On the relative importance of thermal and chemical buoyancy in regular and impact-induced melting in a Mars-like planet

final version

9 July 2017

published in:


1 Institute of Planetology, Westfälische Wilhelms-Universität, Münster, Germany

2 Institute of Planetary Research, German Aerospace Center (DLR), Berlin, Germany

An edited version of this paper was published by AGU. Copyright (2017) American Geophysical Union. That version of record is available at [http://dx.doi.org/10.1002/2016JE005221](http://dx.doi.org/10.1002/2016JE005221)
On the relative importance of thermal and chemical buoyancy in regular and impact-induced melting in a Mars-like planet

Thomas Ruedas*
Institute of Planetology, Westfälische Wilhelms-Universität Münster, Germany
Institute of Planetary Research, German Aerospace Center (DLR), Berlin, Germany

Doris Breuer
Institute of Planetary Research, German Aerospace Center (DLR), Berlin, Germany

Key points

- Compositional buoyancy stabilizes melting-induced anomalies at the base of the lithosphere.
- The anomalies may be detectable by gravimetry but not by seismics with a sparse station network.
- Neglecting impact-induced mantle anomalies results in overestimation of the local crustal thickness.

Abstract

We ran several series of two-dimensional numerical mantle convection simulations representing in idealized form the thermochemical evolution of a Mars-like planet. In order to study the importance of compositional buoyancy of melting mantle, the models were set up in pairs of one including all thermal and compositional contributions to buoyancy and one accounting only for the thermal contributions. In several of the model pairs, single large impacts were introduced as causes of additional strong local anomalies, and their evolution in the framework of the convecting mantle was tracked. The models confirm that the additional buoyancy provided by the depletion of the mantle by regular melting can establish a global stable stratification of the convecting mantle and throttle crust production. Furthermore, the compositional buoyancy is essential in the stabilization and preservation of local compositional anomalies directly beneath the lithosphere and offers a possible explanation for the existence of distinct, long-lived reservoirs in the martian mantle. The detection of such anomalies by geophysical means is probably difficult, however; they are expected to be detected by gravimetry rather than by seismic or heat flow measurements. The results further suggest that the crustal thickness can be locally overestimated by up to $\sim 20 \text{ km}$ if impact-induced density anomalies in the mantle are neglected.

1 Introduction

In spite of the obvious and well-known geological evidence for their extent and ubiquity, the connection between meteorite impacts and the convective processes that shape the thermal and chemical evolution of planetary interiors still tends to be underappreciated. Among the first numerical studies of mantle convection that considered the effects of large impacts are those by Reese et al. (2002, 2004), which investigate how a basin-forming impact could introduce a strong thermal anomaly and trigger regional, potentially long-lived convection in a mantle that would otherwise convect only sluggishly. These authors already pointed to the potential of such an impact to seed a long-lived, stable superplume or other sort of thermal anomaly that feeds a major volcanic center, in their case the Tharsis province on Mars, as proposed earlier on the basis of geological evidence (e.g., Schultz and Glicken, 1979). They also observed that impacts entail strong transient increases in melt production that result in the formation of thickened crust in the area affected by the impact and proposed that this may explain the areoid and topography of Tharsis.

Reese et al. (2002, 2004) started their simulations with the impact into a mantle in a simple, symmetric state assumed to represent pre-impact conditions and described the heating and geometry of the impact as summarized by Melosh (1989). In a more recent study, Watters et al. (2009) have refined the physical model of the impact, in particular with regard to the question how the depth-dependent properties of the target body affect shock propagation and, as a consequence, heating. They also account for the fact that the isobaric core of the impact lies at a certain depth below the pre-impact surface. Their models demonstrate the dynamics of impact-induced plumes: while small events are obliterated quickly in the general flow of the mantle, large impacts cause a strong localized upwelling that results in a large anomaly that spreads beneath the lid and can sweep away cold downwellings. The upwelling induces a flow field that may attract plumes from its neighborhood and merges them, but on the other hand, the anomaly may reduce the convective vigor in the mantle as a whole. Roberts and Arkani-Hamed (2012) have extended this approach.

*Corresponding author: T. Ruedas, Institute of Planetology, Westfälische Wilhelms-Universität, Münster, Germany (t.ruedas@uni-muenster.de)
to the effects of multiple basin-forming impacts and also studied the effects of impact size and initial condition in some more detail. \cite{Golabek2011} investigated the effects of an even larger, dichotomy-forming impact, but their focus lies more on core formation and the very earliest stages of martian evolution.

Except for \cite{Reese2004} and \cite{Golabek2011}, the previous workers did not consider compositional effects of impact-induced melting, and the paper by \cite{Reese2004} applies very simplistic models for both the impact, which is part of their initial condition, and the dependence of depletion and melt content on melting. On the other hand, various numerical convection studies of martian mantle evolution did include compositional effects to some extent (e.g., \cite{Schott2001, Kiefer2003, Li2007, Keller2009, Ogawa2011, Ogawa2012, Ruedas2013a, Plessa2014, Breuer2016}, but some of them do so only with regard to heat source concentrations, and none include impacts. In particular, \cite{Schott2001, Ogawa2012}, \cite{Plessa2014} observed that the compositional density variations produced by melting may be crucial in establishing and preserving long-lived chemical heterogeneities in the martian mantle; a similar observation was made by \cite{Rolf2017}, who studied how basin-forming impacts may have altered lunar evolution. Furthermore, the review on the chemical reservoirs in the martian interior by \cite{Breuer2016} includes some examples of simple local structures with anomalous density or viscosity and discusses their evolution, but these structures do not belong to any specific geological category.

In this study we consider both the thermal effects and the compositional changes from regular melting processes and in particular from melting induced by basin-forming impacts of different magnitudes, whereby the compositional changes are derived from a detailed mineralogical model. We compare the relative importance of the individual contributions of temperature and composition to buoyancy on the short-term as well as on the long-term evolution of a planet, which we assume to share many properties with Mars. In addition to the dynamical aspects, we also try to identify the effect of compositional variations on observables, in particular density and seismic velocity.

2 Method

2.1 Convection and mineral physics models

Our method builds upon or extends the approach described in detail in \cite{Ruedas2013a, Ruedas2013b}, so we limit ourselves to a brief overview and describe shortly the more important changes and extensions. The convection code is a modified version of the code STAGYY \cite{Tackley1996, Tackley2008} and solves the equations of conservation of mass, momentum, and energy in the compressible, anelastic approximation on a two-dimensional spherical annulus grid \cite{Hernlund2015} with 512 × 128 points and stress-free, isothermal top and bottom boundaries. Melting and the transport of trace components such as water and radionuclides are modeled with tracer particles, with 50 tracers per cell. In impact sites as well as in any other partially molten region, we extract all melt down to the extraction threshold $\varphi_r$ of 0.7% if the melt is less dense than the matrix and if there is a contiguous pathway of partially molten rock up to the base of the lithosphere. In terms of the redistribution of incompatible components, the use of such a threshold corresponds to the continuous or dynamic melting discussed in detail by \cite{Shaw2000}. As in our previous work, all extracted melt is instantaneously added to the top to build the crust. Coupled to the convection model is a detailed model of the petrological properties of the mantle consisting of a parameterization of experimental phase diagrams of the martian mantle, its melting behavior, and its thermoelastic properties such as the density, from which the buoyancy forces are derived. The thermoelastic properties are derived from the mineralogical model starting from mineral endmembers and assuming ideal mixing; we replaced the polynomial-based parameterization of these properties from \cite{Ruedas2013b} with a formalism based on Mie–Grüneisen–Debye theory and used a database compiled from literature data from numerous experimental studies; a detailed description of several aspects of the mineralogical, rheological, and chemical model is given in the Supporting Information. Moreover, we have now also implemented a petrological model of basalt/eclogite and its properties at a similar level of detail as the peridotite. The calculation of thermal conductivities takes advantage of the new database by \cite{Hofmeister2015}, which covers many of the relevant minerals, and applies effective medium theory to determine the bulk conductivity, compared with the former dataset, it results in somewhat lower conductivities especially in the deeper parts of the mantle. The mineralogical model is also used for calculating melting partitioning coefficients, whereby we attempted to improve the previous work of \cite{Ruedas2013a} by including dependence on pressure, temperature, and composition where applicable and possible. In particular, a better parameterization of iron partitioning is implemented for several mineral phases, which is an important factor in the relation between depletion and compositional buoyancy. Moreover, the lattice strain model is applied to radiomucide partitioning in clinopyroxenes and garnets (e.g., \cite{Wood2014}: a brief summary is given in the Supporting Information. The viscosity of the mantle depends on pressure, temperature, composition, and melt fraction as in the previous work (see Supporting Information). The density of the melt, which needs to be evaluated for several purposes, especially in order to determine whether it is extractable, is now calculated using the new deformable hard-sphere model from \cite{Jing2011, Jing2012} rather than the less versatile model from \cite{Jing2009} used in our previous work. As our petrological and melting model is not yet complete enough to yield the complete composition of the melt, we use the martian crust major oxide composition from \cite{Taylor2009} (Tab. 6.4) as input for the melt density model. Furthermore, we have revised our solidus parameterization using new data (see Appendix A).
The core enters the model as a thermal boundary condition on the interior boundary that represents the core–mantle boundary (CMB) and is described with an energy balance model following Nimmo et al. (2004) and Williams and Nimmo (2004) but with a more elaborate one-dimensional thermodynamical model of its properties that includes the calculation of adiabatic profiles of temperature and various physical properties for the core; the model is largely identical to that used in our previous work, with the difference that we now use material properties directly derived from experimental data not only for liquid iron, but also for liquid Fe–S alloys (Kaiura and Toguri, 1979; Sanloup et al., 2000; Balog et al., 2001, 2003; Nishida et al., 2011).

2.2 Impact shock heating and melt production

The impact process is not included via coupling or interfacing the convection algorithm with a fully dynamical hydrocode simulation but is described using a combination of simpler analytical or semi-empirical approximations. With regard to the energy input from the impact, we follow the approach pioneered by Watters et al. (2009) in many respects. When considering a large impact event such as the one that created Utopia, we start from the known diameter $D_f$ of the final crater and use scaling laws for complex craters to work back to determine the diameter of the transient crater that is formed as an immediate consequence of the impact,

$$D_{tr} = \left(0.8547D_fD_{2c}\right)^{0.131} = 0.87028D_f^{0.885}D_{2c}^{0.115}$$

(Melosh, 2011), and of the impactor,

$$D_{imp} = 0.83 \left(\frac{\rho}{\rho_{imp}}\right)^{0.43} D_{tr}^{1.28} v_{z,imp}^{-0.56} g^{0.28}$$

(e.g., Werner and Ivanov, 2015); the constant pre-factor for the transient crater diameter accounts for the fact that the measured final (complex) crater diameter, which results from the collapse of the transient cavity, is measured from rim to rim rather than at the level of the reference surface and reduces it correspondingly (Melosh, 2011). Here, $v_{z,imp} = v_{imp} \sin \phi$ is the vertical component of the impactor velocity, $D_{2c}$ is the diameter of the transition between simple and complex craters, which is ~5.6 km (global mean from Robbins and Hynek, 2012, Tab. 3) for Mars, $g$ is the acceleration of gravity, and $\rho$ and $\rho_{imp}$ are the densities of the target and the impactor, respectively; a list of the important symbols is given at the end of this paper. The impactor is assumed to be a rocky (S-type) asteroidal object hitting Mars with an absolute velocity of 9.6 km/s (Ivanov, 2001). In the absence of more specific information, the angle $\phi$ is always set to 45°, which is the most probable impact angle for an isotropic impactor flux (e.g., Gault and Wedekind, 1978). In the use of the scaling laws, we do not distinguish between complex craters and multi-ring basins, because it seems that the assumption of proportional growth is valid for the latter as well (Spudis, 1993), and numerical models indicate that many characteristics derived from scaling laws apply to large basins without modification (Potter et al., 2015).

![Figure 1. Key variables of the impact model. Final and transient crater depths are not to scale.](image)

The most extreme conditions prevail in the central part of the impact region, which has been called “isobaric core” by some authors or “near field” by others and which is characterized by the near-constancy of the pressure to which the shocked material is subjected. Many previous workers used empirical equations by Pierazzo et al. (1997) to determine the radius and depth of this region. We follow them in using their formula for the depth of its center below the surface (cf. Figure 1),

$$z_{ic} = 0.1525D_{imp} v_{z,imp}^{0.361}$$

but apply one of the new parameterizations of the recent low-$v_{imp}$ data from Monteux and Arkani-Hamed (2016) by Ruedas (2017) to describe the dependence of shock pressure on distance, because the decay law from Pierazzo et al. (1997) becomes increasingly erroneous at the low $v_{z,imp}$ needed here. These parameterizations have the advantage of describing $p_s(r)$ accurately with a smooth function for all $r$ and without imposing the impedance-match solution at
\( r = 0 \). For an impact of a dunite meteorite onto a dunite target, the “inverse-\( r \)” parameterization by \cite{Ruedas2017} can be written in the form

\[
p_{s}(r) = \frac{a}{b + \left( \frac{2r}{v_{z,imp}} \right)^n}
\]

with

\[
a(v_{z,imp}) = 0.081p_{IM}^{1.476}, \tag{4a}
\]
\[
b(v_{z,imp}) = 0.266v_{z,imp}^{1.161}, \tag{4b}
\]
\[
n(v_{z,imp}) = -0.203 + 1.954 \log v_{z,imp} \tag{4c}
\]

and \( v_{z,imp} \) in km/s. We caution that the coefficients for \( a, b, \) and \( n \) work reasonably well for the \( v_{z,imp} \) value of approximately 6.8 km/s assumed here but may not be suited for much higher \( v_{z,imp} \); the reader is referred to \cite{Ruedas2017} for a detailed discussion.

Previous workers often also used the empirical definition by \cite{Pierazzo2017} for the radius of the isobaric core,

\[
r_{PVM} = 0.225D_{imp} v_{z,imp}^{0.211}, \tag{5}
\]

but as the models from which this formula was derived do not seem to cover the remote parts of the shocked regions for low \( v_{imp} \), we suspect that the quality of eq. 5 deteriorates in those cases. Therefore, we prefer to define the size of the isobaric core by the position of the inflection point of \( p_{s}(r) \):

\[
r_{\text{inf}} = \frac{D_{imp}}{2} \sqrt{\frac{n-1}{n+1} b} \tag{6}
\]

The parameter \( p_{IM} \) is the impedance-match solution, which is used as a scaling factor and is calculated directly from the material properties using the Hugoniot equations as

\[
p_{IM} = \rho(v_{B} + S_{h} u_{IM}) u_{IM}, \tag{7}
\]

with the density of the target, \( \rho \), in g/cm\(^3\), the speed of sound in the target, \( v_{B} \), in km/s, and the Hugoniot slope \( S_{h} = (1 + \gamma)/2 \) being related to the Grüneisen parameter \( \gamma \) \cite{Melosh1989}. The particle velocity \( u_{IM} \) at the interface with the impactor is given by

\[
u_{IM} = -B_{u} + \sqrt{B_{u}^{2} - 4A_{u}C_{u}}
\]

\[
2A_{u}
\]

\( (e.g., \cite{Melosh2011}) \); the properties used here are those of the unshocked material, and the subscript “imp” indicates a property of the impactor. The speed of sound of the target is derived from the thermoelastic properties as given by the mineralogical model for the planet via \( v_{B} = \sqrt{K_{S}/\rho} \), where we use values for \( K_{S} \) and \( \rho \) averaged over the isobaric core volume. The speed of sound in the impactor is simply determined using Birch’s law, \( v_{B,imp} = 2.36 \rho_{imp}^{-1} - 1.75 \) \( (e.g., \cite{Poirier2000}) \), using an assumed density \( \rho_{imp} \) of 2.7 g/cm\(^3\) for mostly rocky (S-type) asteroidal impactors \( (e.g., \cite{Fernández2015}) \).

The propagation of the shock into the target and subsequent isentropic unloading put a certain amount of energy into the target and heat it. Reformulating an estimate by \cite{Gault1963, Watters2009} write this impact heating in terms of shock pressure and lithostatic pressure \( p_{l} \) as

\[
\Delta T = \frac{1}{c_{p}} \left[ \frac{\Delta p}{2\rho S_{h}} \left( 1 - \frac{v_{B}}{v_{B}^{\Phi}} \right)^{2} \Phi^{-1} \right]
\]

with

\[
\Phi = -\frac{2S_{h} \Delta p}{v_{B} \rho} \left( 1 - \frac{4S_{h} \Delta p}{v_{B} \rho} + 1 \right)^{-1}, \tag{9a}
\]
\[
\Delta p = p_{s} - p_{l}, \tag{9b}
\]

where \( c_{p} \) is the specific heat capacity, \( \rho, c_{p}, v_{B}, S_{h}, \) and \( p_{l} \) are derived directly from the convection and petrological model and are always positive; they are therefore functions of position, in particular of depth, which corresponds to the climbing shock model of \cite{Watters2009}. We note in passing that the form of eq. 9 entails some limitations
on $p_s$ and $p_l$ in order for $\Delta T$ to be finite, positive, and real. Inspection of the expressions indicates that this requires $\Phi > 0$ and that this condition is always met for $p_s > p_l$. For $p_s = p_l$, $\Phi = 1$ by de l'Hôpital’s rule, and hence $\Delta T = 0$. For $p_s < p_l$, there are still real solutions as long as $p_l < p_s + \nu_\omega(4S_h)$, but $\Delta T < 0$, and so we set the thermal perturbation to zero as $p_s$ drops below the lithospheric pressure.

In principle, melt production itself could be estimated from scaling laws (e.g., [Bjorkman and Holsapple 1987]; Pierazzo et al. 1997; Abramov et al. 2012), but as this approach gives only bulk melt volumes but no information about the spatial distribution of the melt, it is not usually the approach taken in convection models. We consider directly where eq. [9] leads to supersolidus temperatures and then calculate the amount of melt basically in the same manner as in regular melting processes. However, as the temperatures and melting degrees may be much higher in a giant impact, there will usually exist a region which is much more pervasively molten than normal melt-producing mantle, possibly beyond the rheological solid–liquid transition or even the liquidus. As convection in such a melt pond cannot be handled within the framework of a mantle convection code, we assume that the volume of the model shock-heated above the solidus melts according to the melting parameterization up to an imposed limit of 60% at which we assume that only a residue without any other fusible components would remain; the most extreme conditions are only reached at most in a small part of the shocked volume. The melt in excess of the extraction threshold $\varphi_s$ is added to the top in the usual way as material that has cooled to surface temperature upon eruption. The source region is finally set to the temperature corresponding to $\varphi_s$, i.e., just above the solidus, for the aforementioned technical reasons; a similar numerical crutch for handling the high post-impact temperatures has also been used by previous workers (e.g., Watters et al. 2009; Roberts and Arkani-Hamed 2012). Physically, this procedure can be justified by a combination of several processes: much of the waste heat corresponding to eq. [9] is consumed by melting and vaporization; the post-impact melt pond is expected to cool much more rapidly thanks to its vigorous convection than the duration of a single timestep of the mantle convection algorithm, which usually corresponds to at least several thousand to a few tens of thousands of years (e.g., Reese and Solomatov 2006); and a considerable fraction of the heat is extracted from the mantle along with the erupted melt and implicitly assumed to be lost to space as a consequence of our constant-$T$ boundary condition.

### 2.3 Impact crater and ejecta layer formation

Prior to or simultaneously with causing these thermal effects at the impact site, the impact also forms a crater, a process that includes a modification of the local topography and the permanent redistribution of some of the material at the impact site. The transient crater created in the course of the impact is rapidly modified through slumping of the walls and similar processes in the case of larger events that form complex craters. Although the fixed geometry of the numerical grid, the relatively coarse resolution of global models and the complexity of ejecta ballistics prevent us from modeling these processes in any detail, we attempt to obtain an order-of-magnitude estimate of their basic features, because they potentially affect locally the thickness and composition of the crust, which in turn influences certain observables such as heat flow or gravity. We start by determining from which part of the transient crater material is ejected by calculating a number of mass parcel trajectories within the transient crater using the semi-analytical Z model with a buried source (Maxwell 1977; Craft 1980; Richardson et al. 2007); material expelled by jetting during the first contact between impactor and target is not considered, as these jets seem to account for only a few percent of the total emitted material (Johnson et al. 2014). The transport of the ejecta outside the crater is approximated very simplistically in terms of ballistics of independent particles as suggested by Tauber et al. (1978), whereby the contributions of different particle sizes to the accumulated ejecta is given by empirical size–frequency distributions (O’Keefe and Ahrens 1985; 1987). The result of these calculations are estimates of the ejecta thickness as a function of distance from the impact site and of the ejected and displaced volumes; specifically, we can assess which part of the ejecta ends up within the boundaries of the future final crater. The collapse of the transient crater into the final crater cannot be modeled in any detail in our framework, and so we simply ignore the details of the final topography and impose that in the end the transient crater has transformed into a final crater with a level surface, in which slumped material from the transient crater walls and ejecta are all mixed into a uniform body, with the tracer particles being redistributed according to the final distribution of that material; the deposition of ejecta is treated as an influx of material with surface temperature. Each of these simple models is cheap enough in terms of computational effort to be included into the convection model as a subroutine. We note in passing that Rolf et al. (2017), who also gave consideration to the role of ejecta, describe ejecta blanket formation more straightforwardly in terms of empirical scaling laws but chose to assume zero thickness for ejecta inside the basin.

### 3 Results

#### 3.1 Model set-up

In recent years, models of the martian interior based on the analysis of geodetical and gravity data combined with petrological models have helped to tighten the constraints on the size of the martian core. Khan and Connolly (2008) compiled some interior models, some of which featured a core size small enough to allow for a basal layer of perovskite (pv) and ferropericlase (fp) in the mantle. However, their own preferred model does not include such a layer, and the
Table 1. Model parameters. Default values of variable parameters are printed in italics.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planetary radius, $R_P$</td>
<td>3389.5 km</td>
</tr>
<tr>
<td>Total planetary mass, $M$</td>
<td>$6.4185 \cdot 10^{23}$ kg</td>
</tr>
<tr>
<td>Surface temperature</td>
<td>218 K</td>
</tr>
<tr>
<td>Mantle</td>
<td></td>
</tr>
<tr>
<td>Mantle thickness, $z_m$</td>
<td>1659.5 km</td>
</tr>
<tr>
<td>Initial potential temperature, $T_{pot}$</td>
<td>1700 K</td>
</tr>
<tr>
<td>Initial core superheating</td>
<td>150 K</td>
</tr>
<tr>
<td>Surface porosity, $\varphi_{surf}$</td>
<td>0.2</td>
</tr>
<tr>
<td>Melt extraction threshold, $\varphi_{r}$</td>
<td>0.007</td>
</tr>
<tr>
<td>Bulk silicate Mars Mg#</td>
<td>0.75</td>
</tr>
<tr>
<td>Present-day K content</td>
<td>305 ppm</td>
</tr>
<tr>
<td>Present-day Th content</td>
<td>56 ppb</td>
</tr>
<tr>
<td>Present-day U content</td>
<td>16 ppb</td>
</tr>
<tr>
<td>Initial bulk water content</td>
<td>36, 72, 144 ppm</td>
</tr>
<tr>
<td>Core (average material properties)</td>
<td></td>
</tr>
<tr>
<td>Core radius</td>
<td>1730 km</td>
</tr>
<tr>
<td>Sulfur content</td>
<td>16 wt.%</td>
</tr>
<tr>
<td>Thermal expansivity $^{1}$, $\alpha_c$</td>
<td>$3.5-4.3 \cdot 10^{-5}$ 1/K</td>
</tr>
<tr>
<td>Isobaric specific heat, $c_{pc}$</td>
<td>750 J/(kg K)</td>
</tr>
<tr>
<td>Thermal conductivity $^{2}$, $k_c$</td>
<td>$23.5-25.1$ W/(m K)</td>
</tr>
</tbody>
</table>

$^1$Varies with time.

more recent investigations by [Konopliv et al. 2011] and by [Rivoldini et al. 2011] also point to a rather large core that would preclude the existence of a (pv+fp) layer; the latter study finds that such a layer would be marginally possible only in a hot mantle. Moreover, the impacts under consideration are not large enough to penetrate to depths that allow them to affect the core–mantle boundary region directly, so that the immediate dynamical effects would unfold in regions quite remote from a (pv+fp) layer. For these reasons, we decided to focus on models with a large core and no basal layer.

The chemical model assumed for the bulk mantle and crust is the one by [Wänke and Dreibus 1994], although we do allow for some variability in the water content, having in mind that the water content of the martian mantle is an unsettled issue and may be on the order of several hundred parts per million (e.g., [McCubbin et al. 2010, 2012]), i.e., much higher than the 36 ppm suggested by [Wänke and Dreibus 1994]. We assume this latter value in the major set of models, i.e., a fairly dry initial bulk silicate water content. In addition to these models, whose mantle quickly becomes almost dry, we also consider two further model series with initial bulk water contents of 72 and 144 ppm, i.e., two and four times the default value, respectively. This is enough to leave water contents at the tens-of-ppm level in the mantle, even though it lies lower than the values proposed for instance by [McCubbin et al. 2012]; it may therefore correspond more closely to the scenarios suggested by [Wänke and Dreibus 1994] or more recently by [McCubbin et al. 2016] for the shergottite sources. At these concentrations, the effect of water on thermoelastic properties is not considered to be significant, and even the effect on the position of the solidus will be minor (although we do include it), but it is expected to lower the viscosity of the material noticeably, which is our main motivation for giving it consideration. For the core, we assume a sulfur content of 16 wt.%, more similar to those favored by [Rivoldini et al. 2011], i.e., slightly higher than [Wänke and Dreibus 1994].

We assume an initial state prevailing at 4.4 Ga that is defined by mantle that is compositionally homogeneous and well-mixed on large scales but is also already depleted by the production of the initial primordial crust; this state is implemented as a random pattern of low-degree depletion (∼0.01 on average) throughout the mantle that signifies that the entire mantle has experienced melting or has been mixed with residue from melting thoroughly in the transition from the magma ocean. The thickness of the resulting initial crust is not well constrained by observations but it is mostly dependent on the initial temperature.

As the objective of this study is the comparison of the contributions of thermally and compositionally induced density variations to the buoyancy, the models come in pairs that include one model each in which the compositional contribution to the buoyancy is suppressed by assigning the relevant physical properties of the unmolten material to the depleted material. In other words, in the “thermal-only” (T) model of each model pair, it is assumed that melting does not change the mineralogy and the iron content and hence also leaves density, thermal expansivity, etc. of the residue unchanged, whereas in the other model of the pair (TC), all these properties are functions of the melting degree $f$. However, we do treat the melting material in the T models as depleted residue in the usual sense as far as trace component partitioning or rheology are concerned, i.e., radionuclides and water are removed from the residue in all models.

The reference setup is a Mars-like planet with an initial potential mantle temperature $T_{pot}$ of 1700 K, which is in the range deemed reasonable, e.g., with regard to the resulting crustal thickness, by previous studies (e.g., [Hauck and...
we also give the radius according to Pierazzo et al. (1997) (eq. 5) for reference.

In most of the models with an impact, the reference model is struck by an S-type asteroid with a velocity of 9.6 km/s at an angle of 45° at 4 Ga, which corresponds to 400 My model evolution. The age of the impact is reasonably close (considering uncertainty) to the ages of three martian craters of different size whose parameters we used to define the model impacts, namely, Utopia, Isidis, and Huygens (see Table 2); these three impacts are not intended to be faithful representations of the eponymous craters but serve as representatives covering a range of magnitudes that reaches from having a marginal effect on the mantle to the largest observed basin-forming event. Moreover, the planet has already evolved to some extent at this time and has formed a thermal lithosphere of a certain thickness; the most vigorous stage of regular melt production has already come to an end, so that the impact effects will stand out more clearly. Given crustal thicknesses around 79 km at the time of impact and a melt production zone for regular melting reaching down to \( \gtrsim 330 \) km but reaching the strongest depletion at depths \( \lesssim 200 \) km, these three impacts have the centers of their isobaric core toward the bottom of the main melting zone (Utopia), near its top (Isidis), and within the crust (Huygens). The isobaric cores correspondingly extend into the mantle below the melting zone in the case of Utopia, cover the crust and melting zone in the case of Isidis, and are limited to the crust in the case of Huygens, which should permit to reveal a possible effect of the relative positions of the pre-existing regular global melting zone and the impact-generated anomaly. An effort was made to place the site of every impact so that it is located neither directly above a hot upwelling nor above major cold downwellings; as of the pre-existing regular global melting zone and the impact-generated anomaly. An effort was made to place the site of every impact so that it is located neither directly above a hot upwelling nor above major cold downwellings; as demonstrated by Roberts and Arkani-Hamed (2012), the pre-existing convective features at the impact site can modify the effects of the impact, although its influence for instance on the impact-generated anomaly in the global heat flow does not exceed 10%. For this reason, the impacts appear at different positions in different models, as the different buoyancy mechanisms and water contents result in different convection patterns. An overview of all models is given in Table 3.

The timing of an impact is potentially important, especially in very young planets in which the lithospheric cooling rate is still high. In order to probe this dimension of parameter space, we ran two additional model series in which the impact occurred at 50 or 150 My model time; the uncertainties in crater-statistics dating certainly warrant assuming different times. Because of the small effect of Huygens-sized events, we considered only Isidis- and Utopia-sized impacts in those cases. Apart from the timing, the model parameters are identical to the 400 My cases. Physically, the difference in the “earlier” cases is that the impactor meets a target that is already quite similar to that in the normal Isidis case as far as the extent and depletion of the regular melting zone and the thickness of the crust are concerned, but the thickness of thermal lithosphere is substantially smaller, such that the target volume is hotter on average.

### 3.2 Dynamical evolution

The reference models start with a regular pattern of upwellings and downwellings that breaks down after a few tens of millions of years and transitions into a rather irregular flow; the impacts of the “early” model pairs occur at a time when this transition begins. The Ref-36 models (see Table 3) slowly evolve from the chaotic stage toward the final convection pattern with five to seven stable plumes beginning after \( \sim 500 \) My, i.e., shortly after the time when the regular impacts would strike, whereas this final pattern begins to emerge several hundreds of millions of years later with increasing water content of the mantle.

In terms of the (effective) Rayleigh number \( Ra_{eff} = Ra_0/\eta_{mean} \), the evolution of the reference models can be divided into two major phases (Figure 2): an early period of vigorous but declining convection that lasts until \( \sim 1.5 \) Gy and a later period of nearly constant and calm flow that encompasses the remaining almost 3 Gy. Irrespective of the water content, the T model variant generally convects a bit more energetically during the early stage, whereby the contrast with the TC variant is more pronounced in the model with low water content (36 ppm). The Ref-36 models convect at an approximately constant vigor during the first 1 Gy, whereas the Ref-72 and Ref-144 models show a steep two-stage decline in their Rayleigh numbers, the first of which coincides essentially with the end of their principal period of crust formation. The mean mantle temperatures differ by less than \( \lesssim 10 \) K in most cases and decrease almost linearly throughout the evolution. Only the T variant of Ref-144 follows a somewhat different path, because the combination of the generally lower viscosity and the relatively high density of the cooling uppermost mantle let it become gravitationally unstable and develop numerous downwellings, whereas its TC counterpart is stabilized by the additional compositional buoyancy in the melting zone.

An impact introduces an instantaneous local thermal disturbance that results in enhanced local heating and melting. This disturbance is visible as a spike signal in several parameters that characterize the thermal evolution, e.g., in the
Table 3. Model pairs

<table>
<thead>
<tr>
<th>Model pair name</th>
<th>Impact size</th>
<th>$C_{H_2O}$ (ppm)</th>
<th>time of impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref-36</td>
<td>–</td>
<td>36</td>
<td>–</td>
</tr>
<tr>
<td>H-36</td>
<td>Huygens</td>
<td>36</td>
<td>400 My</td>
</tr>
<tr>
<td>I-36</td>
<td>Isidis</td>
<td>36</td>
<td>400 My</td>
</tr>
<tr>
<td>I-36m</td>
<td>Isidis</td>
<td>36</td>
<td>150 My</td>
</tr>
<tr>
<td>I-36e</td>
<td>Isidis</td>
<td>36</td>
<td>50 My</td>
</tr>
<tr>
<td>U-36</td>
<td>Utopia</td>
<td>36</td>
<td>400 My</td>
</tr>
<tr>
<td>U-36m</td>
<td>Utopia</td>
<td>36</td>
<td>150 My</td>
</tr>
<tr>
<td>U-36e</td>
<td>Utopia</td>
<td>36</td>
<td>50 My</td>
</tr>
<tr>
<td>Ref-72</td>
<td>–</td>
<td>72</td>
<td>–</td>
</tr>
<tr>
<td>I-72</td>
<td>Isidis</td>
<td>72</td>
<td>400 My</td>
</tr>
<tr>
<td>I-72m</td>
<td>Isidis</td>
<td>72</td>
<td>150 My</td>
</tr>
<tr>
<td>I-72e</td>
<td>Isidis</td>
<td>72</td>
<td>50 My</td>
</tr>
<tr>
<td>U-72</td>
<td>Utopia</td>
<td>72</td>
<td>400 My</td>
</tr>
<tr>
<td>U-72m</td>
<td>Utopia</td>
<td>72</td>
<td>150 My</td>
</tr>
<tr>
<td>U-72e</td>
<td>Utopia</td>
<td>72</td>
<td>50 My</td>
</tr>
<tr>
<td>Ref-144</td>
<td>–</td>
<td>144</td>
<td>–</td>
</tr>
<tr>
<td>I-144</td>
<td>Isidis</td>
<td>144</td>
<td>400 My</td>
</tr>
<tr>
<td>I-144e</td>
<td>Isidis</td>
<td>144</td>
<td>50 My</td>
</tr>
<tr>
<td>U-144</td>
<td>Utopia</td>
<td>144</td>
<td>400 My</td>
</tr>
<tr>
<td>U-144m</td>
<td>Utopia</td>
<td>144</td>
<td>150 My</td>
</tr>
<tr>
<td>U-144e</td>
<td>Utopia</td>
<td>144</td>
<td>50 My</td>
</tr>
</tbody>
</table>

Figure 2. 10-My averages of the effective Rayleigh number as a function of time for the reference models and the models with a Utopia-like impact at 400 My. Models with thermal and compositional buoyancy (TC) are plotted with solid lines and models with purely thermal buoyancy (T) with dashed lines.

curves of $Ra_{eff}$ in Figure 2, whereby the magnitude of the signal grows with increasing water content/decreasing viscosity, because a low-viscosity system reacts more strongly to any perturbation. In the models with an earlier impact, the evolution of $Ra_{eff}$ is essentially the same, although in the 50 My series the spike of the impact is absorbed into the general initial steep increase of this parameter. As a consequence of the impact, a compositional anomaly characterized by enhanced depletion develops, as can be seen in the composition ($f$) images in Figure 3. The sequences in that figure and the corresponding animations in the Supporting Information show how the compositionally buoyant impact anomaly essentially stays in place for billions of years in the TC models, whereas it is dispersed within a few hundreds of millions of years in the T models. The density decrease due to heating and, in the TC models, due to the loss of higher-density constituents triggers a localized upwelling whose influence can reach far beyond the volume immediately affected by the impact itself. In the larger impacts, a hot pulse rapidly spreads out beneath the lithosphere and reinforces melting and crust formation if the center of the isobaric core lies deep enough. In a very large impact, the upwelling modifies the flow field on a regional scale such that pre-existing plumes may be drawn toward the impact site and merge. In most models with larger impacts, this effect persists for several hundreds of
milliseconds of years, i.e., much longer than the actual lifetime of the thermal anomaly of a few tens of millions of years, suggesting that a very large (Utopia-sized) impact can enforce a fairly stable regional flow field that is self-sustaining to some extent under favorable conditions; the TC model from U-36 establishes such a long-lived plume for billions of years, as can be seen in Figure 3 (top row). In a more vigorous convective regime as in its wetter counterparts from U-72 and U-144, the enhanced stirring disturbs the stability of the structure and shortens its lifetime the more the higher the water content is, although the direct hot-material surge from the impact spreads further due to the lower viscosity. In the model series with earlier impacts, there is a tendency of globally reinforced crust production and delamination, but the effect is insignificant except for the 144 ppm water model of the Utopia-sized event, in which a substantial amount of crustal material is reintroduced into the mantle. In the corresponding T models, the local effect decays much faster, on timescales of several tens rather than hundreds of millions of years because of the absence of the self-stabilizing compositional anomaly, although the general reinvigoration of convection tends to be a bit more persistent than in the TC models, judging from Figure 2 Figure 4 (bottom left) shows that there is no plume near the former impact site 2 Gy after the impact, although the convection pattern is nonetheless not as regular as that of the reference model shown in Figure 4 (bottom right). This underscores that the compositional buoyancy of the impact-induced anomaly must contribute significantly to the establishment of the stable flow field. An effect is also well visible in the TC models of Isidis-size impacts, but it is less long-lived and extensive, whereas it is insignificant in the T versions. In the Huygens-size events (H-36), which are essentially confined to the thermal lithosphere, there is no effect beyond a small, short-lived local disturbance.

A large part of the crust – approximately 59 km – is produced in the initialization step as “primordial” crust, and crustal growth is strongest at the beginning and declines toward zero afterwards. In the reference models the crust usually reaches its final thickness after 1–1.5 Gy, reaching values higher by a few kilometers in models with higher mantle water contents. The amount of crust added after initialization lies between approximately 20 and 25 km in most cases; only the T variants of the models with 144 ppm water experience more voluminous and protracted crust formation thanks to the reinforcement of convection by the cold downwellings. In some regions that have been fed by plumes over extended periods, the crust eventually grows thick enough to delaminate; nonetheless, delamination is generally not a quantitatively important process, and the small amount of delaminated material is mixed into the deeper mantle within a few tens of millions of years. In Ref-144/T and its impact-affected derivatives delamination is more extensive, because the downwellings drag not only lithospheric mantle but also a part of the thick crust with them, but it does nonetheless not reach a global scale that affects the entire crust.

The enhanced melting and crust formation at an impact site may also result in a thicker crust, but on the other hand, the removal of crustal material by cratem formation may compensate to some extent the growth of the crust. In all models, we find that the T variant develops a thicker post-impact crust at the impact site than its TC counterpart, because the buoyant residue from the impact was less stable and hence less efficient at preventing the ascent of fertile material into the melting zone. The post-impact thickening is largely due to the immediate melt production from the impact in the TC models and therefore finishes within a few tens of millions of years, whereas in the T models, the enhanced crustal formation persists for hundreds of millions of years. The extent to which the crust grows depends on the water content of the mantle and the time of the impact: early impacts and/or higher water contents generally entail a thicker crust, because the lower viscosity allows the mantle to convect more vigorously and produce more melt. In the two larger sorts of impact, the post-impact crust is always thicker than the mean of the reference model due to the additional melt production, whereas the Huygens-sized events do not produce enough melt to compensate the removal of crust by the impact itself. Crustal thickness variations exist directly after the impact, but they tend to flatten out soon by spreading laterally as a consequence of the buoyancy forces of the crust and the underlying mantle anomaly, as well as the action of shear forces from divergent mantle flow at its base. Post-impact melt production and crust formation are facilitated by thermal “rejuvenation” of the lithosphere caused by the impact, which produces a local pit-like low-viscosity zone with a lifetime of several tens of millions of years in the boundary region between crust and mantle. An example for this phenomenon is visible at the impact sites of the I-36 models in Figure 3. This process seems to be more efficient in the T models, where the root is less protected by a stable density anomaly in the underlying mantle. Figure 5 summarizes the principal stages of the post-impact evolution after a large basin-forming impact.

### 3.3 Observables

A variety of geophysical and geochemical observables can be derived from the models. Although it is not our primary goal to reproduce in detail observations from Mars, we will draw comparisons with results from missions to that planet in the following and concentrate on some observables for which data are relatively abundant or which are expected to become available in the near future as results from the InSight mission.

#### 3.3.1 Surface heat flow, $q_s$, and elastic thickness of the lithosphere, $z_{el}$

The heat flux estimates from spacecraft missions have so far usually been derived from flexure or admittance models and are mean heat flows through the elastic lithosphere; comparable fluxes from numerical models could be derived using the temperature difference between the top and bottom of the elastic lithosphere, its thickness, and its average
Figure 3. Evolution of temperature and composition in the model pair I-36. Left: model with thermal and compositional buoyancy; right: model with purely thermal buoyancy. In the images of composition, a positive melting degree $f$ indicates depleted material; the crust is displayed as strongly “enriched” (dark blue). The colorbars are clipped for clarity. Animations of the evolution of both models are available online as Supporting Information.

thermal conductivity, as done by [Ruedas et al. 2013a,b]. The HP$^3$ experiment on the InSight mission, by contrast, is a real heat flux probe designed to determine the actual surface heat flux at the landing site; this is not identical to the mean lithospheric heat flow, but it can be approximated in a numerical model by a finite-difference formulation using the outermost grid cell layer. In this study, we consider the surface heat flow and the elastic thickness of the
lithosphere, which is determined as an intermediate step in the calculation of the mean lithospheric heat flow and should be more readily comparable to independent data as it requires fewer assumptions.

The determination of the elastic thickness of the lithosphere from flexure models usually involves the use of non-Newtonian rheology, which is not part of our numerical model. A direct comparison is therefore not possible, but one can still attempt an estimate by assuming an average lithospheric strain rate $\dot{\varepsilon}$ and a yield stress $\sigma_y$ typical for lithospheric materials. We follow Grott and Breuer (2008), who assumed a strain rate of $10^{-17} \text{s}^{-1}$ and yield stress of 15 MPa and then solve the Arrhenius-type flow law for non-Newtonian creep for temperature:

$$T_{bd} = \frac{Q}{R_{\text{gas}}} \left[ \ln \left( \frac{B\sigma_y^{\nu}}{\dot{\varepsilon}} \right) \right]^{-1},$$

(10)

Figure 4. Temperature field of the TC model from U-36 at the time of the impact, and temperature fields of models U-36 (TC), U-36 (T), and Ref-36 (TC) about 2 Gy later. The white arrows mark the impact sites. The colorbar is clipped for clarity.
Figure 5. Post-impact evolution of a thermochemical anomaly generated by a basin-forming impact. a) Shortly after the impact, the crust around the impact site is removed or shattered, and a large region around the center of the isobaric core is shock-heated, partly to temperatures above the solidus. b) The melt extracted from the supersolidus region leaves behind a depleted residue in the mantle and has formed a new crust together with pre-impact crustal material from the final crater. In the left part of this and the following stages, the new crust is thicker than average, in the right part it is thinner; which variant develops in any given crater depends on the magnitude of the impact and the amount of melt produced by it. c) The thermochemical anomaly rises due to its buoyancy and may trigger a plume from the CMB or attract one from nearby. With time the thermal component of the anomaly fades, whereas the compositional is not much diminished. The rising plume may produce additional crust. d) Final stage: the thermal anomaly and the plume have disappeared, whereas the compositional anomaly has spread out at the base of the lithosphere/crust and remains in place. Variations in the thickness of the crust may be levelled to some extent by creep. The figures are not to scale.
where the factor $B$, the activation enthalpy $Q$, and the exponent $\nu$ are $3.1 \times 10^{-20} \text{Pa}^{-3.05} \text{s}^{-1}$, $276 \text{kJ/mol}$, and $3.05$ for wet crustal diabase and $2.4 \times 10^{-16} \text{Pa}^{-3.5} \text{s}^{-1}$, $540 \text{kJ/mol}$, and $3.5$ for dry mantle olivine, respectively; in this context, “wet” means that the material is saturated with water, which is already the case for the volumetrically (and thus rheologically) important minerals at very low concentrations under the low pressure conditions of the crust. We allow for the possibility of an incompetent layer separating the crust and mantle segments of the elastic lithosphere but did not actually observe such a configuration in our models. Therefore, the elastic thickness of the lithosphere is tied to a single temperature $T_{bd}$ as defined in Eq. (10) associated with the brittle–ductile transition of the mantle ($\sim 1065 \text{K}$ in our case), similar to Ruedas et al. (2013a, b), whose transition temperature was a few dozen kelvins lower; the depth at which this temperature is reached is taken to represent $z_1$. In some instances, structures like cold downwellings make it difficult to determine reasonably the local elastic thickness, and it seems safer to discard such regions when computing the global mean; this problem arose in the models with low water content especially in the time interval from $\sim 250 \text{My}$ to $\sim 1.5 \text{ Gy}$, and therefore the corresponding curves in Figure 6 should be regarded as approximate only in that time interval.

The surface heat flow of the reference models (Figure 6) shows a small increase at the earliest stage of evolution before declining smoothly toward present-day values close to $26.5 \text{ mW/m}^2$; the curves are almost the same in all reference models except for that initial phase (Figure 3). In general, the heat flux does not show large variation in the Reference models: deviations from the global mean by more than $1–2 \text{ mW/m}^2$ are rare. The elastic thicknesses grow almost linearly for most of the time toward a modern value of a bit more than $200 \text{ km}$, except for some minor variability during the first few hundreds of millions of years; the elastic thicknesses in the Ref-144 pair are some $20 \text{ km}$ larger than those of the Ref-36 pair, but the difference decreases as the planet evolves. All models are broadly consistent with most literature data, especially for the first $1.5 \text{ Gy}$ (Figure 6b); in such comparisons it must be kept in mind that several of the heat flow estimates were made for specific regional structures such as volcanic edifices or grabens and that their age is very uncertain.

Impact events produce a sudden change in the conditions near the surface in the surroundings of the impact site: they shock-heat the interior beneath the point of impact and trigger melt production, which leads to magmatic activity. In our treatment, we assume that 1.) the erupted material loses heat quickly and can therefore be set to surface temperature; and that 2.) a large volume of ejecta is produced, much of which is deposited as a layer of several kilometers thickness and with surface temperature within the crater. The surface porosity within the crater is assumed to be zero due to melt filling the voids. The heat flux in the affected area, i.e., within the final crater is strongly reduced but returns to higher levels with time, although the levels of undisturbed areas are not reached again; only in the case of the Huygens-sized events is the drop preceded by a spike-like maximum, which in that case is due to the direct heating effect of the impact itself. The temporal heat flux evolutions for any given impact size are quite similar for all three impact times in terms of the amplitude of the disturbance and the duration of its decay and result in almost identical regional present-day heat flows. Moreover, the regional heat flux evolutions of the Isidis-sized and the Utopia-size events at $400 \text{ My}$ are very similar at any given water content, such that the curves for the Utopia-sized models in Figure 6 are fairly representative for both. The reason for the reduction of the heat flow, which is contrary to intuition and to the findings of other authors (e.g., Roberts and Arkani-Hamed 2012), is that these large events produce a substantial amount of ejecta and erupted material that is assumed to have surface temperature when being deposited as a layer, whose ejecta component alone can be more than $10 \text{ km}$ thick. This material is represented as a cold influx that shifts isotherms downward and thus causes a very shallow temperature gradient that reduces the
heat flow, especially inside the final crater. This effect of ejecta and other shattered material was not taken into consideration in most previous studies. Rolf et al. (2017) do not observe the cooling effect in their models of the Moon, because they treat ejecta deposition in a different way and assume that ejecta are absent within the basin; they find a reduced heat flux just beyond the final crater rim, which in their case is due to the reduced thermal conductivity of the relatively thick ejecta deposits there. Moreover, the crust formed in the crater in our models is produced from high-degree melting of a source that was already partly depleted, so that its content in heat-producing elements is several times lower than the average, resulting in a reduced crustal contribution to the surface heat flux. The hot pulse from the mantle diffuses only slowly upward and is then consumed in heating the cool ejecta-thickened crust. The present-day elastic thickness of the lithosphere beneath the crater in models with impacts is usually not significantly different from the global mean of impact-free models, reaching values between 200 and ~230 km, irrespective of the time of impact; the only stronger deviation from this pattern is the U-144 model pair for an impact at 50 My model time, in which z_{el} reaches values around 250 km.

On a global scale, even large impacts produce only a small instantaneous negative excursion from the values of the Reference models, followed by an exponential-like return toward the heat flux of the corresponding reference model within a few tens of millions of years. The thermal disturbance from a single impact early in the planet’s history decays quickly on the timescale of planetary evolution and leaves no distinguishable trace today. In this context it is worth pointing out that a calculation of global mean values by computing the surface heat flow for every single grid column, summing, and dividing by the number of columns as done for the Reference models would be misleading in two-dimensional models with impacts and is therefore omitted here. The reason is that a single local anomaly like an impact site occupies a certain fraction of the domain that is larger than the true fraction of the entire planet that is affected by an impact in three dimensions, because the two-dimensional geometry implicitly extends that anomaly along half a meridian in the third dimension, making its spatial extent potentially much larger than it would be in three dimensions. As a consequence, it contributes more to the mean than it would in three dimensions, thus resulting in an overestimate. Although this applies, in principle, to all major intrinsically three-dimensional structures in two-dimensional models, the effects of structures that are ubiquitous in the model and can make positive or negative contributions to the mean may cancel out reasonably well for the purpose of calculations such as mean global heat flow estimates; this is unfortunately not the case with a single local structure such as an impact. Appendix B gives a simple estimate of this geometrical error and indicates that the true amplitudes of the impact signal would be one to two orders of magnitude smaller even for large impacts.

### 3.3.2 Density and seismic velocities

Our thermoelastic properties model directly yields the density ϱ of the crust and mantle at different depletions. Subtraction of the laterally averaged density–depth profiles of the T model from the TC model of a model pair indicates that the compositional density difference due to melting is up to 30 kg/m^3; locally, as in impact-generated anomalies with very high depletion, it may exceed 50 kg/m^3. This additional contribution to buoyancy has certain dynamical implications that become especially important on a local scale in the models with impacts, as can be seen by comparing the evolutions of the two models of a given pair. In impacts that are large enough to reach into the mantle, the anomaly created by the strong heating and depletion has a large buoyancy and rises toward the base of the lithosphere, where it tends to spread, as mentioned above (cf. Figure 3). Contrary to the thermal anomaly, which vanishes as the excess heat diffuses away, the depletion anomaly could only be destroyed by remixing. This does indeed happen in the T models, but the compositional contribution to the buoyancy active in the TC models helps to stabilize the anomaly in those cases. The signature of the impact is clearly visible in present-day depth profiles of depletion of the TC models at the site, where the preserved depletion from the event is still nearly up to three times as high as the average maximum of impact-free models. It is correlated with density anomalies (Figure 7b), i.e., it has survived 4 Gy of planetary evolution, whereas it has largely vanished in the T models. The inset figure, which covers the uppermost 400 km/5 GPa of the mantle, shows clearly the density deficit of the compositional anomalies of the two larger impact categories relative to the Reference model, whereas the uppermost slice of mantle below the Huygens site happens to be less depleted even than the Reference and is therefore a bit denser. By contrast, the purely thermal models all have a higher and almost identical density in this depth range; the two high-density segments in models U-36/T and U-144/T downward from 100 km are due to eclogitic crustal material that is beginning to founder. The profiles for the 72 and 144 ppm water models look very similar and show the same features, except that they seem to reach a bit less deep; the reason is probably that the generally lower viscosity in these models allows density anomalies to rise and spread out laterally more easily. The same pattern is observed in the models with earlier impact times.

As the thermoelastic model is based on an equation of state, it also provides the isothermal bulk modulus K_T directly, although it does not yet include the shear modulus and its derivatives. Nonetheless, as a first step toward application to seismic data, we calculate the bulk sound speed in the infinite-frequency approximation,

\[ v_B = \sqrt{\frac{K_S}{\rho}} = \sqrt{v_T^2 - \frac{4}{3} v_s^2}; \]  

the adiabatic bulk modulus K_S can be calculated directly from the available thermoelastic properties. This relation permits the direct comparison with data from a seismometer on Mars.
own strong buoyancy. The long-lived compositional component of that buoyancy, which is not destroyed by diffusion, basin-forming impacts, rise toward the base of the lithosphere, where they spread out under the influence of their melt production. Moreover, local features with extreme density contrasts, exemplified here by anomalies created by that in Mars density contrasts between fertile and depleted mantle rock may indeed reach a magnitude that throttles compositional changes the mantle material undergoes upon melting, confirm the finding of those authors and suggest fertile rock exceeds a certain value. Our models, in which the density contrast is not imposed but calculated from the extent and thus result in a thinner crust if the density difference between depleted material in the melting zone and the depleted low-density layer is dynamically stable enough to suppress melting and crust formation to a significant density contrasts on convection in a more idealized Mars-like planet with imposed density contrasts and found that asthenosphere, and local structures therein.

Plesa and Breuer (2014) have studied the effect of different compositional depletion (a) and density (b) at 4.4 Gy for selected models. The profiles of the reference models are laterally averaged profiles, the profiles of the models with impacts were taken at the impact site; the depletion profiles of the latter set were smoothed for clarity. Also shown are experimentally determined densities from Bertka and Fei (1998) for a composition modeled after Wanke and Dreibus (1994).

Figure 7 shows representative depth profiles of TC models from the 36 ppm water series. The profiles for models with impacts correspond to the impact site and show the signature of the remnants of the impact in the topmost mantle. The magnitude of the seismic anomaly can be assessed by comparison with the laterally averaged profile for model Ref-36, which is virtually identical with the laterally averaged profiles of the impact-affected models; the seismic $v_B$ anomaly is on the order of a few tens of meters per second. The anomalies in the other model series are of a similar magnitude; as we did not consider possible effects of water due to limited data and the generally low water concentrations in these models, the water-related effects on seismic velocity are not included here, but would likely be negligible in our models. The seismic profile shows a slightly thickened crust at the impact site in models I-36 and U-36, but given the limitations on spatial resolution, the thickening should not be interpreted quantitatively.

4 Discussion

The most interesting dynamical feature of the models of this study are certain effects of the compositional contribution to the buoyancy of mantle material on the character of convection and the stability of the lithosphere, the asthenosphere, and local structures therein. Plesa and Breuer (2014) have studied the effect of different compositional density contrasts on convection in a more idealized Mars-like planet with imposed density contrasts and found that the depleted low-density layer is dynamically stable enough to suppress melting and crust formation to a significant extent and thus result in a thinner crust if the density difference between depleted material in the melting zone and fertile rock exceeds a certain value. Our models, in which the density contrast is not imposed but calculated from the compositional changes the mantle material undergoes upon melting, confirm the finding of those authors and suggest that in Mars density contrasts between fertile and depleted mantle rock may indeed reach a magnitude that throttles melt production. Moreover, local features with extreme density contrasts, exemplified here by anomalies created by basin-forming impacts, rise toward the base of the lithosphere, where they spread out under the influence of their own strong buoyancy. The long-lived compositional component of that buoyancy, which is not destroyed by diffusion,
makes it possible for them to survive until they are integrated into the lithosphere, facilitating their preservation for the entire remaining history of the planet. Rolf et al. (2017) came to a similar conclusion in their models of the Moon.

As Mars has been hit by many impacts large enough to penetrate into the asthenosphere or even deeper, impacts should be expected to have produced substantial lateral chemical heterogeneity in the uppermost martian mantle, because they would remelt material in the regular melting zone to different degrees according to their sizes and depths of penetration. The largest events would even affect mantle material that lies below the regular melting region and may therefore be more pristine. The resulting ancient melts from the impact would span a wide range of compositions. Moreover, their source regions in the mantle, which are depleted to different degrees, would themselves form a set of heterogeneous source regions for later volcanism if they become reactivated as melt sources, e.g., by plumes. All this heterogeneity may be reflected in the chemical variability observed in the different types of martian meteorites. Although the impacts we have modeled postdate the supposed time around 4.5 Ga (e.g., Borg et al., 1997, 2016) at which the major heterogeneity distinguishing the different classes of martian meteorites was established by several hundred millions of years, the mechanism as such could have worked at those earlier times as well; the model pairs with impacts at earlier times indicate that it is viable even at the more vigorous convection when the lithosphere was thinner and the mantle even hotter, and the higher number of giant impacts would offer more opportunities to form the corresponding strong anomalies. Such a scenario would come close to Option 3 of Borg et al. (2016, Fig. 6), in which a major part of the mantle of the recently accreted and differentiated early Mars is molten during giant impacts and undergoes silicate differentiation, which produces separate shergottite and nakhlite source regions by ~ 4.5 Ga. Apart from these regional scenarios, we note that even regular melting could induce some degree of compositional layering with a density structure that counteracts rehomogenization, especially at higher initial temperatures than assumed here, as shown already before by Plesa and Breuer (2014). That scenario resembles the one proposed by Jones (2003), who has suggested that the mantle sources of nakhlites and the more depleted shergottites lie at different depths and are largely isolated from each other. Finally, if a part of the impactor material merges with the mantle target, this effect may permit it to leave its signature in those preserved anomalies, although we did not include the potentially different composition of the impactor in the model.

Apart from the obvious crater at the surface, how can old impact sites be detected? The most conspicuous feature
of an impact, the crater, is being obliterated with time and may all but disappear. Indeed, even for very large impacts, it is not quite clear how many have occurred in the ancient past of Mars: an analysis by Frey (2008) tallied 20 circular depressions with more than 1000 km diameter that are considered traces of giant impacts, whereas a more recent revision by Frey and Mannoia (2014) came up with 31 candidates. These analyses were based on topography, modeled crustal thicknesses, and surface geology, all of which depend on assumptions or may be affected by later influences such as rebound or the spreading and leveling of crustal thickness undulations. Under such circumstances, additional criteria provided by gravimetry or seismology could, in principle, be helpful.

Figure 7 shows some representative density profiles that quantify the density contrast that characterizes the self-stabilizing impact-created anomalies, which at that stage is entirely compositional. The present-day density deficit associated with very large impacts such as Isidis or even Utopia reaches its maximum of \( \sim 30 \text{ kg/m}^3 \) at a depth of 120 or 220 km, respectively, and may be overlain by anomalous structures such as a thickened crust or a thin slice of less depleted mantle. Neglecting curvature, we make a rough estimate of the expected gravity anomaly by approximating the deficit mass of the mantle anomaly by a flat, thick cylinder of radius \( R_a \), height \( h_a \), and homogeneous density contrast \( \Delta \rho \) extending downward from a depth \( z_a \). The vertical component of the gravity anomaly on its axis relative to the reference model mantle is given by

\[
\Delta g_z = 2 \pi G_0 \Delta \rho \left( h_a + \sqrt{z_a^2 + R_a^2} - \sqrt{(z_a + h_a)^2 + R_a^2} \right)
\]

and can reach values of \(-129\) and \(-194\) mgal for I-36/TC (\( h_a = 135\) km, \( R_a \approx \pi R_P / 6 \)) and U-36/TC (\( h_a = 195\) km, \( R_a \approx \pi R_P / 3 \)), respectively (Fig. 9), assuming \( \Delta \rho = -25\) kg/m\(^3\) and \( z_a = 85\) km in both cases. We find a crust thickened by \( \sim 20\) km in both impact models, but notice that the vertical resolution of our models does not permit a much better determination of the crustal thickness than the grid point spacing (\( \sim 13\) km). To estimate the crustal contribution, we construct a three-cylinder model consisting of a deep cylinder (\( \Delta \rho = -12 \) or \(-20\) kg/m\(^3\) for Isidis and Utopia) representing the excess crust and two stacked cylinders (\( \Delta \rho = 10 \) or \(250\) kg/m\(^3\)) each spanning half of the depth range of normal crust and representing the denser, pore-free crust in the final crater; the radius of the crustal anomaly is taken to be that of the final crater. The resulting positive anomalies will result in fairly similar total gravity signals from these two impacts, and they would at any rate be visible from the ground as well as from an orbit at the elevation of the Mars Reconnaissance Orbiter (MRO) (\( \sim 400\) km) (Figure 9). However, due to the non-uniqueness of potential fields, such signatures cannot be taken as proof of the presence of an impact-related mantle anomaly. To illustrate this point, one can try to mimic the impact-generated mantle anomalies by crust thickened by a certain amount over an area of the dimensions of our assumed mantle anomaly but overlaying an undisturbed mantle. As the gray curves labeled cqI-36 and cqU-36m in Figure 9 show, thickening by 15 or 22 km of a crust with \( \Delta \rho = -210\) kg/m\(^3\) would almost perfectly match the I-36/TC and U-36/TC mantle anomalies, respectively. An important implication of this trade-off is that neglecting negative densities anomalies in the mantle such as those caused by ancient basin-forming impacts may result in an overestimate of the crustal thickness. However, we emphasize that this calculation for a thickened post-impact crust is only an example based on the larger impacts in this study; depending on the parameters of the impact, the post-impact crust could also be thinner. Net crustal thinning is often invoked in the interpretation of gravity signatures of impact basins. Therefore, our approach to extract almost all of the melt generated directly by the impact and assume that it contributes to the formation of the post-impact crust only puts an upper bound on the crustal thickness but may be an overestimate. The complex processes occurring during the rebound and collapse of the transient crater, the crystallization of a post-impact magma pond, and the different modes of melt migration under these conditions cannot be modeled within the framework of our global mantle convection models. Dedicated investigations of the impact process itself will be needed to provide a parameterization that allows to determine the fraction of impact-generated melt that actually forms new post-impact crust in a modeling setup like ours. At any rate, present models of the global crustal thickness variation on Mars (e.g., Zuber et al. 2000; Wieczorek and Zuber 2004) are based on the assumption of a constant mantle density beneath the crust and may therefore produce inaccurate thickness estimates.

A seismic survey targeted at a potential impact site will certainly map the crust–mantle boundary and thus resolve deviations in crustal thickness, but the detection of the bulk velocity anomaly in the mantle, which is not expected to exceed 50 m/s (cf. Figure 5), would require a very good coverage of the shallow mantle with seismic rays and a reliable reference model; whether actual P or S wave anomalies are more readily seen is yet to be investigated. Although such observations are in principle possible, the observational framework is not expected to come into existence in the next few decades.

The global surface heat flow curves are very similar for all models over almost the entire evolution of the planet, as we stuck to one radionuclide composition; this is in agreement with the findings by Plesa et al. (2015), who arrive at similar values for their models with the Wänke and Dreibus (1994) concentrations. The heat flow experiment HP3 on the InSight mission, which is set to arrive at Mars in 2018, will hopefully answer the question whether the radionuclide composition by Wänke and Dreibus (1994) is correct, but it will not be able to discern between other characteristics of interior models such as viscosity or water content due to the insensitivity of \( q_h \). Estimates of the elastic thickness of the lithosphere and the mean lithospheric heat flow, which are already available, are burdened with more assumptions but provide more means to distinguish between different scenarios, especially with regard to the water content and its effect on convective vigor. We did not consider local heat flow variations in more detail, because they are poorly
Figure 9. Estimated on-axis gravity anomaly at the TC models for I-36 and U-36 as a function of height above the martian surface according to eq. (12). Also shown are the signals of positive crustal thickness anomalies above normal mantle (cqI-36 and cqU-36) that would be almost identical to the impact-generated mantle anomalies in I-36 and U-36, respectively, as discussed in the text.

resolved and depend on a variety of assumptions. For instance, local variations arise in our models as a consequence of crustal temperature modifications after impacts, whereas in the lunar models by Rolf et al. (2017), they are caused by variations in the thickness of ejecta layers with a low thermal conductivity. We agree with those authors in the sense that both of us assigned an insulating effect to ejecta or regolith in general, but contrary to Rolf et al. (2017), we assume that the permanent meteorite flux had produced an insulating global regolith layer from the beginning, which is implemented as a pressure-dependent layer of reduced thermal conductivity and density. Similar to them, we reset the porosity of the surface layer to zero within the final crater by invoking magmatic processes that fill the pore space; this results in positive mass anomalies in the crust within the crater. As it is impossible to quantify plausibly the ratio of intrusive to extrusive magmatism in the aftermath of the impact or the detailed temperature distribution of the ejecta and the newly formed crust in the framework of these global models, we set the temperature of the crater infilling to the surface value as the simplest solution but caution that this results in an underestimate of the post-impact crustal temperature; the crater surface heat flows in Fig. [I] should therefore be regarded as lower bounds. At any rate, local variations of conductivity at the surface are not expected to have any substantial effect on the dynamics and processes in the deeper interior, especially in the direct aftermath of the impact, and on the resulting density anomalies.

5 Conclusions

The compositional contribution to buoyancy in convection models with melting can produce a stable, depleted layer beneath the lithosphere that acts to reduce melt production and may help to establish separate chemical reservoirs. The same factor appears to be crucial for anchoring highly depleted compositional anomalies, which may for instance be produced by large impacts, beneath the lithosphere and stabilize them for timespans of hundreds of millions of years, allowing them to be permanently included into the growing lithosphere. Such stable anomalies may appear as different reservoirs in the geochemical record.
In spite of this mechanism of preservation, anomalies such as those created by giant impacts are heavily modified by mantle convection, and their geophysical signature is expected to degrade with time. As the chemical anomaly entails a density anomaly, it may be possible to detect it in gravity data; by contrast, the prospects for detecting it with seismic means seem rather poor, and the heat flux anomaly directly caused by the impact is nil. Thermal signals from single impacts up to the size considered here do not have a long-lasting, discernible effect on the global heat flow; whether that is different for the cumulative effects of several basin-forming impacts will depend on their magnitudes and timing and will be investigated in a future study.

A Martian mantle solidus

We have revised the mantle solidus derived in [Ruedas et al. (2013b)], which was based on data by Bertka and Holloway (1994) and Schmerr et al. (2001) from experiments on a K-free martian model peridotite similar to the compositional model from Wänke and Dreibus (1994); we decided not to use the results from Bory and Draper (2003) and Agee and Draper (2004) although it would be desirable to have more data at higher $p$, because these authors used a material with somewhat different iron and alkali contents. More recently, Matsukage et al. (2013) and Collinet et al. (2015) performed new melting experiments on martian mantle analogue materials after Wänke and Dreibus (1994), the former also on a K-free composition. As [Collinet et al. (2015)] pointed out, the lack of K in the other studies would lead to an increase of the solidus by $\sim 30$ K; this estimate is consistent with the effect of K on the solidus observed in the K-doped peridotite from [Wang and Takahashi 2000], if the difference in K content is accounted for by a linear dependence. Therefore, we reduced the temperature values measured at upper mantle pressures from the other studies by 30 K. As [Collinet et al. 2015] did not provide subsolidus data, we extrapolated the melting degree–temperature relations from their experiments linearly at each pressure and weighted the resulting solidus estimate doubly. In view of the high-$p$ values from [Schmerr et al. (2001)], it seems that the solidus determined by [Collinet et al. (2015)] themselves is unsuited for extrapolation beyond $\sim 3.5$ GPa, and so we decided to fit a selection of all available data for pressures up to the appearance of pv+fp anew, which gave:

$$T_s = \begin{cases} 0.118912p^3 - 6.37695p^2 + 130.33p + 1340.38 & \text{for } p < 23 \text{ GPa} \\ 62.5p + 975 & \text{otherwise}, \end{cases}$$

for $p$ in GPa and $T_s$ in K. The linear fit to the highest-$p$ points from [Schmerr et al. (2001)] for the pv+fp stability field was only slightly shifted to avoid a kink at the transition but was otherwise left unchanged from our previous work, because the depressing effect of K on the solidus was observed to disappear at lower-mantle conditions ([Wang and Takahashi 2000]; in our work, we rarely approach the solidus at that depth anyway, and pv+fp does not appear at all in the models of this study.

Apart from the data for the solidus itself, the work of Matsukage et al. (2013) and Collinet et al. (2015) was also used to readjust estimates for the temperatures at which certain mineral phases are exhausted during melting, especially for the garnet stability field.

Kiefer et al. (2015) discussed the effect of Na, K, and Ca on the martian solidus and proposed shifting the original terrestrial peridotite solidus by Hirschmann (2000) down by 35 K in order to account for that effect. However, Katz et al. (2003, section 4.1) had already pointed out that the original Hirschmann (2000) solidus, which was based on intermediate fertile terrestrial compositions, seems to overestimate the solidus temperature as a consequence of the poor coverage of low-degree melting by experimental data and had suggested lowering it by 35 K, which they found to give a better fit to the experimental data; in other words, their improved fit for terrestrial peridotite is practically identical to the function of Kiefer et al. (2015) suggesting for martian peridotite. Following the argument of Kiefer et al. (2015) concerning the effect of Na, K, and Ca, which is still valid, then implies that it is the function from Katz et al. (2003) that should be reduced by further 35 K in order to produce a martian solidus.

Figure 10 shows the functions proposed here and demonstrates that a further reduction of the Kiefer et al. (2015) solidus would result in a function that agrees very well with ours especially in the most critical pressure range up to 3 GPa. We have plotted the solidus from Hirschmann et al. (2009) instead of the one from Hirschmann (2000), because the former can be used at pressures above 10 GPa while being identical to the latter below that pressure. The solidus modified from Kiefer et al. (2015) applies the additional shift and uses the Hirschmann et al. (2009) solidus (shifted down by 70 K) at $p > 10$ GPa to make it applicable to higher pressures.

All of these solidi have used experimental data on nominally anhydrous materials. However, as recently pointed out by Sarafian et al. (2017), even nominally anhydrous samples that have been dried in order to remove water may still contain traces of it at the level of a few dozen or hundreds of parts per million. Those authors assert, on the basis of experiments of their own that apply a novel technique for monitoring water content, that previous studies systematically underestimated solidus temperatures because of such trace water. The considerations above concerning the effect of Na, K, and Ca, which is still valid, then implies that it is the function from Hirschmann (2000) solidus was based and deduced that an increase of the solidus temperature by $\sim 60$ K is necessary to correct this
error. This would imply that the solidus originally determined by Hirschmann et al. (2009) for terrestrial anhydrous peridotite would in fact have to be lowered by mere 10 K to apply to Mars. However, we think that the data to devise a more generally applicable correction are not yet sufficient at this point, and so we stick to the conventional datasets.

B Estimate of the geometrical error in 2D global mean calculations

We consider a single axisymmetric anomaly on the surface of a spherical planet. In two dimensions, such an anomaly would be represented by a cross-section through the center of the sphere and the center of the anomaly, and the anomaly would appear as an arc of length $\theta$, i.e., it occupies a fraction $f_2 = \theta/(2\pi)$ of the perimeter of the cross-section. By contrast, in three dimensions the anomaly is a spherical cap that occupies a fraction

$$f_3 = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\frac{\theta}{2}} \sin \theta \, d\theta \, d\phi = \frac{1}{2} \left( 1 - \cos \frac{\theta}{2} \right)$$

(14)

of the total surface of the sphere. Hence, the amplitude marking the impact in the two-dimensional model should be multiplied by a factor

$$\frac{f_3}{f_2} = \frac{\pi}{\theta} \left( 1 - \cos \frac{\theta}{2} \right),$$

(15)

with $\theta$ in radians; the discrepancy is worst for small anomalies. For the Huygens, Isidis, and Utopia events, we have angles $\theta$ of 1.26°, 3.64°, and 9.1°, respectively, and so the correction factors would be 0.0086, 0.025, and 0.062, respectively.
Notation

The subscript “imp” indicates a property of the impactor throughout the text.

\( a, b \) scaling and shift parameters for \( p_s \) after \( \text{Ruedas} \) (2017)
\( B \) dislocation creep pre-factor
\( c_p \) heat capacity
\( C_i \) concentration of \( i \)
\( D_f \) final crater diameter
\( D_{imp} \) impactor diameter
\( D_{tr} \) transient crater diameter
\( D_{s2c} \approx 5.6 \text{ km} \) simple/complex transition crater diameter (Mars)
\( f \) melting degree/depletion
\( G_0 = 6.67408 \times 10^{-11} \text{ m}^3/(\text{kg s}^2) \) Newton’s constant of gravitation
\( g = 3.72 \text{ m/s}^2 \) surface gravity acceleration (Mars)
\( \Delta g_z \) vertical component of gravity anomaly
\( K_S \) adiabatic bulk modulus
\( n \) decay exponent for \( p_s \) after \( \text{Ruedas} \) (2017)
\( p_{IM} \) shock pressure from impedance-match solution
\( p_l \) lithostatic pressure
\( p_s \) peak shock pressure
\( Q \) activation enthalpy
\( q_s \) surface heat flow
\( R_{gas} = 8.3144598 \text{ J/(mol K)} \) universal gas constant
\( R_P = 3389.5 \text{ km} \) planetary radius (Mars)
\( r_{infl} \) radius of isobaric core defined by inflection point of \( p_s \)
\( r_{PVM} \) radius of isobaric core after \( \text{Pierazzo et al.} \) (1997)
\( R_{00}, R_{0\text{eff}} \) initial, effective Rayleigh number
\( S_h \) Hugoniot slope
\( T \) temperature
\( T_{pot} = 1700 \text{ K} \) initial potential temperature
\( T_s \) solidus temperature
\( u_{IM} \) particle velocity from impedance-match solution
\( v_B \) speed of sound
\( v_{imp} = 9.6 \text{ km/s}, v_{z,imp} = 6.8 \text{ km/s} \) impact velocity and its vertical component (Mars)
\( v_P, v_S \) P-, S-wave velocity
\( z_{el} \) elastic thickness of the lithosphere
\( z_{ic} \) depth to center of isobaric core
\( \dot{\varepsilon} = 10^{-17} \text{ s}^{-1} \) strain rate
\( \eta_{\text{mean}} \) volumetric mean of viscosity, nondimensionalized
\( \phi \) porosity
\( \varphi_t = 0.067 \) threshold porosity for melt extraction
\( \gamma \) Grüneisen parameter
\( \nu \) exponent of creep law
\( \varrho \) density
\( \sigma_y = 15 \text{ MPa} \) yield stress

Acknowledgments

Anne Hofmeister and Zhicheng Jing kindly gave additional explanations of various aspects of their respective work. We thank Martin Knapmeyer for helpful remarks on some seismological aspects and Sabrina Schwinger for helping to clarify a question concerning the water content of the crust. The comments by Gregor Golabek and an anonymous referee are gratefully acknowledged. TR was supported by grant Ru 1839/1-1 from the Deutsche Forschungsgemeinschaft (DFG), with additional initial funding from the Helmholtz Alliance project “Planetary evolution and life”. DB was supported by the DFG (SFB-TRR 170). This is TRR 170 publication no. 28. The numerical calculations were carried out on the computational resource ForHLR II at the Steinbuch Centre for Computing, Karlsruhe Institute of Technology, funded by the Ministry of Science, Research and the Arts Baden-Württemberg and DFG. Some earlier code development first applied in this study benefited from support by the NASA Planetary Geology and Geophysics Program through grant NNX11AC62G and the use of computing resources provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center, award SMD-11-2549. TR is grateful to Sean Solomon for having made this support from NASA possible. The data used in this investigation
are presented in this paper and the Supporting Information; data sets used in plots and figures can be obtained from Figshare at https://doi.org/10.6084/m9.figshare.c.3777686.

References


Stixrude, L., and M. S. T. Bukowinski (1990), Fundamental thermodynamic relations and silicate melting with implications for the constitution of D'', *J. Geophys. Res.*, 95(B12), 19,311–19,325.


1 Thermoelastic model

The starting point for the calculation of the thermoelastic properties of the solid silicate part of the planet is the set of mineral endmember properties, which we assume to form ideal solid solutions. We have collected a large number of experimental data for many endmembers of the most common (basaltic) crust and mantle minerals and fitted them to the third-order Birch–Murnaghan equation of state and to the Mie–Grüneisen–Debye equation of state in order to determine the key properties, which are listed in Table 1. Possible effects of water on these properties have been neglected at this point, because the water concentrations of the models in this paper are generally rather low, and for many minerals water-related effects are poorly known. Some hypothetical endmembers that are needed for the purpose of calculation but do not actually exist have also been included; their properties were inferred from systematics, as were properties of phases for which no or insufficient data were available.
**Theoretical framework**  The thermoelastic properties of the mineral endmembers are formulated in the framework of quasi-harmonic theory. In the harmonic approximation, the internal energy of a crystal at temperature \( T \) is split into a “cold” contribution that accounts for the static pressure and compression and a vibrational thermal part (Grüneisen, 1912):

\[
E = E_c + E_{\text{vib}} = E_c + \sum_{i=1}^{3N} \hbar \omega_i \left( \frac{1}{2} + \frac{1}{\exp \left( \frac{\hbar \omega_i}{k_B T} \right) - 1} \right),
\]

where the sum is taken over the \( 3N \) modes \( \omega_i \) of an \( N \)-oscillator crystal (e.g., Ashcroft and Mermin, 1976). A similar additive representation also holds for the Helmholtz potential, from which all thermodynamic properties can be derived. The cold compressional contribution at pressure \( p \) to the Helmholtz potential depends on the choice of equation of state; we use the third-order Birch–Murnaghan equation of state,

\[
p = \frac{3}{2} K T_0 \left[ \left( \frac{V_0}{V} \right)^{7/3} - \left( \frac{V_0}{V} \right)^{5/3} \right] \left\{ 1 + \frac{3}{4} (K' T_0 - 4) \left[ \left( \frac{V_0}{V} \right)^{2/3} - 1 \right] \right\},
\]

which has been the most popular choice in the literature for decades; \( V, K_T, \) and \( K'_T \) are the molar volume, the isothermal bulk modulus, and its pressure derivative, respectively, and the subscript 0 indicates the value at standard pressure.

The thermal contributions are described by Mie–Grüneisen–Debye theory. The Debye model of phonons (Debye, 1912) introduces a simplified description of the phonon spectrum based on a parabolic density of states that is cut off at a characteristic frequency \( \omega_D \), corresponding to a temperature \( \theta_D = \hbar \omega_D / k_B \), the Debye temperature; \( \hbar \) and \( k_B \) are Planck’s and Boltzmann’s constants, respectively (the former divided by \( 2\pi \)). With this simplification, the vibrational energy becomes

\[
E_{\text{vib}} = 9NRT \left( \frac{T}{\theta_D} \right)^3 \int_0^{\frac{\theta_D}{T}} \frac{t^3}{\exp t - 1} \, dt = 3NRT D_3 \left( \frac{\theta_D}{T} \right),
\]

i.e., it approaches \( 3NRT \) at \( T \gg \theta_D \), and with \( F = E - TS \) the Helmholtz free energy becomes

\[
F_{\text{vib}} = 9NRT \left( \frac{T}{\theta_D} \right)^3 \int_0^{\frac{\theta_D}{T}} \ln(1 - \exp(-t)) t^2 \, dt; \quad (4)
\]

\( D_3 \) in eq. 3 is the Debye function of order 3. The characteristic temperature is assumed to depend explicitly only on volume (Grüneisen, 1926) in a manner described by the thermal Grüneisen parameter

\[
\gamma = \gamma_0 \left( \frac{V}{V_0} \right)^q \quad (5a)
\]

\[
\theta_D = \theta_{D0} \exp \left( \frac{\gamma_0 - \gamma}{q} \right) = \theta_{D0} \exp \left\{ \frac{\gamma_0}{q} \left[ 1 - \left( \frac{V}{V_0} \right)^q \right] \right\}; \quad (5b)
\]

the parameter \( q \) is often close to 1 (e.g., Stixrude and Bukowinski, 1990). The second
temperature derivative of $F_{\text{vib}}$ yields the isochoric heat capacity,

$$c_V = -T \left( \frac{\partial^2 F}{\partial T^2} \right)_V = 9NR \left( \frac{T}{\theta_D} \right)^3 \int_0^{\theta_D} \frac{t^4 \exp t}{(\exp t - 1)^2} dt = 9NR \left( \frac{T}{\theta_D} \right)^3 \int_0^{\theta_D} t^2 \mathcal{E}_2(t) dt,$$

(6)

where $\mathcal{E}_2(t)$ is the Einstein function of order 2. The volume derivative of $F_{\text{vib}}$ gives the thermal pressure,

$$p_{\text{th}} = \frac{\gamma}{V} (E_{\text{vib}}(V,T) - E_{\text{vib}}(V,T_0)),$$

(7)

and the thermal contribution to the isothermal bulk modulus,

$$K_{T,\text{th}} = \frac{\gamma}{V} \frac{c_p K_T}{c_V} = K_T(1 + \gamma \alpha T),$$

(8)

(9)

(10)

(11)

As closed-form solutions of the integrals in eqs. 3, 4, and 6 are not available, approximate solutions based on summation formulae are used. The integral in the expression for the vibrational energy (eq. 3) can be expressed as

$$\int_0^{\theta_D} \frac{t^3}{\exp t - 1} dt = \begin{cases} \left( \frac{\theta_D}{T} \right)^3 \left[ \frac{1}{3} - \frac{1}{8} \frac{\theta_D}{T} + \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k+3)(2k)!} \left( \frac{\theta_D}{T} \right)^{2k} \right] & \theta_D \lesssim 0.5 \\ 6 \zeta(4) - \sum_{k=1}^{\infty} \exp \left[ -k \frac{\theta_D}{T} \right] \left[ \frac{1}{k} \left( \frac{\theta_D}{T} \right)^3 + \frac{3}{k^2} \left( \frac{\theta_D}{T} \right)^2 + \frac{6}{k^3} \left( \frac{\theta_D}{T} \right) + \frac{6}{k^4} \right] & \text{else,} \end{cases}$$

(12)

where $B_{2k}$ are the Bernoulli numbers and $\zeta$ is the Riemann zeta function (Abramowitz and Stegun, 1964, eqs. 27.1.1–3 with $n = 3$).

The integral in the expression for the specific heat (eq. 6) belongs to a class of functions that are definite integrals of the Einstein functions of order $s$,

$$I_n(s, x_D) = \int_0^{x_D} \frac{x^n \exp x}{(\exp x - 1)^s} dx = \int_0^{x_D} x^{n-s} \mathcal{E}_s(x) dx,$$

(13)

for which Mamedov et al. (2009) have derived the summation formula

$$I_n(s, x_D) = \sum_{m=0}^{\infty} (-1)^m \binom{-s}{m} H_{nm}(s, x_D)$$

(14)

with

$$H_{nm}(s, x_D) = \frac{n!}{(m + s - 1)^{n+1}} \left\{ \frac{n}{(m + s - 1)^{n+1}} \left[ 1 - \exp[-x_D(m + s - 1)] \sum_{k=0}^{n} \frac{(x_D(m + s - 1))^k}{k!} \right] \right\}.$$
Using
\[
\begin{align*}
\left( \frac{-s}{m} \right) &= (-1)^m \binom{s + m - 1}{m} = (-1)^m \frac{(s + m - 1)!}{m!(s - 1)!} \\
\text{(14b)}
\end{align*}
\]
and setting \( n = 4 \) and \( s = 2 \) as in eq. 6, eq. 14 can be simplified to
\[
I_4(2, x_D) = \sum_{m=0}^{\infty} (m + 1) H_{4m}(2, x_D).
\text{(14c)}
\]

The integrals for the vibrational Helmholtz potential and its temperature derivative, the entropy \( S \), can be expressed with similar formulas using partial integration if desired.

**Estimates of unconstrained parameters**  In some cases, there are no data for a given endmember, for instance if that endmember is not thermodynamically stable and hence hypothetical, as is the case with Fe-wadsleyite, or if there are simply no experimental data. In such cases, the parameters have to be estimated using available data for similar minerals. In order to estimate the molar volume, we consider the ratios of known \( V_0 \) of minerals with similar structure or composition; for instance, the molar volume of Fe-wadsleyite is estimated from Mg-wadsleyite, ringwoodite, and ahrensite assuming that
\[
\frac{V_{0,\text{Fe}w}}{V_{0,\text{Mg}w}} = \frac{V_{0,\text{ah}}}{V_{0,\text{ri}}}. \text{(15)}
\]
For estimates of the bulk modulus, we use the empirical relation
\[
MV = \text{const.} \Leftrightarrow MV = M_{\text{ref}}V_{\text{ref}} \text{(16)}
\]
(e.g., Duffy and Anderson, 1989), where \( M \) is an elastic modulus. The \( p \) derivative is unfortunately not well captured by any empirical relation, but differentiating with \( M = K_T \)
yields
\[
K_T' = (K_T'_{\text{ref}} - 1) \frac{V_{\text{ref}}}{V} + 1; \text{(17)}
\]
it generally does not show very large variation. The thermal parameters could be estimated based on the assumption that volume is the only factor that affects \( \gamma \) and \( q \) so that the Grüneisen parameter could be estimated using eq. 5a with \( \gamma_0 \) and \( V_0 \) from the known substance and \( V \) for the unknown substance, i.e.,
\[
\gamma = \gamma_{\text{ref}} \left( \frac{V}{V_{\text{ref}}} \right)^{q_{\text{ref}}}, \text{(18)}
\]
with \( q = q_{\text{ref}} \), as \( q \) would be poorly constrained by any available data. We emphasize that this is an ad hoc ansatz, and we will investigate in more detail below how well it works.

To determine \( \theta_D \), we first consider the (elastic) Debye temperature:
\[
\theta_D = \frac{h}{k_B} \left( \frac{3NN_A}{4\pi V} \right)^{\frac{1}{3}} v_D = \frac{h}{k_B} \left( \frac{9NN_A}{4\pi V \left( \frac{1}{v_P^2} + \frac{2}{v_S^2} \right)} \right)^{\frac{1}{3}} = \frac{h}{k_B} \left( \frac{9NN_A}{4\pi V \left[ 2 + \left( \frac{9}{v_P} \right)^3 \right]} \right)^{\frac{1}{3}} v_S \text{(19a)}
\]

\(4\)
(Poirier, 2000, eq. 3.36), where \( v_D, v_p, \) and \( v_S \) are the Debye, compressional, and shear velocity, respectively. Assuming that the mineral is a Cauchy solid \( (v_p = \sqrt{3}v_S) \), the results of the parameters already determined for the cold compression can be used together with

\[
v_D = \sqrt{\frac{3K_S}{5\varrho}} \left( \frac{3}{2 + \left(\frac{1}{3}\right)^{\frac{3}{2}}} \right)^{\frac{1}{3}} \approx 0.86 \sqrt{\frac{K_S}{\varrho}} \approx \sqrt{\frac{K_TV}{m_{mol}}},
\]

(19b) to deduce the unknown Debye temperature:

\[
\frac{\theta_D}{\theta_{D,ref}} = \left( \frac{N}{N_{ref}} \right)^{\frac{1}{3}} \left( \frac{K_{T,nmol,ref}}{K_{T,ref,m_{mol}}} \right)^{\frac{1}{2}} \left( \frac{V}{V_{ref}} \right)^{\frac{1}{6}}; \tag{19c}
\]

often one will choose the reference mineral so that \( N = N_{ref} \).

In order to assess the quality of such estimates and test what sort of analogue materials are suitable for applying such systematics, we can make such systematic comparisons for endmembers for which there are actual measurements to compare with; the data are taken from the database of which Table 1 shows a subset. The plots in Fig. 1 show predicted vs. measured values for several materials, grouped by mineral structure and composition. The groups are Olivine (forsterite and fayalite), Orthopyroxene (orthoenstatite and orthoferrosilite), Clinopyroxene (2–2) (Clinopyroxenes with one of the bivalent cations Mg\(^{2+}\), Fe\(^{2+}\), and Ca\(^{2+}\) on each of the M1 and M2 sites), Clinopyroxene (1–3) (Clinopyroxenes with Na\(^{+}\) as the monovalent cation and one of the trivalent cations Al\(^{3+}\), Fe\(^{3+}\), and Cr\(^{3+}\) on the M1 and M2 sites), Al–Cr spinel (Spinel with Mg\(^{2+}\) or Fe\(^{2+}\) as the bivalent ion and Al\(^{3+}\) or Cr\(^{3+}\) as the trivalent ion), Si spineloid (the high-\(p\) polymorphs of olivine, i.e., Mg-wadsleyite, ringwoodite, ahrensite), Al–Cr–Fe garnet (the various garnets with the bivalent cations Mg\(^{2+}\), Fe\(^{2+}\), or Ca\(^{2+}\) on the X site and the trivalent cations Al\(^{3+}\), Fe\(^{3+}\), and Cr\(^{3+}\) on the Y site), Mg–Al/Si garnet (pyrope and Mg-majorite), Ilmenite (Mg-akimotoite, corundum, eskolaite), Oxide (periclase and wüstite), and Perovskite/postperovskite (bridgmanite, Ca-perovskite, Mg-postperovskite).

In order to estimate an unknown \( V_0 \) from similar minerals, three reference materials are needed according to eq. 15, which limits the number of data available for testing. For the three mineral groups for which there are sufficient data, the estimates are very accurate (Figure 1a).

For the other parameters, only one reference mineral is required. In Figures 1b–e, data are shown for two sets of calculations. In the first one, which is plotted with solid symbols, only the target parameter is assumed to be unknown but the other properties, especially \( V_0 \) are derived from experiments. In the other one, which is plotted with open symbols, all properties of the target materials are assumed to be unknown, i.e., \( V_0, K_{T0}, K'_{T0}, \gamma_0, q, \) and \( \theta_D \) are derived from systematics, implying that errors propagate through the steps of the calculation. In most calculations, especially those of the first set, we chose the reference material so that it differs from the target material only in one type of ions. For some mineral groups, we also considered reference materials differing by two types of ions or having a slightly different crystal structure (Si spineloid, Perovskite/postperovskite); this subset is plotted in a paler hue of the same color.

The estimates for \( K_{T0} \) indicate that eq. 16 works quite well for several mineral groups, even if the reference mineral differs by more than a single ion (Figure 1b). The reliability of the formula deteriorates towards materials with higher \( K_{T0} \), which are typical for higher-pressure phases. Unfortunately, the performance does not generally remain at that level.
for $K'_{T_0}$ (Figure 1c). The reason is at least in some cases that the $K'_{T_0}$ values of the reference materials are themselves not very well constrained. The estimates for $\gamma_0$ and $\theta_D$ are satisfactory in many cases (Figure 1d,e), but it is probably always safer to make several estimates with different minerals if possible, and to observe whether they cluster around certain values.

**Database assembly and fitting procedure** The thermodynamical calculations are based on a large set of experimental data for the endmembers. The database was assembled according to the following principles:

- It was tried to use observables that can be measured as directly as possible, i.e., without intermediate steps or additional assumptions or data. Specifically, this restricts the data to direct measurements of volume $V$ (or rather the lattice parameter) as a function of $p$ and/or $T$, e.g., by X-ray crystallography, and calorimetric measurements that yield $c_p$ or the differential enthalpy $\Delta H = H(T) - H(T_{\text{ref}})$. Density measurements from immersion (Archimedean) experiments were admitted if they agreed reasonably well with the other STP data and/or the sample had a very low ($\ll 1\%$) porosity; however, if X-ray densities were available for the same sample, the X-ray density is preferred.

- In addition to the STP measurements from compression or expansion studies, an effort was made to include many additional measurements conducted independently only at standard conditions, e.g., for structure refinement. The purpose is to constrain the reference volumes $V_0$ as tightly as possible rather than treating it as a freely adjustable parameter in the cold compression fit, because that would give the low-$T$, high-$p$ data an influence on $V_0$ but not the low-$p$, high-$T$ data, which does not seem to be justified.

- Results from shock experiments are not admitted, because the temperature is usually poorly constrained.

- Results from ab-initio, molecular dynamics, or similar numerical calculations are not admitted.

- If possible, only data from chemically pure endmember samples were used, whereby “pure” is defined as “containing less than 1 mol% components that are not part of the ideal endmember composition”. Only if pure samples were not available, samples with more impurities were also taken into consideration.

- Frequently, the temperature at which isothermal compression experiments are conducted is only specified as “room temperature”, or no information is given about the temperature at all. In such cases, the experiments are assumed to have been carried out at “room temperature”, which is assumed to be 298.15 K (25°C).

- Some fits were restricted if necessary due to poor convergence. Most importantly, $K'_{T_0}$ was never allowed to be smaller than $5/3$ in order to enforce a thermodynamic condition for $p \to \infty$ (e.g., Stacey and Davis, 2004), nor was it allowed to exceed 9. We emphasize that calculations at high $p$ for minerals in which the fit gave such extreme values should be done with particular care if they cannot be avoided.
Figure 1: Observed vs. predicted parameters from systematic comparisons of endmembers of different mineral types for $V_0$ (a), $K_{T0}$ (b), $K_{T0}'$ (c), $\gamma_0$ (d), and $\theta_D$ (e). Solid symbols indicate estimates based on the assumption that only the target variable is unknown, open symbols indicate estimates assuming that all parameters of the target mineral are unknown. In most cases, the target mineral and the reference mineral differ only in one cation; the cases in which they differ in two ions (Al–Cr spinel, Al–Cr–Fe garnet, ilmenite) or structure (perovskite/postperovskite) are shown in a paler hue of the same color.
• Data from tabulated, “smoothed” datasets such as the data tables by Robie et al. (1978) or Robie and Hemingway (1995) are avoided, because the fitting model necessarily introduces a certain bias towards that model.

• If possible, original data were taken from tables in the respective publication. In numerous cases, however, authors did not tabulate their data but provided only plots. In this case, the data points were taken from a digital version of the plot with a dedicated tool trying to preserve the highest accuracy permitted by the quality of the plot. We expect the accuracy of these readings to be generally better than experimental error, especially in compression experiments.

• Only data from 1950 or later are admitted.

The final material parameters are determined in several steps:

1. If volume measurements at standard conditions were available, \( V_0 \) was determined as the arithmetic mean of all accepted \( V_0 \) values and kept fixed afterwards.

2. The room-temperature compression data are fitted to the isothermal equation of state with a Levenberg–Marquardt algorithm in order to determine \( K_{T0} \) and \( K'_{T0} \).

3. If differential enthalpy measurements are available, each such dataset is recalculated into a heat capacity dataset individually. In the literature, it is common to give a mean isobaric specific heat by a finite difference form of its relation with the enthalpy,

\[
c_p = \left( \frac{\partial H}{\partial T} \right)_p \Rightarrow c_{p,\text{mean}} = \frac{H(T) - H(T_{\text{ref}})}{T - T_{\text{ref}}},
\]

(20)

\[
giving a mean specific heat; the actual \( c_p \) dependence is often cast in the form of empirical polynomial formulae whose coefficients are determined by integration and fitting to the \( \Delta H \) data. As we do not use such formulae, we chose to approximate the actual \( c_p(T) \) by

\[
\frac{\Delta H_2 - \Delta H_1}{T_2 - T_1} = \frac{H_2 - H_1}{T_2 - T_1} \approx \left( \frac{\partial H}{\partial T} \right)_p = c_p
\]

(21)

or indeed by more accurate higher-order finite-difference expressions (e.g., Fornberg, 1998); note that the expressions must be suitable for unequally spaced samples, because experimental data are often not equally spaced so that the familiar FD operators for equispaced grids cannot be used. This method is found to yield satisfactory results if the dataset is of sufficiently high quality.

4. A Simulated Annealing algorithm is used to fit the standard-pressure datasets of volumetric expansion, thermal equation of state, and isobaric heat capacity separately and together.

5. All fits as well as all curves derived from systematics are controlled visually, and parameters are adjusted by trial and error if necessary, often with the aid of systematics. In the case of the isothermal fits, this was not usually necessary, but the thermal fits frequently required such adjustments because the results of the fitting procedure did not produce physically acceptable matches to all available data. The adjustments had the goal to approximate all datasets as well as possible visually and physically; in particular,
• volumetric expansion should be a monotonically increasing function whose slope approaches zero for $T \to 0$;
• isobaric heat capacity should be a non-decreasing function if possible at all, or at least show only a minimal decline at $T \gg \theta_D$.

Especially the $c_p$ datasets frequently make the thermal fits difficult, because often $\theta_D$ is not, or only barely, reached, and the data are insufficient for $\theta_D$ to be found properly by the fitting algorithm; as a result, the algorithms would often find too small values. Volumetric expansion is fairly insensitive towards the choice of $\theta_D$ and does not seem very useful to constrain this parameter.

The thermoelastic parameters determined by this approach that we used in our modeling study are summarized in Table 1, but we emphasize that this compilation and the fits are continually improved as new data become available and that we hope to remove shortcomings of the current fit in the future.

**Properties of the bulk rock** In order to determine the properties of the bulk mantle material from the endmember properties, two more steps have to be completed, namely the calculation of the properties of the solid solution minerals that represent the individual mineral phases and the derivation of the properties of the composite formed by them.

For the solid solution properties we follow the practice of many previous authors to assume ideal solid solution of endmembers, which often is a fairly good approximation and accurate enough for our purposes. Ideal mixing implies that the molar volume of the solid solution, $V_{ss}$, is the linear combination of the molar volumes of the endmembers, $V_i$, with the weights given by their respective molar fractions $X_i$ as determined from bulk chemistry and, if melting has occurred, with the help of the partitioning coefficients described in more detail in Section 3. The density of the solid solution is calculated from $V_{ss}$ and the mole-weighted average of the end-member molecular masses, and the thermal expansion coefficient and the isothermal and adiabatic bulk moduli then follow from their thermodynamic definitions:

$$\alpha_{ss} = \sum X_i \frac{V_i}{V_{ss}} \left( \frac{\partial}{\partial T} \right)_p \left( \sum X_i V_i \right) = \frac{\sum X_i V_i \alpha_i}{V_{ss}},$$

$$K_{T,Sss} = \left[ \sum X_i \frac{V_i}{V_{ss}} \left( \frac{\partial}{\partial p} \right)_{T,S} \left( \sum X_i V_i \right) \right]^{-1} = \frac{V_{ss}}{\sum \frac{X_i V_i}{K_{Ti}}},$$

Assuming complete relaxation, which should be applicable to long-duration processes in planetary interiors, the expansivity and bulk moduli of the composite material are then estimated as isostress (Reuss) averages of the properties of the component minerals (e.g., Stacey, 1998):

$$\alpha_{mc}^R = \frac{1}{m_{mol} \sum X_i \frac{V_i}{m_{mol_i}}} \left( \frac{\partial}{\partial T} \right)_p \left( m_{mol} \sum X_i \frac{V_i}{m_{mol_i}} \right) = \varrho_{mc} \sum X_i \frac{V_i}{\varrho_i} = \sum \tilde{X}_i \alpha_i$$

$$K_{mc}^R = \frac{1}{\varrho_{mc} \sum \frac{X_i}{\varrho_i K_i}} = \frac{1}{\sum \frac{\tilde{X}_i}{K_i}},$$

where $X_i$ and $\tilde{X}_i$ are the mass and volume fractions of the constituents, respectively. For the calculation of seismic velocities, it is not appropriate to assume relaxation, however. One
and to compute the effective composite modulus as the arithmetic mean of the Reuss and Voigt bound. The bulk rock density is determined from the densities of the constituent minerals:

$$
\rho_{mc} = \left( \sum \frac{X_i}{\rho_i} \right)^{-1} = \sum \frac{X_i}{\rho_i}. 
$$

Table 1: Thermoelastic parameters of mineral endmembers for the third-order Birch–Murnaghan and Mie–Grüneisen–Debye equations of state. Properties inferred from systematics, in particular of hypothetical phases, are italicized.
There are tighter bounds, such as the Hashin–Shtrikman bounds, but for the time being, we use the VRH average for the calculation of the bulk sound velocity.

The thermal conductivity $k$ is linked to some extent to the thermoelastic properties and was also calculated in two steps. $k$ and the thermal diffusivity $\kappa = k/(\varrho c_p)$ were determined for each (solid-solution) mineral phase individually using a representation of the temperature and composition dependence based on the $\kappa$ parameterizations from Hofmeister and Branlund (2015) for the phonon contribution and adding the (relatively unimportant) pressure-independent radiative component as given by Hofmeister (1999). The pressure dependence of the phononic part of $k$ and $\kappa$ is not well constrained experimentally, and it was assumed here that $k(p)$ scales with the bulk modulus (e.g., Hofmeister and Branlund, 2015). The conductivity of the bulk material was taken to be the mean of the generalized Hashin–Shtrikman bounds for a composite of the individual minerals (Berryman, 1995).

Near the surface, the material is generally assumed to have a certain porosity. We keep the approach already used by Ruedas et al. (2013a) to adopt a description by Leone et al. (2011) based on work by Head and Wilson (1992), which provides pressure-dependent formulae for the density and thermal conductivity of porous materials:

$$\varrho(p) = \varrho_{\text{solid}}(1 - \varphi), \tag{28}$$

$$k(p) = k_{\text{solid}}(1 - 1.738\varphi + 0.8228\varphi^2), \tag{29}$$

where $\varphi = \varphi_{\text{surf}} \exp(-11.8p)$ is the porosity, $\varrho_{\text{solid}}$ and $k_{\text{solid}}$ are the density and thermal conductivity of the compact solid material, respectively, and the pressure $p$ is in GPa. For the seismic velocities, we calculate the bulk modulus based on that formula for the density of the porous material and the thermodynamic definition, $K = \varrho \partial p/\partial \varrho$, for physical consistency.

## 2 Rheological model

The viscosity $\eta$ in our models is the same as in Ruedas et al. (2013b) and depends on temperature, pressure, melt, iron, and water content, and mineralogy. We used the general law for diffusion creep given by Hirth and Kohlstedt (2003, eq. 1), scaled with an additional factor to take into account the effect of an elevated iron content, after Zhao et al. (2009). The resulting formulation is therefore a function of pressure, temperature, porosity (in this case corresponding to the melt content) $\varphi$, Mg#, $C_{\text{OH}}$, and mineralogy $X$:

$$\eta(p, T, \varphi, \text{Mg}#, C_{\text{OH}}, X) = \eta_p(X) \frac{b^3}{AC_{\text{OH}}} \left( \frac{1 - \text{Mg}#_{\text{ref}}}{1 - \text{Mg}#} \right)^{1.5} \exp \left[ \frac{E + pV + c(\text{Mg}#_{\text{ref}} - \text{Mg}#)}{RT} - a\varphi \right]; \tag{30}$$

in the dry case, $C_{\text{OH}}$ is dropped from the formula, and the parameters for dry material are used. In practice, $\eta$ is calculated for both the dry and the wet case, and the smaller value is used; this corresponds to a transition from wet to dry material if the concentration drops below a few ppm water.

We base the rheology on the model of a load-bearing framework, which means that the rheological properties of the volumetrically dominant mineral phase govern the behavior of the whole rock. In a peridotite, this would be olivine, and hence, it is the water concentration in olivine as determined via the partition coefficients (see Section 3) rather than the bulk
water concentration that is to be used for $C_{OH}$. For the reference case of a lherzolitic upper-mantle mineralogy, we used the values for dry or wet olivine ($A = 1.5 \cdot 10^9$ or $10^6$, $E = 375$ or $335$ kJ/mol, $V = 6 \cdot 10^{-6}$ or $4 \cdot 10^{-6}$ m$^3$/mol, respectively, and $a = 30$, Hirth and Kohlstedt, 2003); for this reference case, the petrological term $\eta_p(X)$ is 1. Grain size $b$ was fixed at 1 cm, and $\text{Mg}^{\#}_{\text{ref}}$ is 0.9. Deviations from this reference rheology due to the presence of basalt or eclogite or the transition to high-pressure polymorphs were treated by calculating volume-weighted averages with petrological factors of 2.5 (high-$p$ polymorphs of the lower mantle layer, after Jin et al., 2001), 0.25 (basalt, after Mackwell et al., 1998), or 1.2 (eclogite, after Jin et al., 2001), respectively.

3 Partition and exchange coefficients

The partition coefficient describes how a trace component $X$ is distributed between two phases $a$ and $b$:

$$D_{a/b}^X = \frac{C_b^X}{C_a^X},$$

(31)

where $C$ is the concentration (usually mass fraction); in many cases of interest, phase $b$ is silicate melt. In reality, $D$ is a function of $p$, $T$, and the composition of the solid phases and the melt, but often the available data are too scarce to provide robust information on these dependencies. This problem is particularly pressing in the case of most high-$p$ minerals, for which there seem to be hardly any melt partitioning experiments. However, mineral–mineral partitioning information is available more frequently, and if mineral–melt partitioning is known for one mineral $a$, $D$ for another mineral $b$ may be estimated by

$$\frac{C_b^{\text{liq}}}{C_b^X} = \frac{C_a^X}{C_a^{\text{liq}}} = D_{a/b}^X = \frac{D_{a/\text{liq}}^X}{D_{b/\text{liq}}^X}.$$

(32)

This equation also enables us to calculate how a component is distributed between various mineral phases, if the mineral–melt partition coefficients are known; this is useful especially in the case of water, where the unequal distribution between the different mineral phases affects the physical properties, e.g., the viscosity, of the bulk mixture of the mineral constituents.

Apart from these relations for a single component and one or more minerals, it can also be useful to consider the partitioning, or exchange, of two components $X$ and $Y$ between two phases $a$ and $b$. This exchange coefficient is the ratio of the respective partition coefficients,

$$K_{X/Y}^{a/b} = \frac{D_{X}^{a/b}}{D_{Y}^{a/b}} = \frac{D_{X}^{a/\text{liq}}}{D_{Y}^{a/\text{liq}}},$$

(33)

where phase $b$ is often melt; the partition coefficients may be either mass or molar (marked with an asterisk) thanks to canceling terms (Jones, 1995). $K$ tends to be less sensitive to $T$ and composition, and we use it in some cases for calculating the change in Mg# upon melting.

In the following, we give a brief overview of the parameterizations we used for minerals in our models. In some cases, it has been possible to put them on a physical foundation such as lattice-strain theory, but often only a simple phenomenological description is possible; it is hoped that these parameterizations will continually be improved and that formulations
more suitable for use in geodynamical models will become more common in the geochemical literature. The main purpose of the parameterizations here is to capture the general trends in the variability of the partition coefficients with \( p \), \( T \), and composition and thus improve upon models that assume constant coefficients, but the formulae given below, unless they are cited from the geochemical literature, may not always be accurate enough for use in detailed geochemical models. They are intended for use in our geodynamical model that can only include chemistry at a very primitive level.

Unless mentioned otherwise, pressures are given in GPa and temperatures are given in K. In several cases, compositional information for the mineral phases under consideration is needed; generally, the basis for deriving estimated values for use in those formulae is the silicate bulk composition of the mantle and crust of the planet.

**Olivine**

\( K_{\text{Fe/Mg}}^{\text{ol/liquid}} \) can be expressed in terms of \( p \), \( T \), the molar silica content of the melt \( \bar{X}_{\text{SiO}_2} \), and the Mg\# of the olivine as

\[
K_{\text{Fe/Mg}}^{\text{ol/liquid}} = (3.6 \bar{X}_{\text{SiO}_2} - 0.22) \exp \left( \frac{350p - 7.34T - 3766.035 - 6000\text{Mg\#}}{RT} \right)
\]

(rewritten from *Toplis*, 2005, eq. 10), where we determine \( \bar{X}_{\text{SiO}_2} \) from the bulk composition of the crust if possible.

**Taura et al.** (1998) measured partitioning between olivine and melt for various elements including K at \( T \) between 1873 and 2273 K and \( p \) between 3 and 15 GPa. Leaving out the probably spurious point for KLB-28, their data can be fitted by

\[
D_{\text{K}}^{\text{ol/liquid}} = \exp \left( 5.16279p + \frac{-10933.9p + 51567.8 - 28.3621T}{T} \right).
\]

For \( T \ll 1750 \text{ K} \) and low \( p \), this fit will quickly become unreliable, and extrapolation at the high-\( p, T \) end is not recommended either; we use the values at the experimental bounds for \( p, T \) outside this range and cap \( D_{\text{K}}^{\text{ol/liquid}} \) at 0.1, close to the highest value from *Taura et al.* (1998) deemed reliable.

For Th\(^{4+}\) and U\(^{4+}\) there are not enough data to establish a quantitative model; we use a value of \( D_{\text{Th}}^{\text{ol/liquid}} = 6 \times 10^{-5} \) and the ratio \( D_{\text{U}}^{\text{ol/liquid}} / D_{\text{Th}}^{\text{ol/liquid}} = 6.3 \) recommended by *Blundy and Wood* (2003, Tab. 6).


\[
D_{\text{H}_2\text{O}}^{\text{ol/liquid}} (p, T, C_{\text{H}_2\text{O}}^{\text{ol}}) = 0.115 \sqrt{C_{\text{H}_2\text{O}}^{\text{ol}}} \exp \left( \frac{259.0635 + 108.249p}{T} \right).
\]

**Orthopyroxene**

We determined \( D_{\text{MgO}}^{\text{ol/liquid}} \) according to the formulae in the previous section and eq. 33 and use the result in the relation

\[
D_{\text{FeO}}^{\text{ol/liquid}} = 0.129D_{\text{MgO}}^{\text{ol/liquid}} + 0.264
\]

For K, we use a constant value of $D_{K}^{\text{opx/liq}} = 4.85 \times 10^{-3}$ (van Kan Parker et al., 2010). Th$^{4+}$ and U$^{4+}$ partitioning studies are limited to basaltic compositions and indicate that $D_{U}^{\text{opx/liq}} \geq D_{\text{Th}}^{\text{opx/liq}}$, with a best estimate of $D_{U}^{\text{opx/liq}}/D_{\text{Th}}^{\text{opx/liq}} \approx 2.52$ (Blundy and Wood, 2003; Wood and Blundy, 2014).

Partitioning of water essentially follows O’Leary et al. (2010, eqs. 12, A8, A9), who parameterize it as

$$D_{\text{H}_2\text{O}}^{\text{opx/liq}} = \exp \left( -5.66 + 8.4 \hat{X}_{\text{Al}}^{\text{opx}} + 10 \hat{X}_{\text{Ca}}^{\text{opx}} \right),$$

where the atomic fraction of Al$^{3+}$ on the M1 site of opx, $\hat{X}_{\text{Al}}^{\text{opx}}$, is determined from

$$\hat{X}_{\text{Al}}^{\text{opx}} = \begin{cases} 0.032 + 0.056p, & p \leq 2.83 \text{GPa} \\ \max(0, 0.4 - 0.2 \ln p), & \text{else} \end{cases}$$

In their parameterization, $\hat{X}_{\text{Ca}}^{\text{opx}}$ has the constant value 0.074 atoms per formula unit, whereas we solve the Ca-in-opx thermometer from Brey and Köhler (1990) for it:

$$\hat{X}_{\text{Ca}}^{\text{opx}} = \exp \left( - \frac{6425 + 264p}{T} + 1.843 \right).$$

Clinopyroxene From parameterizations of the partition coefficient for MgO and FeO from Langmuir et al. (1992, tab. C1), we estimate $D_{\text{Fe}}^{\text{cpx/liq}}/D_{\text{Mg}}^{\text{cpx/liq}} \approx 0.24$.

For K, we constructed a model formula of the form used for partitioning of Na by Blundy et al. (1995) and fitted it to experimental data (Harlow, 1997; Hauri et al., 1994; Chamorro et al., 2002; Brooker et al., 2003; Dygert et al., 2014):

$$D_{K}^{\text{cpx/liq}}(p,T) = \exp \left( \frac{-6738.9 + 1974.2p + 110.35p^2}{T} - 1.6249 - 0.88006p - 0.064808p^2 \right);$$

this fit seems to work best for $p \lesssim 8 \text{GPa}$ but should not be considered better than an order-of-magnitude approximation.

For $D_{\text{Th}}^{\text{cpx/liq}}$, we use the $p$-$T$ parameterization by Landwehr et al. (2001) and derive $D_{U}^{\text{cpx/liq}}$ from the result using lattice-strain theory, applying a relation for the partition coefficients of ions of different radius but same valence (4+ in this case) (Wood and Blundy, 2014, section 3.11.2). For details, the reader is referred to those papers.

The partitioning of water is calculated again according to the formulae given by O’Leary et al. (2010, modified after eqs. 11, A5–A7):

$$D_{\text{H}_2\text{O}}^{\text{cpx/liq}} = \exp \left( -5 + 6.3 \hat{X}_{\text{Al}}^{\text{cpx}} - 1.2 \hat{X}_{\text{Ca}}^{\text{cpx}} + \frac{1600}{T} \right),$$

where the atomic fraction of Al on the tetrahedral site in clinopyroxene is

$$\hat{X}_{\text{Al}}^{\text{cpx}} = \begin{cases} 0.067 + 0.03p, & p \leq 2.6 \text{GPa} \\ \max(0, 0.26 - 0.12 \ln p), & \text{else} \end{cases}$$

and the Ca fraction on the M2 site is

$$\hat{X}_{\text{Ca}}^{\text{cpx}} = \min[\max(0, 0.7 - 0.23 \ln p), 1 - \hat{X}_{\text{Na}}^{\text{cpx}}].$$

14
Garnet  The starting point for the partitioning of Th, U as well as Fe is again lattice-strain theory. van Westrenen et al. (2001) and Wood and Blundy (2002) have developed a detailed parameterization for trivalent cations that also includes a parameterization for $D_{\text{Mg}}^{\text{gt/liq}}$. The latter can be combined with a formula for the cpx–gt exchange coefficient of Mg and Fe,

$$K_{\text{Fe/Mg}}^{\text{gt/liq}} = 0.24 \exp \left( 2.33 - \frac{3686 + 283.5p}{T} \right)$$

(Råheim and Green, 1974), and eq. 33 to determine the partitioning of Fe. Partitioning of (tetravalent) Th and U extrapolates from the parameterization for trivalent ions following Wood and Blundy (2014, eq. 29) with $\epsilon = 12$ for Mg-rich, Fe,Ca-bearing aluminous garnet (Shannon, 1993) and the ionic radii from Wood et al. (1999) for Th$^{4+}$ and U$^{4+}$; for details, the reader is referred to the original papers, especially the comprehensive review by Wood and Blundy (2014), which has been our guide in many respects.

The few data for water do not show a clear trend and are insufficient to derive a $p\cdot T$-dependent formulation, so we a constant value of 0.0026 from averaging various experimental data (Hauri et al., 2006; Aubaud et al., 2008; Tenner et al., 2009; Novella et al., 2014).

Wadsleyite  Experimental data for high-$p$ minerals are scarce. $K_{\text{Fe/Mg}}^{\text{wd/liq}}$ may vary substantially with $p$ and $T$, but $T$ for the two data points listed by Jones (1995) differs by only 60 K, and so we assumed that the variation is mostly due to $p$: 

$$K_{\text{Fe/Mg}}^{\text{wd/liq}} = \max(1.25, 5.304 - 0.2457p).$$

For Th and U, we take the values 0.013 and $1.5 \times 10^{-5}$, respectively, from Mibe et al. (2006), and for water, we multiply the partition coefficient for olivine at 13.1 GPa with $D_{\text{ol}}^{\text{H$_2$O}} = 4.95$ (Chen et al., 2002) to determine at least a constant estimate for $D_{\text{H$_2$O}}^{\text{wd/liq}}$.

Ringwoodite  We know of no data for $K_{\text{Fe/Mg}}^{\text{rw/liq}}$, so we assume that the structural similarity with wadsleyite results in a similar partitioning behavior and use a constant value of 0.39, which is the higher-$p$ value given by Jones (1995).

Some data exist for a limited $p$ range for water–melt partitioning in iron-free ringwoodite. They indicate a twofold increase of $D_{\text{H$_2$O}}^{\text{rw/liq}}$ from 20 to 23 GPa, but a clear $T$ dependence cannot be deduced (Ohtani et al., 2000). We fitted these data with 

$$D_{\text{H$_2$O}}^{\text{rw/liq}} = 5.75 \times 10^{-3}p - 0.09125,$$

which should be satisfactory, given the scarce data, because extrapolation far outside the $p$ range from 20 to 23 GPa will not be necessary, given the narrow stability field of the mineral.

Majorite  Jones (1995) has compiled some $K_{\text{Fe/Mg}}^{\text{mj/liq}}$ data for majorite from different sources. There is no clear $p$ trend but some indication for a decrease of $K_{\text{Fe/Mg}}^{\text{mj/liq}}$ with increasing $T$. We fitted these data with 

$$K_{\text{Fe/Mg}}^{\text{mj/liq}} = 1.126 - 3.14 \times 10^{-4}T,$$

but the trend for $D_{\text{Fe}}^{\text{mj/liq}}$ is even less clear, with data in the range 16–26 GPa/2173–2473 K scattering around $\sim 0.35$.  

15
Corgne and Wood (2004) have measured partitioning between majorite and melt for K, Th, and U. The partitioning of water can be inferred in two steps: from $D_{\text{H}_2\text{O}}^{\text{rw/st}} \approx 521$ and $D_{\text{H}_2\text{O}}^{\text{mj/st}} \approx 270$ (Bolfan-Casanova et al., 2000) follows $D_{\text{H}_2\text{O}}^{\text{rw/mj}} = 1.93$, and so the result for $D_{\text{H}_2\text{O}}^{\text{rw/liq}}$ in eq. 47, divided by 1.93, yields:

$$
D_{\text{H}_2\text{O}}^{\text{mj/liq}} = 2.95 \times 10^{-3} p - 0.04679.
$$

(49)

4 Description of supplementary datasets

The files in the following list can be retrieved from Figshare at https://doi.org/10.6084/m9.figshare.c.3777686.

**f.dat** Present-day depth–depletion profiles of Figure 7a of the main text. The first two data blocks contain global lateral averages of the Reference models Ref-36 (TC and T), the subsequent blocks contain local radial profiles taken at the impact sites of models H-36 (TC and T), I-36 (TC and T), U-36 (TC and T), I-72 (TC and T), and I-144 (TC and T); these local profiles are smoothed (downsampled by a factor of 3 or 4) for clarity when plotting. The columns contain depth $z$ in m, pressure $p$ in Pa, and the depletion variable $f$ ($f > 0$: depleted, $f < 0$: enriched/basaltic).

**I-36-T.f.dat** Data of the depletion ($f$) snapshots of model I-36 (T) in the rightmost column of Figure 3 of the main paper. Each data block contains the data of one snapshot. The columns contain the lateral coordinate ($\phi$) in degrees, the depth ($z$) in m, and the depletion.

**I-36-T.T.dat** Data of the temperature ($T$) snapshots of model I-36 (T) in the center-right column of Figure 3 of the main paper. Each data block contains the data of one snapshot. The columns contain the lateral coordinate ($\phi$) in degrees, the depth ($z$) in m, and the temperature in K.

**I-36-TC.f.dat** Data of the depletion ($f$) snapshots of model I-36 (TC) in the center-left column of Figure 3 of the main paper. Each data block contains the data of one snapshot. The columns contain the lateral coordinate ($\phi$) in degrees, the depth ($z$) in m, and the depletion.

**I-36-TC.T.dat** Data of the temperature ($T$) snapshots of model I-36 (TC) in the leftmost column of Figure 3 of the main paper. Each data block contains the data of one snapshot. The columns contain the lateral coordinate ($\phi$) in degrees, the depth ($z$) in m, and the temperature in K.

**marssol.dat** Compilation of experimental values used in Appendix B and Figure A.1 of the main text. The columns contain the pressure in GPa, the temperature in °C, and (for informational purposes only) the extant phases if known.

**qs_t.dat** Time evolution of the surface heat flow $q_s$ of the Reference, Huygens, and Utopia (impact at 400 My) model series for initial water contents of 36 and 144 ppm (TC and T). The columns contain time $t$ in My and $q_s$ in W/m².
**Raef10_t.dat** Time evolution of 10 My-averages of the effective Rayleigh number of the Reference and Utopia (impact at 400 My) model series for initial water contents of 36, 72, and 144 ppm (TC and T) (Figure 2 of the main text); the values are centered on the middle of the averaging interval. The columns contain time \( t \) in My and the effective Rayleigh number. One spurious average value was masked.

**Ref-36-TC_T.dat** Data of the temperature \( (T) \) snapshot of model Ref-36 (TC) in the lower right quarter of Figure 4 of the main paper. The columns contain the lateral coordinate \( (\phi) \) in degrees, the depth \( (z) \) in m, and the temperature in K.

**rho.dat** Present-day depth–density profiles of Figure 7b of the main text. The first two data blocks contain global lateral averages of the Reference models Ref-36 (TC and T), the subsequent blocks contain local radial profiles taken at the impact sites of models H-36 (TC and T), I-36 (TC and T), U-36 (TC and T), I-72 (TC and T), and I-144 (TC and T). The columns contain depth \( z \) in m, pressure \( p \) in Pa, and density \( \varrho \) in kg/m\(^3\). The very last data block contains the experimental data from Bertka and Fei (1998) as retrieved from their Fig. 1, with \( p \) in GPa and \( \varrho \) in kg/m\(^3\).

**U-36-T_T.dat** Data of the temperature \( (T) \) snapshot of model U-36 (T) in the lower left quarter of Figure 4 of the main paper. The columns contain the lateral coordinate \( (\phi) \) in degrees, the depth \( (z) \) in m, and the temperature in K.

**U-36-TC_T.dat** Data of the temperature \( (T) \) snapshots of model U-36 (TC) in the upper row of Figure 4 of the main paper. Each data block contains the data of one snapshot. The columns contain the lateral coordinate \( (\phi) \) in degrees, the depth \( (z) \) in m, and the temperature in K.

**vB.dat** Present-day depth–bulk sound velocity profiles of Figure 8 of the main text. The first data block contains global lateral averages of the Reference models Ref-36 (TC), the subsequent blocks contain local radial profiles taken at the impact sites of the TC variants of models H-36, I-36, U-36, I-72, and I-144. The columns contain depth \( z \) in m, pressure \( p \) in Pa, density \( \varrho \) in kg/m\(^3\), and the Voigt–Reuss–Hill average of the adiabatic bulk modulus \( K_S \) in Pa; the bulk sound velocity can then be calculated as \( v_B = \sqrt{K_S/\varrho} \).

**zle_t.dat** Time evolution of the elastic thickness of the lithosphere \( z_{el} \) of the Reference, Huygens, and Utopia (impact at 400 My) model series for initial water contents of 36 and 144 ppm (TC and T). For data blocks 1–12, the columns contain time \( t \) in My and \( z_{el} \) in km. The data blocks from 13 onward contain minimum and maximum estimates for the ages in Ma, for \( q_6 \) in mW/m\(^2\) (for informational purposes only, unused here), and for \( z_{el} \) in km. The ages used for plotting are the arithmetic means of the minimum and maximum ages. For \( z_{el} \), the arithmetic mean of the minimum and maximum was used where available, otherwise only the minimum value was given; the errorbars in the plot indicate the minimum and maximum values. Absence of data is indicated by a dot.

**I-36_TC.mp4** Animation of the temporal evolution of temperature and composition (depletion) in model I-36/TC. Snapshots from this animation are shown in the left panel of Figure 3 of the main paper.
I-36_T.mp4 Animation of the temporal evolution of temperature and composition (depletion) in model I-36/T. Snapshots from this animation are shown in the right panel of Figure 3 of the main paper.

References


