks G. Hulbert, H. Pollack, L. Ruff, and K. Satake). Funded by the David and Lucile Packard Foundation and NSF grant EAR-9496185. S.Z. is supported by a Texaco Fellowship at the California Institute of Technology. Some computations were performed at the Pittsburgh Supercomputer Center. Contribution 5451 of the Division of Geological and Planetary Sciences of the California Institute of Technology.

## Methodology

Except for the Japanese example below, all text shown here has been reduced to a 27-character alphabet (uppercase A through Z plus space). The alphabet size only weakly affects the ultimate outcome (22), so there is no loss of generality; consistent results are also obtained for a range of  $n$ -gram lengths (22). I have worked with 5-grams for the English language examples, and 6-grams for the Japanese (23), but it should be borne in mind that the size of the alphabet and the  $n$ -gram length are both flexible.

Given an alphabet and value of  $n$ , a naive calculation of the number of possible  $n$ -grams can be misleading. It is immaterial that, for example,  $27^5 = 14,348,907$  distinguishable 5-grams can be formed using 27 characters because most of them are never encountered. Huge reserves of computer memory for  $n$ -gram statistics are therefore unnecessary.

An entire document can be represented as a vector whose components are the relative frequencies of its distinct constituent  $n$ -grams (the exhaustive list of constituent  $n$ -grams comprises all  $n$ -character sequences produced by an  $n$ -character-wide window displaced along the text one character at a time, and contains many duplications). Let the document contain  $J$  distinct  $n$ -grams, with  $m_i$  occurrences of  $n$ -gram number  $i$ . Then the weight assigned to the  $i$ th vector component will be

$$x_i = \frac{m_i}{\sum_{j=1}^J m_j} \quad (1)$$

where

$$\sum_{j=1}^J x_j = 1 \quad (2)$$

Because both the size of the alphabet and the length of the  $n$ -grams are arbitrary, document vectors can be stored conveniently by indexing ["hashing" (24)] each  $n$ -gram in a consistent manner; numerical values of vector components are stored and retrieved using these indices as pointers to memory. For the present work, I have used an 18-bit index (hash key) and ignored collisions (relatively infrequent instances of different  $n$ -grams being mapped to the same key).

Documents are characterized as follows: (i) Step the  $n$ -gram window through the document, one character at a time. (ii) Convert each  $n$ -gram into an indexing key. (iii) Concatenate all such keys into a list and note its length. (iv) Order the list by key value [efficient algorithms will do this in linear time (25)]. (v) Count and store the number of occurrences of each distinct key while removing duplicates from the list. (vi) Divide the number of occurrences of each

# Gauging Similarity with $n$ -Grams: Language-Independent Categorization of Text

Marc Damashek

A language-independent means of gauging topical similarity in unrestricted text is described. The method combines information derived from  $n$ -grams (consecutive sequences of  $n$  characters) with a simple vector-space technique that makes sorting, categorization, and retrieval feasible in a large multilingual collection of documents. No prior information about document content or language is required. Context, as it applies to document similarity, can be accommodated by a well-defined procedure. When an existing document is used as an exemplar, the completeness and accuracy with which topically related documents are retrieved is comparable to that of the best existing systems. The results of a formal evaluation are discussed, and examples are given using documents in English and Japanese.

I report here on a simple, effective means of gauging similarity of language and content among text-based documents. The technique, known as Acquaintance, is straightforward; a workable software system can be implemented in a few days' time. It yields a similarity measure that makes sorting, clustering, and retrieving feasible in a large multilingual collection of documents that span an unrestricted range of topics. It makes no use of words per se to achieve its goals, nor does it require prior information about document content or language. It has been put to practical use in a demanding government environment over a period of several years, where it has demonstrated the ability to deal with error-laden multilingual texts.

Sorting and categorizing the enormous amount of text now available in machine-readable form has become a pressing problem. To complicate matters, much of that text is imperfect, having been derived from existing paper documents by means of an error-prone scanning and character recognition process.

Over the past few decades, many document categorization and retrieval methods [for example, (1-3) and references therein] have relied on the self-evident utility of

words, sentences, and paragraphs for sorting, categorizing, and retrieving text (4), and various means of suppressing uninformative words, removing prefixes, suffixes, and endings, interpreting inflected forms, and performing related tasks have been developed. Depending on the application, these methods share a number of potential drawbacks: They require a linguist (or a polyglot) for initial setup and subsequent tuning, they are vulnerable to variant spellings, misspellings, and random character errors (garbles), and they tend to be both language-specific and topic-specific.

A potentially more robust alternative, the purely statistical characterization of text in terms of its constituent  $n$ -grams (sequences of  $n$  consecutive characters) (5, 6), has sporadically been applied to textual analysis and document processing (7). Recent examples include spelling and error correction (8-14), text compression (15), language identification (16, 17), and text search and retrieval (18-21).

The literature offers no convincing evidence of the usefulness of either approach for the purpose of categorizing text according to topic in a completely unrestricted multilingual environment, that is, an environment that encompasses many different documents containing a nonnegligible number of character errors. The present paper is intended to provide such a demonstration.

The author is at the Department of Defense, Fort George G. Meade, MD 20755-6000, USA.