Degree-1 mantle convection and the crustal dichotomy on Mars

Shijie Zhong*, Maria T. Zuber

Abstract

The surface of Mars consists of an old, heavily cratered and elevated southern hemisphere and younger, resurfaced and depressed northern hemisphere, a feature often termed the crustal dichotomy. The global crustal structure [Zuber et al., 2000] revealed by topography and gravity data from the Mars Global Surveyor spacecraft, and the possible late formation of the boundary zone between the hemispheres [McGill and Dimitriou, 1990], have been proposed to indicate an endogenic origin for the dichotomy. However, degree-1 mantle convection that is required for any endogenic process to be viable cannot be produced with conventional mantle convection models [Schubert et al., 1990]. We have studied the role of radially stratified viscosity on mantle deformation by using Rayleigh–Taylor instability analyses in a spherical shell geometry. Our analyses reveal that when mantle viscosity is stratified with a weak asthenosphere, deformation at long wavelengths is more efficient than that at short wavelengths. The weaker the asthenosphere, the longer the wavelength at which the deformation is the most efficient. A thicker asthenosphere also favors the deformation at long wavelengths. Both the Rayleigh–Taylor instability analyses and numerical modeling of mantle convection show that degree-1 convection can be produced within the Martian mantle provided that the mantle had a weak asthenosphere (~500 km thick and ~10^2 times weaker than the underlying mantle) early in planetary history. The degree-1 convection causes preferential heating of one hemisphere that may explain the primary features associated with the dichotomy in crustal structure. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: mantle; convection; Mars; mantle rheology; crust

1. Introduction

Since the Martian crustal dichotomy was first identified by the Mariner 9 spacecraft [4], both endogenic [5,6] and exogenic [7,8] processes have been proposed to explain its origin. Exogenic processes include single [7] and multiple [8] hemispheric-scale impacts. An impact-related mechanism implies that the dichotomy is a primordial feature [2] formed when large impactors in planet-crossing orbits were still present in the post-accretional solar system. However, an impact origin may be difficult to reconcile with recent Mars Global Surveyor (MGS) geophysical observations that indicate pole-to-pole gradual variations in topography [9] (Fig. 1a) and crustal thickness [1] and a mismatch between the surfacial expression...
of the dichotomy and the crustal structure (Fig. 1b) [1]. In addition, geological mapping suggests that at least part of the boundary formed later in Martian history (Late Noachian or Early Hesperian) [2,10] than would be expected if produced by a mega-impact or impacts.

Endogenic processes could conceivably satisfy the observed geological and geophysical constraints. An early study suggested on the basis of qualitative arguments that the thinner crust in the northern hemisphere could be a consequence of sub-crustal erosion caused by unspecified degree-1 mantle convection [5]. More recently, Sleep [6] proposed that the dichotomy boundary may
represent relic boundaries associated with plate tectonics in the northern hemisphere. The confirmation of the impact origin of the 1500-km diameter Utopia structure [2] in the northern hemisphere from the MGS topography [9] and gravity [11] observations suggest that if the hemispheric difference was a consequence of plate tectonics, it must have originated early in solar system history. The recent discovery of zones of alternating magnetization in the southern hemisphere [12] has been interpreted as evidence of spreading of the early Martian crust [13]. However, the crust in this region does not have a structure that is obviously suggestive of plate tectonics [1] and other tectonic processes, such as magmatic intrusion [14], may alternatively explain the magnetic lineations. On the other hand, if the average crustal thickness is $V50$ km [1], it is unclear whether mantle convection could result in subduction of such a thick crust, given that the basalt-eclogite phase transition, which aids the subduction, does not occur until a depth of $\sim200$ km.

In order for an endogenic process (such as but not restricted to plate tectonics) to be a viable mechanism to explain a global feature like the crustal dichotomy, planet-scale (degree-1) mantle convection is necessary [3]. However, mantle convection models with either isoviscous [3] or temperature-dependent viscosity [15,16] structure do not produce degree-1 convection. Core formation may result in degree-1 convection [3,17], but if this process produced the dichotomy, then evidence for early core formation on Mars [18] would also imply a primordial origin for the dichotomy. Such a scenario would require additional mechanisms to explain both the resurfacing of the northern hemisphere [19] and the relatively young age for at least part of the dichotomy boundary [2,11]. The dynamics of solid-state phase changes may produce a single thermal upwelling plume (i.e., approximately degree-1 feature), which has been used to explain the Tharsis volcano-tectonic province on Mars [20,21]. However, a single upwelling plume can only form after $2\times10^9$ years [20,21], which may be too late to explain the formation of dichotomy [2]. In this study we explore endogenic processes that lead to degree-1 convection that can explain the general characteristics of the global topography and crustal structure.

2. Physical models

To study the generation of degree-1 structure, we have employed two different approaches: Rayleigh–Taylor instability analyses and finite element models of thermal convection. We describe our model approaches in this section.

2.1. A Rayleigh–Taylor instability analysis

A Rayleigh–Taylor instability analysis may provide important insights into the conditions under which degree-1 convection may occur [22]. Our analytic methods are based on a recently developed model for viscoelastic stress relaxation in a spherical shell geometry [23]. For our Rayleigh–Taylor instability analysis, the mantle is considered to be purely viscous. In addition, the mantle is assumed incompressible with two layers of different viscosity and density. The dynamics is determined by the equations of continuity and motion:

$$\nabla \cdot \mathbf{u} = 0$$  \hspace{1cm} (1)

$$-\nabla P + \nabla \cdot [\eta (\nabla \mathbf{u} + \nabla^T \mathbf{u})] - \rho g \mathbf{e}_r = 0$$  \hspace{1cm} (2)

where $\mathbf{u}$ is the velocity, $P$ is the pressure, $\eta$ is the viscosity, $g$ is the gravitational acceleration, $\mathbf{e}_r$ is the unit vector in vertical direction, and $\rho$ is the density. The surface and core–mantle boundary (CMB) are subjected to stress-free boundary conditions. Similar analyses in a Cartesian geometry are used in studying the structure of salt domes [24,25]. The density interface is at $r = r_d$. For the upper layer with $r > r_d$, $\rho = \rho_u$ and $\eta = \eta_u$. For $r < r_d$, $\rho = \rho_l$ (i.e., 3400 kg m$^{-3}$ in Table 1) and $\eta = \eta_l$. If $\rho_l < \rho_u$, then the two-layer system is unstable. If a perturbation at a wavelength or spherical harmonic degree is introduced to the topography of the density interface, the topography will grow with time as the instability develops. For our models with three deformable density interfa-
ces including the surface and the CMB, the time evolution of the interfacial topography $H(l,t)$ at harmonic degree $l$ and at $r = r_d$ is controlled by the sum of three exponential functions with different characteristic time $\tau_{li}$ ($i = 1, 2$ and $3$) [23]

$$H(l,t) = \sum_{i=1}^{3} h_{li} \exp(t/\tau_{li})$$ (3)

where $h_{li}$ is the eigenfunction. The characteristic time $\tau_{li}$ depends on the density and viscosity structures of the two-layer model [23]. The analytic techniques to determine the characteristic times and eigenfunctions are given in [23]. For $\rho_l < \rho_u$, there is one positive characteristic time $\tau_{lo}$ that corresponds to the unstable mode. The growth rate of the instability $q(l)$ for degree $l$ is:

$$q(l) = h_{lo}/\tau_{lo}$$ (4)

The assumption of the unstable density stratification is rather arbitrary and is invoked mainly to simplify our stability analyses. How the mantle evolves to give rise to unstable conditions is addressed by our models of thermal convection.

2.2. Finite element models of thermal convection with variable viscosity

A Rayleigh–Taylor instability analysis assumes infinitesimal deformation and does not include the energetics. To fully understand the development of degree-1 mantle structure, it is important to formulate numerical models of thermal convection. In addition to the conservation equations of mass and momentum, Eqs. 1 and 2, we also include the equation of energy balance in our models of thermal convection:

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \kappa \nabla^2 T + Q$$ (5)

where $T$ is the temperature, $\kappa$ is the coefficient of thermal diffusion, $Q$ is the rate of internal heat generation. We assume that the Martian mantle has the same concentration of radioactive elements as that for the Earth’s mantle (Table 1) [24]. We also consider time dependence for the internal heating that results from radioactive decay [24].

In contrast to our Rayleigh–Taylor instability analyses, we assume that the mantle has a homogeneous intrinsic density and that the only source of density anomalies is from thermal expansion:

$$\delta \rho = \rho_0 \alpha (T - T_0)$$ (6)

which drives the mantle flow. In Eq. 6, $\rho_0$ and $T_0$ are reference density and temperature, and $\alpha$ is the thermal expansion coefficient. We will use depth- or/and temperature-dependent viscosity:

$$\eta = A(r) \exp((E + PV)/RT)$$ (7)

where $E$ and $V$ are activation energy and volume, respectively (Table 1), $P$ is the pressure, $R$ is the gas constant, and $A(r)$ determines the radial dependence of viscosity. Our models include an 80-km thick high viscosity lid that does not change with time. Details of viscosity structure are discussed in Section 3.

The surface and CMB are subjected to free-slip and isothermal boundary conditions. With the Boussinesq approximation, we only consider the super-adiabatic temperature in our models. The
initial temperature is assumed to be equal to the solidus temperature (1160°C) at a depth of 80 km and to decrease linearly to the surface temperature at the surface. Below the depth of 80 km, the mantle temperature is initially the same as that at the depth of 80 km. The super-adiabatic temperature at the CMB is fixed at 1320°C. Superimposed on this radial temperature profile is a randomly perturbed temperature with very small amplitude (<0.1 K). When the adiabatic temperature gradient is considered, temperature at the CMB is 1677°C (or 1950 K). This temperature may be consistent with the estimates of melting temperature for the core, which depends on the core sulfur content [26]. For the calculations with temperature-dependent viscosity, the viscosity is determined by the total temperature that includes the adiabatic temperature.

Our models assume a spherically axisymmetric geometry. We use a finite element approach to solve Eqs. 1, 2, and 5 for time-dependent thermal convection [27]. We use 81 and 181 grid points in the radial and azimuthal directions, respectively. Numerical grids are significantly refined near the thermal boundary layers. Numerical tests with higher resolution demonstrate that our models are well resolved.

3. Results

3.1. Rayleigh–Taylor instability analyses

In a previous study on the origin of hemispheric asymmetry of lunar mare basalts (i.e., a degree-1 structure), we have demonstrated that hydrodynamic instabilities for a planet with a small core relative to planetary radius ($R_{\text{core}}/R_{\text{planet}} < 0.15$) can lead to degree-1 mantle convection [22]. However, the relatively large core for Mars ($0.4 < R_{\text{core}}/R_{\text{planet}} < 0.6$) as implied from recent measurement of the moment of inertia [28] dictates that a different explanation is required for the Martian degree-1 convection. Here we apply Rayleigh–Taylor instability analyses to a two-layer mantle model to investigate the effects of radially stratified viscosity on flow structure.

With parameters appropriate for Mars and a thickness of 750 km (i.e., $r_d = 2650$ km) for the top layer, a perturbation at harmonic degree $l$ is introduced as the topography for the density interface (the amplitude of the perturbation is the same for all harmonic degrees). The growth rate for the instability is calculated for each harmonic and is then normalized by the maximum growth rate found for all harmonic degrees. The maximum growth rate occurs at a spherical harmonic degree that decreases with increased viscosity for the bottom layer (Fig. 2a). Degree-1 becomes the most unstable wavelength (MUW) when the viscosity contrast is greater than 100 (Fig. 2a). Similar effects of stratified viscosity on the MUW were observed previously [25,22,29]. Our results indicate that as the viscosity of the lower layer increases deformation is more efficient at longer wavelengths.

![Fig. 2](image-url)

Fig. 2. (a) Dependence of normalized growth rate on spherical harmonic degrees for a two-layer mantle with different viscosity contrast $\eta_u/\eta_l$ between the upper and lower layers. The radius for the boundary between the upper and lower layer $R_d$ is 2650 km. (b) Dependence of the MUW in terms of spherical harmonic degree on $\log_{10}(\eta_u/\eta_l)$ and $R_d$. 
Fig. 2b shows the dependence of the MUW on the viscosity contrast and layer thickness (or \( r_d \)). For an isoviscous mantle, the longest possible MUW is degree-4 and it occurs when the two layers are approximately equal in thickness. This behavior partially explains why thermal convection models for isoviscous flow cannot produce degree-1 structure [3]. For isoviscous flow, when a layer becomes too thin, the resisting force from either the surface or the CMB tends to promote short-wavelength deformation. As the bottom layer becomes more viscous, the MUW increases and occurs at degree-1 over a wide parameter space (Fig. 2b). Increasing the viscosity for the top layer also favors long-wavelength structure, and degree-1 may become the MUW when the top layer is sufficiently more viscous than the bottom layer (Fig. 2b). These analyses indicate that degree-1 structure may be produced with stratified viscosity structure over a wide range of model parameters.

### 3.2. Time-dependent thermal convection calculations

The Rayleigh–Taylor instability analyses suggest that viscosity stratification may be the key to producing degree-1 mantle convection (Fig. 2b). Our thermal convection models are used to investigate how degree-1 mantle convection may develop in a dynamically self-consistent manner. The following observations for the Earth’s mantle prompt us to focus on models for Mars in which the mantle viscosity increases with depth. (1) The viscosity in the asthenosphere may be 300 times smaller than that in the lower mantle, as inferred from studies of the Earth’s geoid [30]. The inferred viscosity contrast needs to be \( \sim 600 \) if surface tectonic plates are considered in the geoid models [31]. An even weaker asthenosphere is inferred from studies of earthquake cycles and post-seismic rebound [32] and remote triggering of earthquakes [33]. (2) Observations of seismic anisotropy in the upper mantle (< 300 km in depth) and studies on silicate rheology suggest that mantle deformation at depths smaller than 200 to 300 km is controlled by dislocation creep [34,35], which can substantially weaken the flow with stress. A relatively weak shallow layer may also result from the depth dependence of homologous temperature [24] and partial melting.

We first present three cases to demonstrate the

![Fig. 3. Characteristic thermal structure and flow field for Cases 1 (a), 2 (b), and 3 (c) with different \( \eta_u/\eta_i \). \( R_d \) is 2650 km. The times in (a), (b) and (c) are 4690, 1710, and 450 Ma, respectively. To highlight the velocities in the more viscous lower layer, the amplitudes of velocities are plotted as their square roots. The viscosity for the 80-km thick lid is 20000\( \eta_i \).](image-url)
control of stratified viscosity on the development of degree-1 convection. All three cases have an 80-km thick high viscosity lid, and their mantle viscosity structures are assumed to be time-invariant and only dependent on the radius. Case 1 has an isoviscous mantle with a viscosity of $5 \times 10^{21}$ Pa s. The characteristic thermal structure for Case 1 has wavelengths comparable to the thickness of the mantle (Fig. 3a), consistent with previous studies [3] and predictions from our Rayleigh–Taylor instability analysis for isoviscous mantle (Fig. 2). In Case 2, the viscosity from the base of the lid to a depth of 750 km is reduced to $10^{20}$ Pa s. Considering the difference in gravitational acceleration between Mars and the Earth, the depth of 750 km may correspond to the depths for the asthenosphere and the proposed change in deformation mechanism from dislocation creep to diffusion creep in the Earth’s upper mantle [34,35]. However, our numerical experiments and the Rayleigh–Taylor instability analysis (Fig. 2b) indicate that our conclusions do not depend significantly on the choice of this depth.

The incorporation of a weak layer leads to two different scales in thermal structure: a short-wavelength structure in the weak top layer and a long-wavelength structure associated with the more viscous bottom layer (Fig. 3b). This is consistent with previous studies applied to terrestrial mantle flow with a slightly smaller viscosity contrast [36,37]. Further reduction in viscosity for the top layer to $10^{19}$ Pa s in Case 3 results in a degree-1 flow and thermal structure with one hemisphere that is significantly hotter than the other (Fig. 3c). The effect of a weak upper layer on the flow structure is thus generally consistent with our Rayleigh–Taylor instability analysis.

The viscosity structure affects not only the thermal structure (Fig. 3a–c), but also the mantle temperature and surface heat flux. Although Cases 1–3 have the same high viscosity lid, the averaged heat flux for the first $1.5 \times 10^9$ years is larger for models with a weaker upper layer (for Cases 1, 2, and 3 the heat flux are 25, 37, and 41 mW m$^{-2}$, respectively). The enhanced cooling in such models results in lower mantle temperatures. For Cases 1, 2, and 3, averaged temperatures for the mantle (including lithosphere) for the first $1.5 \times 10^9$ years are 1790, 1717, and 1651 K, respectively. Our results thus not only explain the degree-1 structure, but may also suggest an important role of the length scales of mantle structure on heat transfer. The latter point is particularly interesting given that many parameterized mantle convection models that in general do not explicitly consider the effects of wavelengths of mantle flow on heat transfer predict an overheated Martian mantle [38]. However, our models with layered viscosity structure do not consider the thickening of lithosphere with cooling.

Mantle rheology may depend on temperature and pressure [34]. The effects of this rheology are considered in Case 4. A relatively small activation energy is used (Table 1), because it is consistent with observations of flexural rigidity on Earth’s seafloor [39]. The pre-exponential constant in the rheological equation (Table 1) is chosen such that the viscosity at the solidus temperature and the mid-mantle depth is $10^{20}$ Pa s. The pre-exponential constant is reduced by a factor of 500 from 80-km depth to 750-km depth. The resultant averaged viscosity structure after 400 Ma is similar to that in Case 3. For Case 4 during the early stage, convection mainly occurs in the weak upper layer with relatively small horizontal scale and temperature anomalies (Fig. 4a). As the lower layer heats up, it induces convection throughout the mantle and leads to degree-1 structure at $\sim 230$ Ma (Fig. 4b). The degree-1 structure can be maintained for over $1 \times 10^9$ years (Fig. 4c). In Case 4, the heat flux out of the core is greater than 5 mW m$^{-2}$ only in the first 100 Ma. As the mantle temperature increases, the core heat flux decreases and eventually becomes negative.

Our numerical experiments show that the formation of degree-1 structure is insensitive to initial and boundary conditions. For example, with the same viscosity structure as in Case 3, similar degree-1 structure to that in Case 3 is produced for models with either insulated bottom boundary conditions or the initial temperature taken from Fig. 3a. Numerical resolution tests confirm the robustness of the results. Due to the computational cost for models with large viscosity contrast (e.g., Case 3), fully 3D spherical models have not been computed. However, based on previous
studies of thermal convection in 2D spherically axisymmetric and 3D spherical models [40] and our Rayleigh–Taylor instability analyses (Fig. 2), we believe that our results will not change significantly in fully 3D models.

4. Discussion and conclusions

Our Rayleigh–Taylor instability analyses and finite amplitude thermal convection calculations demonstrate that hemispherically asymmetric (i.e., degree-1) thermal and flow structures may develop in a dynamically self-consistent manner in the Martian mantle, if the early Martian mantle has a weak asthenosphere (i.e., \( \sim 100 \) times weaker than the underlying mantle; the requirement is less strict for the thickness of asthenosphere). We think that such a weak asthenosphere is very likely, based on our understanding of Earth’s mantle viscosity structure [30,33,34] and the effects of partial melting on mantle viscosity. The degree-1 mantle convection may provide an explanation to the dichotomy in crustal structure revealed by the MGS gravity and topography observations [1]. Our models with a constant thickness lithosphere show that the degree-1 mantle convection enhances the cooling of the Martian mantle, but the significant core heat flux is only possible in the first 100 Ma. The time evolution of degree-1 mantle convection and the core heat flux in our models depend on choices of mantle viscosity, the heat production rate, and initial mantle temperature distribution [38,41]. A mantle with initially smaller temperature or smaller heat production (i.e., possibly as a result of crustal formation that leads to more concentrated radioactive elements in the crust) would lead to larger core heat flux over a longer time. Our models are broadly consistent with early shutdown of the Martian magnetic field suggested by crustal remnant magnetic anomalies [12,13]. In addition, the time scales are consistent with later formation of the dichotomy boundary [2,10], although we recognize that the boundary zone between the smooth northern lowlands and rough southern highlands has also been influenced by other processes such as erosion and perhaps impacts.

We suggest that degree-1 mantle convection may lead to the dichotomy in crustal structure in the following scenario. The hemisphere with the upwelling (e.g., Fig. 3c) would form the north-
ern hemisphere. This scenario requires the lower part of the primordial crust to be deformable, which is likely given the inferred 50-km average crustal thickness [1] that would lead to relatively high temperature in the crust to facilitate creep deformation in the lower crust. Dynamic stresses associated with convection should thin the crust in the hemisphere with the upwelling and thicken the crust in the other hemisphere. Accompanying crustal thinning in the hemisphere above the upwelling should be surface subsidence and volcanism, which would produce volcanic resurfacing. As the crust and lithosphere cool with time, this would strengthen the lower crust to maintain the crustal structure throughout the Martian history. A possible test for this model is to use the MGS data to examine to what extent volcanic resurfacing has affected the northern hemisphere [42].

Alternatively, it is also conceivable that the southern hemisphere may form above the upwelling with excessive crustal materials that are produced by the upwelling and are retained in the southern hemisphere to produce the high elevation there. However, this scenario requires that no significant volcanic resurfacing occur in the southern hemisphere while the crustal materials are produced by the upwelling. Future studies are needed to investigate the dynamic interaction between degree-1 mantle convection and crustal formation and deformation and to document geological and geophysical evidence to distinguish between these competing scenarios.

Our models for degree-1 thermal convection may also have implications for the formation of supercontinents including the Pangea and Rodinia on the Earth. Because mantle flow is significantly influenced by surface plates [43], supercontinents and the distribution of subduction zones during the Pangea [44] imply a degree-1 structure of mantle flow. While supercontinents may form as a result of continental collisions as suggested by models of thermal convection in a 2D Cartesian geometry [45], supercontinents may also be attributed to the degree-1 thermal convection in the mantle. Our models suggest that such degree-1 thermal convection may be generated if the Earth’s asthenosphere is sufficiently weak or/and thick. Supercontinents and degree-1 mantle structure may break up because the Earth’s mantle convection is more vigorous. This leads to a periodicity in the formation of supercontinents. Future studies are needed to explore the model parameters for which degree-1 thermal convection can occur for the Earth’s mantle.

Acknowledgements

We thank E.M. Parmentier and S. Karato for helpful discussions, L. Kellogg for permission to use the finite element code, and O. Aharonson for assistance with Fig. 1. We also thank H. Harder and F. Nimmo for constructive reviews. This research was supported by the NASA Planetary Geology and Geophysics Program and the NASA Mars Global Surveyor Project.[SK]

References
