1
On the Origin of the Upper Mantle Seismic Discontinuities

Shun-ichiro Karato and Jeffrey Park

ABSTRACT

Recent high-frequency seismological studies revealed enigmatic structures in the upper mantle: the mid-lithosphere discontinuity (MLD) in the old continents and the oceanic lithosphere-asthenosphere boundary (LAB) at relatively shallow depth. Both discontinuities occur at relatively low temperatures (~1000°C) making it difficult to attribute them to partial melting. Also, the inferred relatively sharp and large velocity drop at these boundaries is difficult to explain with conventional smooth temperature-dependent elastic and anelastic properties from an absorption-band $Q$ model. Models that go beyond these two mechanisms are needed to explain these characteristics of the MLD and the LAB. Three models are reviewed: (a) layered anisotropy, (b) layered composition, and (c) elastically accommodated grain-boundary sliding (EAGBS). A layered anisotropy model is difficult to reconcile with the presence of these discontinuities for diverse ray geometries, and a layered composition model is not supported by the petrological/geochemical observations. EAGBS can explain globally observed substantial velocity drop at a modest temperature. This model implies that the link between seismic-wave velocity and the long-term creep strength is indirect, and predicts a peak in attenuation near these boundaries. However, experimental studies on EAGBS and geophysical tests are incomplete. Directions of future studies to test these models are suggested.

1.1. INTRODUCTION

Major seismic discontinuities both bound the transition zone (~410 to ~660 km) and occur within it. In this depth range, the seismic discontinuities are attributed mostly to phase transformations in the major mantle minerals [e.g., Ringwood, 1991]. In contrast, the volumetrically dominant upper-mantle mineral olivine suffers no phase transformations shallower than ~410 km. (There is a phase transformation in pyroxene ((Mg,Fe)SiO$_3$), the second most abundant mineral in the upper mantle, but the influence of this phase transformation on seismic-wave velocity is relatively minor [e.g., Matsukage et al., 2005].) Several seismic features interpreted as discontinuities, however, have been reported in the upper mantle [e.g., Thybo and Perchuc, 1997; Kawakatsu et al., 2009; Rychert and Shearer, 2009; Tauzin et al., 2010; Beghein et al., 2014; Hopper and Fischer, 2015; Wei and Shearer, 2017], among which we focus on two well-documented discontinuities in the upper mantle: the lithosphere-asthenosphere boundary (LAB) in both oceanic and continental upper mantle and the mid-lithosphere discontinuity (MLD) in the continental upper mantle.

At both the LAB and the MLD seismic velocities decrease with depth. Both the LAB and the MLD have been attributed to the onset of partial melting [e.g., Anderson and Spetzler, 1970; Thybo, 2006], layering in anisotropy [e.g., Yuan and Romanowicz, 2010b], layering in chemical/mineralogical composition [e.g., Selway et al., 2015], the smooth temperature dependence of seismic velocities including the influence of anelastic relaxation [e.g., Gueguen and Mercier, 1973; Faul and Jackson, 2005], and physical dispersion associated with temperature and water-dependent anelasticity facilitated by elastically accommodated grain-boundary sliding...
These models differ in their implications for the geological significance of the LAB and the MLD. Therefore it is important to understand which mechanisms may cause these discontinuities.

In this chapter, we review geophysical observations on upper mantle discontinuities and their interpretation in terms of elastic and nonelastic properties of rocks. In the next section (section 1.2), we review seismological observations on these discontinuities and emphasize the importance of using a broad range of mutually complementary seismological approaches (e.g., high-frequency body waves, low-frequency surface waves). Also, we will briefly summarize magnetotelluric (MT) observations on electrical conductivity as far as it is related to the MLD and the LAB (see also Chapters 2 and 5). In section 1.3 we review geological (and petrological) observations related to the MLD and the LAB. We review various models of the LAB and the MLD in section 1.4, including chemical layering, partial melting, and solid-state relaxation. A detailed account of the EAGBS (elastically accommodated grain-boundary sliding) model is provided in section 1.5. In section 1.6 we evaluate each model by comparing seismological (and other) observations with the predictions for each model. Finally in section 1.7 we discuss possible directions of future studies.

**1.2. Seismological Observations Relevant to the LAB and the MLD**

1.2.1. General Introduction: Long Wavelength Versus Short Wavelength Seismology

In order to understand the origin of various discontinuities, several seismological observations can be used to complement geological observations of rocks in Earth’s interior and the MT observations on electrical conductivity. These seismological observations include (a) the depth, (b) the sharpness, and (c) the magnitude of velocity change. In addition, (d) distribution and nature (geometry) of seismic anisotropy and (e) distribution of seismic attenuation will also provide data to evaluate models. In this section, we provide a review of various seismological approaches to address these issues, with specific examples from both oceanic and continental upper mantle.

The sharpness of and the velocity contrast across the mantle discontinuities play important constraints on the models to explain them. Their resolution is highly sensitive to the method used, particularly to the frequency (wavelength) of the seismic waves. In body-wave seismology, relatively high frequencies (short wavelengths) are used, whereas in surface-wave seismology to reveal global structure, low-frequency (long-wavelength) waves are used. For example, with a seismic wave of 1 s period, the wavelength for shear waves is ~4–5 km in the upper mantle, whereas for a seismic wave of 100 s period, the wavelength will be ~400–500 km. The spatial resolution of Earth’s structures that can be revealed by each type of wave is different, as shown in Figure 1.1.

Different types of seismic waves are used to investigate different regions, and consequently it is essential to understand the resolution of each study to develop a comprehensive model for the mantle. For example, as we discuss in the next section, early studies on the oceanic upper mantle used long-wavelength surface waves that revealed global structures with poor depth resolution, whereas recent studies using ocean-bottom seismometers utilize short wavelength waves that reveal fine-scale structures.

Scattered waves offer a direct inference of sharp lithospheric interfaces. Revenaugh and Jordan [1991a–c] inferred scattered waves within near-vertical ScS reverberations to infer mantle transition-zone interfaces at 410 km and 660 km and, in oceanic regions, a negative
velocity step at ~60 km depth that they named the Gutenberg discontinuity. ScS reverberations in continental regions showed no evidence of the 60 km feature. Although inferred as a sharp interface, the seismic deconvolution to detect scattered waves used a low-pass filter with a 40 mHz corner frequency (25 s period), so that a 30 km velocity gradient would appear “sharp” [Bagley and Revenaugh, 2008]. The low-passed ScS deconvolution would plausibly miss the thin low-velocity zone features within continental lithosphere inferred from body-waveform modeling [e.g., Given and Helmberger, 1980]. A subsequent ScS study by [Gaherty et al., 1999] argued that the Gutenberg discontinuity is prevalent only beneath oceans. Interpreting a suite of tomographic models from surface waves and long-period shear waves, Gung et al. [2003] proposed that the LAB globally represents a transition from stiff lithosphere into deforming anisotropic asthenosphere, at the Gutenberg discontinuity (60–80 km depth) beneath oceans, and at the Lehmann discontinuity (200–250 km depth) beneath stable continents, see also [Eaton et al., 2009]. Along this line of investigation, no MLD was detected.

Receiver functions have good potential for quantifying the sharpness of lithospheric interfaces, particularly using the higher frequency signals available in P waves (see also Chapter 8). Crustal reverberations, however, can obscure upper-mantle interfaces in P receiver functions [Zhu and Kanamori, 2000; Park and Levin, 2016], so attention to epicentral moveout effects (effects on the traveltimes due to temperature increase [e.g., Kanamori and Press, 1970; Forsyth, 1975, 1977; Yoshii et al., 1976]. Study of the structures of deep regions requires the use of long-wavelength (low-frequency) waves. This poses a serious limit for the depth resolution of the velocity structure, as discussed previously (see Figure 1.1).

The most clear evidence for an abrupt change in elastic properties at the base of the lithosphere, or within it, comes from scattered teleseismic body waves, e.g., precursors to PP and SS waves from reflections at interfaces prior to a free-surface bounce-point, and S-to-P and P-to-S converted phases at interfaces, arriving respectively as a precursor to the direct S wave or within the coda of the P wave. The frequency content of scattered waves is key to resolving the sharpness of boundaries [Olugboji et al., 2013]. If scattering is distributed evenly within a transition layer corresponding to the seismic wavelength at oscillation period T, the observed wave amplitude will be dampened by phase cancellation. Scattering within a transition layer corresponding to half a wavelength will suffer no cancellation, and the observed signal will not easily be distinguishable from a zero-thickness discontinuity. For an SS precursor of 10 s period scattered from the lithospheric mantle ($V_S \approx 4.7 \text{ km s}^{-1}$), accounting for two-way traveltime through a $\sim 24$ km gradient layer will predict substantial cancellation, and a $\sim 12$ km gradient zone will resemble a “sharp” interface. For receiver functions, the accumulation of Ps delay time corresponds roughly to 10 km s$^{-1}$ in the upper mantle, leading to substantial cancellation at 1 s period for a 10 km transition layer, and 5 km thickness as “sharp.” For S receiver functions the Sp precursor time accumulates at the same rate, so that a 5 s period Sp phase cancels over a 50 km transition, and a 25 km transition is “sharp.”

1.2.2. Some Examples: Isotropic Velocity–Depth Models

1.2.2.1. Oceanic Upper Mantle

Using the traveltime records of refracted and reflected body waves is a classic technique to determine Earth’s structure [e.g., Lay and Wallace, 1995]. Where there is a low-velocity layer, however, body-wave raypaths refract downward through the layer without bottoming, leaving the details of the velocity inversion invisible to the traveltime curve. Indeed, the presence of a low-velocity layer in the upper mantle (now referred to as the asthenosphere) was inferred indirectly based on the absence of seismic signals at a certain range of distance between the source and the receiver by Gutenberg [1926].

Consequently, surface waves are typically used to infer the velocity structure of the low-velocity layer such as the asthenosphere. In these studies, records of surface waves for a broad range of frequency are used, and the frequency-dependence of phase velocity is inverted for the depth-dependent velocity [e.g., Kanamori and Press, 1970; Forsyth, 1975, 1977; Yoshii et al., 1976]. Study of the structures of deep regions requires the use of long-wavelength (low-frequency) waves. This poses a serious limit for the depth resolution of the velocity structure, as discussed previously (see Figure 1.1).

The studies cited above show a smooth transition from the lithosphere to the asthenosphere, including the age dependence of the transition depth (Figure 1.2). Such a feature, including the age dependence of the structure of the oceanic lithosphere, can be explained by various models that include a smooth reduction of velocity due to temperature increase [e.g., Gueguen and Mercier, 1973; Schubert et al., 1976; Yoshii et al., 1976; Parsons and McKenzie, 1978; Faul and Jackson, 2005; Priestley and McKenzie, 2013; see also Chapter 6] (Figure 1.3).

High-frequency waves are now used more commonly in investigations of the oceanic upper mantle, partly due to the development of ocean-bottom seismometers (OBS) [e.g., Shimamura and Asada, 1976] that have the ability to detect subtle but sharp changes in seismic-wave velocities. Such studies have revealed more details of the velocity–depth profile of the oceanic upper mantle, including the sharp and large velocity drop at the oceanic LAB.
An example of the velocity profile of the old oceanic upper mantle is shown in Figure 1.4a. In contrast to the results from long-wavelength (long-period) seismic waves, where velocity changes smoothly with depth (Figure 1.2), a sharp and large velocity drop is seen at ~80 km depth in the old oceanic upper mantle (~130 Ma) when short-wavelength (short-period) seismic waves are used. In the old oceans, the depth at which such a sharp and large velocity drop occurs is nearly independent of the age of the ocean floor, whereas the depth increases with age in young oceans [e.g., Kumar and Kawakatsu, 2011; Rychert et al., 2012; Olugboji et al., 2016] (Figure 1.4b). A similar structure was also inferred using waveform analysis of differences in multiply reflected waves from the Pacific to infer a sharp and large velocity drop at ~60 km depth [Tan and Helmberger, 2007]. The temperature corresponding to these depths can be estimated by a thermal model of the oceanic lithosphere to be ~800–1000°C [e.g., McKenzie et al., 2005].

Some more detailed results were reported for the oceanic upper mantle. Mehouachi and Singh [2018] infer from an equatorial Atlantic marine-reflection study a low-velocity channel with 12–18 km thickness and 8.5% drop in $V_p$, the upper surface of which deepens from 72 km in 40 Ma oceanic lithosphere to 88 km in 70 Ma oceanic lithosphere. In the open Pacific Ocean near the Shatsky Rise, Ohira et al. [2017] interpret reflectors at ~50 km depth from marine reflection and refraction data as packages of low-velocity rock associated with what they term “oceanic MLDs.” These features also involve thin layers, likely to be expressed weakly in receiver functions at periods $T > 2$ s, but the global significance of these detailed structures is unknown.

Global-scale studies by Tharimena et al. [2016, 2017a, b] report the effect of lithospheric interfaces in precursors to SS waves. With dominant wave periods ~20 s, Tharimena et al. [2017a] reported 3–15% drops in $V_s$ over potentially a 21 km transition at depths that subside from 30 to 60–80 km depth from mid-ocean ridges to ~36 Ma seafloor, and flattens beneath older seafloor. The feature is characterized as “sharp” because thermal variations with depth are too smooth to explain it. Tharimena et al. [2017b] detected similar features in low-passed stacked SS precursors beneath old continents. These SS precursors imply a large velocity drop at the relatively shallow (and cold) regions, and together with the results shown in Figure 1.4 these observations provide key constraints to understanding the origin of the LAB.

### 1.2.2.2. Continental Upper Mantle

By modeling the traveltimes and waveforms of 1–10 s $P$ waves in the tectonically active southwestern United States, Burdick and Helmberger [1978] favored a $V_p$ model with a lithospheric “lid” that overlies a low-velocity asthenosphere at ~60 km depth. Although some trade-offs in model parameters were acknowledged, Burdick and Helmberger [1978] argued that their $P$ waveforms had resolved a sharp interface, a feature that we would now call the continental LAB. Benz and McCarthy [1994] confirmed the LAB sharpness in this region with data from the Polar Anglo-American Conjugate Experiment (PACE) active-source field experiment. The velocity drops inferred for the asthenosphere in these and other studies of the 1970s...
were substantial, often 5–10%. Parsons and McKenzie [1978] argued that these velocity drops were too large to be explained by thermal effects alone, and suggested that partial melt pervades the asthenosphere.

Later studies have provided a rich data set on the structures of the continental upper mantle. These studies have revealed a large regional variation in seismic structure of the continents. In active regions with extension tectonics (e.g., the Basin and Range region in the United States) the high-velocity lid (the lithosphere) is thin, whereas in the stable continents the lithosphere is thick [e.g., Grand and Helmberger, 1984; Artemieva, 2011].

Figure 1.3 Seismic-wave velocity versus depth relationship. (a) A model of seismic-wave velocity in the oceanic upper mantle based on models of the temperature and pressure dependence of seismic-wave velocity where only the anharmonic effect (effect of thermal expansion) is included. $V_p$ is $P$-wave velocity, $V_s$ is $S$-wave velocity. (After Schubert et al. [1976]. Reproduced with the permission of Wiley.) (b) A $S$-wave velocity versus depth profile of the oceanic upper mantle showing the influence of various processes: AH, anharmonicity (effect of thermal expansion); AN, anelasticity (effect of viscous dissipation). The influence of AN is included only through the absorption band model. In this case the influence of AN is limited to less than $\frac{\Delta V}{V} = Q^{-1}$ and is small (Faul and Jackson [2005] provided essentially the same results as a model of AH + AN, although we argue in the text that their model of grain-size sensitivity of seismic-wave velocity is questionable). (After Karato and Jung [1998]. Reproduced with permission of Elsevier.)
Figure 1.5 shows a reference velocity–depth model of stable (SNA) and tectonically (TNA) active regions of the continental lithosphere, as determined by S and SS waves with 5–10 s period [Grand and Helmberger, 1984]. A small velocity reduction is noted in TNA regions at ~100 km depth and SNA regions at ~150–200 km, however, the velocity–depth relation is nearly flat in the depth range of 50–150 km in the stable continents in this model.

Thybo and Perchuc [1997] reported a substantially different velocity–depth profile. They performed long-baseline refraction surveys in stable continental regions to document a loss of sharp-onset P waves at downrange distances between 8° and 11°, consistent with depressed Vp between ~100 km depth and the Lehmann discontinuity in all old continents studied, including Europe, North America, and Eurasia, and suggested that this is a global structure. Thybo [2006] summarized attempts to model refraction data with complex finite-difference synthetic seismograms, and argued that small-scale scatterers pervade the low-velocity layer at around ~100 km depth in the stable continents (Figure 1.6a).

Thybo’s [2006] inference has been confirmed by a number of later studies using converted waves (RFs; Figure 1.6b) [e.g., Rychert and Shearer, 2009; Abt et al., 2010]. These studies identified interfaces within the depth range 60–130 km at which velocity drops 6–9%. Ford et al. [2010] and Abt et al. [2010] confirmed the observations of such interfaces across Australia and North America with RFs for Sp and Ps and concluded that they were expressions of a MLD distinct from the LAB, which would occur at a greater depth in these regions. Further observations established the global character of the MLD, with RF studies in East Africa [Wölbern et al., 2012], China
average more and deeper discontinuities than those in actively
tectonically stable continents. Seismographic stations in North America—seeking evidence for the LAB and the MLD with RFs and
calibrating average seismic velocities with surface-wave
dispersion measurements into a Monte-Carlo inversion scheme that they applied to data from 30 long-running
seismographic stations in North America—seeking evidence for the LAB and the MLD with RFs and constraining average seismic velocities with surface-wave dispersion. Similar to Gaherty et al. [1999], this study formulates a general taxonomy of upper-mantle seismic discontinuities, reporting a number of velocity inversions in the depth range of 60–150 km. The inferred velocity profiles of stations in tectonically stable continent have more and deeper discontinuities than those in actively deforming terranes, in addition to a generally higher average $V_s$.

The joint inversion of RFs and surface-wave dispersion will likely aid greatly the interpretation of seismic features associated with the LAB and various MLDs. Although upper-mantle velocity inversions based primarily on surface-wave dispersion [Priestley and McKenzie, 2006; Darbyshire et al., 2007; Pedersen et al., 2009; Lin et al., 2016; see also Chapter 6] typically interpret a shallow LAB as gradational or miss the MLD entirely, dispersion observations can also be fit with low-velocity depth ranges that are bounded by sharp interfaces [Dalton et al., 2017]. The take-away message is that surface-wave dispersion can be fit with either sharp or gradational interfaces, and therefore cannot discriminate them well.

Tharimena et al. [2016] used SS precursors to infer velocity inversions below the Ontong Java Plateau (OJP) at roughly 80 km and 280 km depths, matching the pattern of the MLD and LAB under continents. The SS stacks for the OJP are modeled with sharp interfaces but could be fit with broader transitions.

### 1.2.3. Anisotropy

The one-dimensional Preliminary Reference Earth Model (PREM) [Dziewonski and Anderson, 1981] contains an anisotropic layer in the top 220 km where $S_V$ wave (vertically polarized $S$-wave) velocity is slower than that of $S_H$ wave (horizontally polarized $S$-wave) velocity. In their model $S_V$/$S_H$ radial anisotropy does not change at the depths corresponding to the LAB or the MLD.

In one region of the Pacific upper mantle, Tan and Helmberger [2007] infer $S_V$/$S_H$ anisotropy with a sharp upper boundary at the base of their lithospheric lid, but most one-dimensional seismic models lack azimuthal anisotropy. Surface waves, shear-wave birefringence and RFs have detected shifts in the orientation of azimuthal anisotropy with depth in the shallow mantle, under both continents [Levin and Park, 2000; Yuan and Romanowicz, 2010a,b; Yuan and Levin, 2014] and oceans [Beghein et al., 2014; Lin et al., 2016]. We note that Babuška et al. [1998], Babuška and Plomerová [1989], Plomerová and Babuška [2010], and Plomerová et al. [2002] also reported depth-dependent, tilted anisotropic structure in the continent using body-wave traveltime analyses. Although there is reason to associate the LAB with a transition from a relict anisotropy from lithosphere formation to
anisotropy associated with present-day asthenospheric shear, the sharpness of the anisotropic transition is not well constrained. Wirth and Long [2014] and Park and Levin [2016] report cases where RFs detect both a MLD and a change in anisotropy, but at distinct depths.

1.2.4. Attenuation

Different models have different links between velocity reduction and attenuation, and therefore observations on seismic wave attenuation combined with velocity models help identify the mechanisms for the LAB and the MLD. In order to understand how to use the observations on attenuation to evaluate models, let us review the basics of attenuation [e.g., Karato, 2008a; Jackson, 2015].

Attenuation is often characterized by a quantity called $Q$: $Q^{-1}$ provides a measure of attenuation ($Q = 100$ means 1% of energy is lost during one cycle of wave propagation). At a very general level, there is a relation between elastic and anelastic parts of wave propagation, i.e., the Kramers–Kronig relation, but the actual relationship between anelasticity (attenuation) and elasticity depends on the particular physical mechanism. To see this, let us consider two cases (Figure 1.7). Figure 1.7a is a case of a standard linear solid corresponding to a single peak in anelastic relaxation, which corresponds to a single mechanism with a single characteristic time. In this case, low velocity could correspond either to low or high attenuation depending on the frequency.

Figure 1.7b is a case where there is a broad distribution of relaxation peaks leading to a weak frequency dependence of attenuation and velocity ($Q \propto \omega^\alpha, 0 < \alpha < 1$), the “absorption band model”. In this case, the velocity anomaly is related to anelasticity as [e.g., Minster and Anderson, 1981]

$$\frac{\Delta V}{V_\infty} = \frac{1}{2} \cot \frac{\pi \alpha}{2} Q^{-1}(T, P, C_W, d) \approx Q^{-1}(T, P, C_W, d)$$

(1.1)

where $\Delta V$ is the magnitude of velocity reduction, $V_\infty$ is the velocity at infinite frequency, $T$ is temperature, $P$ is pressure, $C_W$ is water content, and $d$ is grain-size. Most previous analyses of velocity and attenuation for the lithosphere–asthenosphere system assumed the absorption band behavior [e.g., Gueguen and Mercier, 1973; Karato and Jung, 1998; Faul and Jackson, 2005; Priestley and McKenzie, 2013].

In this scenario, low velocity always corresponds to high attenuation. Most laboratory studies show this behavior [Jackson, 2015], and many seismological observations also follow this relation [e.g., Anderson and Minster, 1979; Shito et al., 2004]. One important consequence of this relation is that it (with $\alpha \approx 0.3$) predicts a small influence of anelasticity on velocity anomalies. Therefore an absorption-band model cannot explain a large velocity drop at the LAB (or MLD).

There are two cases, however, where power-law dependence in $Q$ (i.e., the absorption band model) is not valid: anelasticity associated with partial melting [Jackson et al., 2004] and EAGBS [Jackson and Faul, 2010; Sundberg and Cooper, 2010; Karato, 2012; Karato et al., 2015].

![Figure 1.7](image)

Figure 1.7 Two models of anelastic behavior: (a) a standard linear solid that corresponds to the case of a single peak; (b) absorption band model that corresponds to a case of distributed peaks. The majority of seismological and laboratory observations are consistent with this model, but some exceptions are also observed.
In both cases, attenuation has a peak at a certain frequency for a fixed temperature (or at a certain temperature for a fixed frequency). Therefore detection of such a peak will help us identify the mechanisms of formation of the LAB or the MLD.

Let us now summarize some key seismological observations on attenuation (Figure 1.8). We note first that measurements of attenuation are far more challenging than those of seismic velocities because distinguishing true (intrinsic) attenuation from scattering (extrinsic) attenuation or other geometrical effects is not straightforward [e.g., Fehler et al., 1992; Jin et al., 1994; Romanowicz and Mitchell, 2007], but we consider that the following observations are robust and important.

1. Generally, attenuation is stronger in the deeper upper mantle (~100–300 km). In the asthenosphere $Q_s$ ($Q$ for S-wave) ~50–100 [e.g., Dziewonski and Anderson, 1981].

2. Attenuation in the oceanic lithosphere is very small: $Q_s$ ~600 [Dziewonski and Anderson, 1981]; $Q_s$ ~3200 [Takeuchi et al., 2017], but substantial attenuation ($Q_s = 100–300$) is reported in the continental lithosphere [e.g., Mitchell, 1995; Dalton et al., 2009].

3. In the asthenosphere, attenuation follows the power law relations, which is consistent with absorption band behavior [Shito et al., 2004].

4. There is evidence for an attenuation peak in the asthenosphere when compared with attenuation in the lithosphere [e.g., Takeuchi et al., 2017]. Observation 4 suggests either EAGBS or a localized melt-rich region both of which would show a peak in attenuation within a thin layer near the LAB. Also, observations of modestly high attenuation ($Q_s = 100–300$) in the continental lithosphere are surprising when considering that temperatures in the continental lithosphere are generally lower than those of the oceanic lithosphere [Sclater et al., 1981]. One possible explanation is the presence of an attenuation peak in the continental lithosphere, as we shall discuss later in section 1.6.
1.3. GEOLOGICAL/PETROLOGICAL OBSERVATIONS RELEVANT TO THE LAB AND MLD

1.3.1. The LAB in the Oceanic Upper Mantle

The plate-tectonics model for the formation of oceanic lithosphere provides an important framework within which any model for a seismic discontinuity must be compatible. Oceanic lithosphere is formed at a mid-ocean ridge where hot materials rise and undergo partial melting. After reaching the thermal boundary layer, hot mantle rock diverts to near-horizontal flow and gradually cools. This cooled and hence mechanically strong region is the oceanic lithosphere.

In addition to cooling, oceanic lithosphere develops a distinct chemical composition from the partial melting under the mid-ocean ridge. The nature of this partial melting is well understood based on experimental studies as well as comparing the chemical compositions of mid-ocean ridge basalt and xenoliths from the oceanic upper mantle [e.g., Ringwood, 1975]. The major component of oceanic lithosphere is harzburgite, a depleted peridotite that contains less FeO, Al₂O₃, and CaO than primitive peridotite, as well as less water.

The changes in the major-element composition listed above do not result in substantial velocity changes (Figure 1.9) [e.g., Matsukage et al., 2005; Schutt and Lesher, 2006]. In contrast, a change in the water content

![Figure 1.9](image-url)

Figure 1.9 The influence of compositional change due to melt extraction on shear-wave velocity. After partial melting, major element composition (mineralogy) of the residual rocks changes (Mg# increases, garnet content decreases), but these changes largely cancel out and the net effect of compositional change is small (< ~1%). More drastic changes in composition are required to change the seismic velocity more than a 1–2%. (After Schutt and Lesher [2006]. Reproduced with the permission of Wiley.)
could result in a large change in seismic velocity and attenuation [e.g., Karato, 1995, 2003, 2012; Karato et al., 2015], but this has not been confirmed by laboratory studies and will be discussed in detail later in section 1.6.

1.3.2. Composition and Evolution of the Continental Lithosphere

The composition of the continental lithosphere can be inferred from the rocks carried from its deep interior by ascending melts, called “xenoliths” [e.g., Pearson et al., 2003]. In the oceanic regions, xenoliths are available only from shallow depths < 60 km, [e.g., Nicolas, 1989; Peslier et al., 2010]. From the continents, there are many xenoliths that originate from depths to ~200 km. In isolated cases, xenoliths from the transition zone or the lower mantle have been identified [Collerson et al., 2000; Pearson et al., 2014; Kaminsky et al., 2015; Nestola, 2017]. We shall focus on xenoliths coming from the lithosphere, down to ~200 km.

Processes that control the composition and mineralogy of continental mantle rocks are more complex than those of the oceanic upper mantle [e.g., Lee, 2003; Carlson et al., 2005; Schutt and Lesher, 2006]. Consequently, there is a larger degree of chemical and mineralogical variety in rocks in the continental upper mantle. As far as the compositions of dominant peridotites are concerned, the influence of composition on seismic velocity is minor (Figure 1.9).

Exceptions are the rocks formed from the freezing of a melt. If a small degree of partial melting occurs, then a large fraction of incompatible elements goes to the melt, including Fe and H (FeO and H2O) [e.g., Ringwood, 1975]. If these melts solidify to form rocks, then seismic velocity could be substantially lower than the original rocks because of high FeO content and a large amount of hydrous minerals; a point discussed later in section 1.6.

Another important constraint on the causes of MLDs is the distribution of mantle-rock ages [Carlson et al., 2005], which elucidate the evolution of continents in general. Figure 1.10 displays examples of Re–Os dating studies showing that the ages of mantle rocks drawn from the cratonic upper mantle do not vary drastically with depth and that the ages do not differ greatly from those of the crust. This implies that after its formation, most of the continental lithosphere has remained intact. Therefore, if extensive metasomatism were to be invoked, it must have occurred before the continental crust was formed in the Archean Eon (before 2.7 Ga).

![Figure 1.10](image-url) Distribution of ages of mantle rocks in South Africa determined by Re–Os isotope measurements. The numbers are the ages (in Ga) determined by Re–Os isotope systematics. In most regions, Re–Os ages in the mantle rocks agree roughly with the ages of crust and show small depth variation. Although there are some depth variations in the age, there is no clear trend to show younger ages at or near the MLD. (After Pearson [1999]. Reproduced with permission of Elsevier.)
1.4. MODELS FOR THE LAB AND THE MLD

1.4.1. Partial Melting

Partial melting is an obvious mechanism to reduce seismic wave velocities [e.g., Mizutani and Kanamori, 1964; Spetzler and Anderson, 1968; Anderson and Spetzler, 1970; Lambert and Wyllie, 1970], but the following conditions need to be met for this model to explain the low velocities observed: (a) the physical conditions must be consistent with the presence of melt; (b) the melt fraction must be sufficient. The critical melt fraction to explain the observed velocity reduction depends greatly on the geometry of melt in a rock [Stockert and Gordon, 1975; Takei, 2002]. For a typical melt geometry in the shallow upper mantle corresponding to the dihedral angle of 20–50° [e.g., Toramaru and Fujii, 1986; Yoshino et al., 2009], 2–3% (3–5%) of melt is needed to explain ~2–5% (5–10%) of velocity reduction (Figure 1.11). (Holtzman [2016] recently proposed that even a very small amount (5–10%) of velocity reduction (10–20% shear modulus reduction), (~20–40° for asthenosphere materials). In order to explain 5–10% velocity reduction (10–20% shear modulus reduction), a melt fraction of 3–5% is needed (\(v_s = \sqrt{\frac{\mu}{\rho}} \), \(\rho\) is density).

(Modified from Takei [2002]. Reproduced with the permission of Wiley.)

For a partial-melt model to explain a large velocity drop, then some mechanisms are needed to accumulate melt. One possibility is melt accumulation at the LAB [e.g., Hirschmann, 2010] (Figure 1.12a), and another is layering of melt-rich regions by deformation [e.g., Kawakatsu et al., 2009] (Figure 1.12b); the plausibility of these mechanisms will be discussed in section 1.6.

1.4.2. Chemical/Mineralogical Layering

Layering in the major element composition and/or mineralogy is an obvious mechanism to explain a sharp velocity change. For this mechanism to be a good model for the LAB or the MLD, there must be a large compositional contrast at that boundary. Existing experimental studies, however, show that seismic-wave velocities are only weakly dependent on major-element chemistry and mineralogy as far as typical ranges following partial melting are assumed (Figure 1.9) [e.g., Lee, 2003; Matsukage et al., 2005; Schutt and Lesher, 2006].

There is scant major-element change in composition expected for the LAB, and therefore this mechanism is unlikely to be a good model. For the MLD, a large contrast in chemistry and/or mineralogy generated by the complex processes of continent formation could be invoked. Behn and Kellem [2006] discussed amphibole enrichment within igneous rocks of deeply exposed arc terranes, which are one of the building blocks of continents. Rader et al. [2015] proposed that metasomatism has occurred globally in the deep upper mantle (~200 km or so), producing volatile-rich melt. The melt is assumed to be buoyant and to migrate upward until it reaches its solidus. The melt would solidify as mantle cools to produce volatile and FeO-rich rocks at the observed depth of the MLD (Figure 1.13).

For this model to explain the MLD, three conditions must be met: (a) the velocity drop observed at the MLD (2–6%) must be consistent with the compositional variation; (b) the depth of compositional variation must agree with the depth of the MLD; (c) the compositional
variation occurs globally at similar depths. Corollary predictions of this model must be consistent with the relevant observations, which include an increase in electrical conductivity across the MLD and geochemical traces of this wet mineralogy in mantle xenoliths. We will review these issues in section 1.6 in detail.

### 1.4.3. Layering in Anisotropy

If there were a sharp change in anisotropy within the upper mantle with scarce compositional changes, it would resemble a sharp discontinuity in isotropic velocity with respect to wave refraction and conversion, but properties would vary with the direction of wave propagation (Figure 1.14). Evidence for a change in anisotropy has been reported by, for example, **Yuan and Romanowicz** [2010b] and **Sodoudi et al.** [2013].

When anisotropy is caused by lattice-preferred orientation (LPO) of minerals, layering in anisotropy may occur either by a change in the flow geometry or in the deformation fabrics, or both. Anisotropy may also be due to shape-preferred orientation (SPO) (of a melt-rich materials), as suggested by **Holtzman et al.** [2003] and **Kawakatsu et al.** [2009]. **Yuan and Romanowicz** [2010b] suggested a change in azimuthal anisotropy with depth corresponding to the change in flow geometry, and proposed that this is a mechanism for a MLD. **Kawakatsu et al.** [2009] suggested that layering in SPO is a cause of the LAB. The validity of such models for the LAB and the MLD is discussed in section 1.6.

### 1.4.4. Temperature Effects

Low seismic velocities in the asthenosphere have been explained classically by the high homologous temperature [**Birch**, 1952; **Schubert et al.**, 1976]. Physical dispersion from anelastic effects amplifies this [**Gueguen and Mercier**, 1973; **Faul and Jackson**, 2005]. These models, however, fail to explain the sharp velocity drop at the LAB (and the MLD) demonstrated by studies using short wavelength body waves. Temperature increases with

---

**Figure 1.12** Possible mechanisms of melt accumulation that might explain the large velocity reduction in the asthenosphere: (a) melt accumulation at the LAB; (b) melt-rich layers in the asthenosphere. (After **Hirschmann** [2010]. Reproduced with permission of Elsevier.)
depth smoothly, and therefore velocity drop caused by temperature is gradual and could not by itself explain the sharp velocity drop shown by RF studies (Figure 1.3).

1.4.5. Temperature and Water Effects

In order to explain a sharp velocity drop, a sharp change in some factor that has a large effect on seismic-wave velocity would need to be invoked. Knowing that the major element chemistry of the upper mantle does not change seismic velocities much [e.g., Lee, 2003; Matsukage et al., 2005], a possibility is layering in water content. This layering likely arises from the large contrast in water (hydrogen) solubility between minerals and melt, and to the slow diffusion of hydrogen. A large contrast in water solubility causes depletion of water in the solid upon partial melting [e.g., Karato, 1986; Hirth and Kohlstedt, 1996]. At mid-ocean ridges the rate of melting increases rapidly at ~65 km depth, so there will be a sharp contrast in water content in a rock at around this depth (water-poor above this depth, and water-rich below). Hydrogen diffusion is slow (diffusion distance is ~1–10 km for ~100 Myr) and therefore this layered structure will be maintained for most of the lifetime of an oceanic plate.

Using a close link between plastic deformation and anelastic relaxation [e.g., Karato and Spetzler, 1990], Karato and Jung [1998] suggested that a sharp change in the water content might explain a sharp velocity drop at the LAB. If restricted to the absorption band model, however, the magnitude of a velocity drop by this mechanism is limited, as seen from equation (1.1) [e.g., Minster and Anderson, 1981; Karato, 1993]. If $Q$ is 80 [Dziewonski and Anderson, 1981] in the asthenosphere, the velocity drop is ~1% or less (Figure 1.3), which is far smaller than the observed velocity drop (5–10%).

Figure 1.13 A frozen-melt model of the MLD. In this model, it is assumed that metasomatism occurred in the deep continental upper mantle at the global scale. The hydrous melts thus formed migrated upward until they reached the place where temperature is below the solidus (~1000°C). Minerals with low seismic-wave velocities (e.g., amphibole, phlogopite, FeO-rich pyroxenes) are formed to cause the velocity drop at the MLD. If this model is correct, then the age of rocks (frozen melt) near the MLD should be younger than the ages of rocks below and above, and rocks from/near the MLD should contain a large amount of low-velocity minerals such as amphibole, phlogopite, or FeO-rich pyroxenes at the global scale. (After Rader et al. [2015]. Reproduced with permission of Elsevier.)
1.5. ELASTICALLY ACCOMMODATED GRAIN-Boundary Sliding Model

The above difficulties led Karato [2010] to invoke “elastically accommodated grain-boundary sliding” (EAGBS) as an alternative model to explain a sharp and large velocity drop at the oceanic LAB. When EAGBS operates, then the close link between attenuation and velocity reduction predicted by the absorption band model (equation 1.1) is no longer valid, and the magnitude of velocity reduction can exceed $V_VV_Q^1$.

EAGBS is a well-established physical process that occurs during the deformation of a polycrystalline material. The basic physics were established in the 1940s, and some modifications have been made on theory in the subsequent years. Experimental studies have also been made in metals, oxides, and silicates. In this section, we will provide a brief review of this model. One major advantage of this model is that it predicts a relatively large velocity drop (a few to 10%) at a modest temperature. Therefore if this process occurs sharply enough, it could explain the key features of LAB (and MLD), namely a large and sharp velocity reduction at a modest temperature on the global scale. In the following, we review the basics of EAGBS.

1.5.1. Deformation of a Polycrystalline Material: the Role of Grain-Boundary Sliding

When a deviatoric stress is applied to a polycrystalline material, each grain will be deformed corresponding to the local stress. When chemical bonding among atoms is strong everywhere, then the magnitude of all atomic displacement is small, and deformation is instantaneous and recoverable, i.e., elastic. In such a case, the elastic constant of a polycrystalline material is some average of the elastic constants of individual mineral grains.

At high temperatures and low frequencies, however, grain boundaries behave like a viscous fluid and fail to sustain shear stresses. Consequently, large stress concentration will occur at the vertices of crystalline grains, leading to excess strain and, effectively, the reduction of elastic constants. Since grain-boundary sliding dissipates energy, seismic waves attenuate when this process occurs.

The processes of grain-boundary sliding are schematically illustrated in Figure 1.15. The key point is that grain-boundary sliding is accommodated in different manners at different stages of deformation. In the initial small strain deformation, accommodation is elastic, leading to high local stress. This EAGBS leads to a peak in attenuation and a reduction in elastic constant (seismic-wave velocity). The local high stress caused by elastic accommodation is gradually relaxed by diffusional mass transport, leading to a distributed relation time and weakly frequency-dependent attenuation (the absorption band behavior). Finally, stress distribution and diffusional accommodation will produce a balance, leading to steady-state diffusion creep. Changes in elastic-wave velocity (elastic constant) and attenuation associated with these processes are shown schematically in Figure 1.16 as a function of frequency of elastic waves.

As illustrated in Figure 1.15, the degree to which an elastic constant is reduced by grain-boundary sliding is determined by the degree of stress concentration at the mineral scale, which in turn is determined by the grain-boundary shape, that is, the orientation of each grain-boundary with respect to the applied stress. Consequently, the magnitude of reduction in elastic constant (seismic-wave velocity) due to EAGSB is independent of grain-size.
Figure 1.15 Schematic diagrams showing the processes of accommodation associated with grain-boundary sliding: $\sigma_n$, normal stress (positive = tension); $\tau_a$, applied shear stress. (a) Grain-boundary shape. (b) Stress distribution for elastic accommodation. When grain-boundary sliding is accommodated by elastic deformation, stress concentration occurs at grain-corners. Sliding eventually stops when stress concentration balances applied stress. (c) Stress distribution for diffusional accommodation. High stress is relaxed by diffusional mass transport. Steady-state stress distribution is shown that corresponds to steady-state diffusional creep. Gradual transition from elastic to diffusional accommodation leads to transient diffusional creep associated with weakly frequency-dependent anelasticity absorption-band behavior. (After [Raj and Ashby, 1971]. Reproduced with the permission of Springer.)

Figure 1.16 A schematic diagram showing the frequency dependence of seismic-wave velocity ($V$) and attenuation ($Q^{-1}$). (From Karato [2012]. Reproduced with permission of Elsevier.). (See insert for color representation of the figure.)
This can be seen in the following equations for the change in shear modulus by grain-boundary sliding [Zener, 1941; Raj and Ashby, 1971; Ghahremani, 1980],

\[
\frac{\mu_{\text{relaxed}}}{\mu_\infty} = \left( \frac{V_{\text{relaxed}}}{V_\infty} \right)^2 = \frac{2 \gamma + 5\nu}{5 \gamma - 4\nu} \quad \text{[Zener, 1941]} \quad (1.2a)
\]

\[
\frac{\mu_{\text{relaxed}}}{\mu_\infty} = \left( \frac{V_{\text{relaxed}}}{V_\infty} \right)^2 = \frac{1}{0.57(1-\nu)+1} \quad \text{[Raj and Ashby, 1971]} \quad (1.2b)
\]

\[
\frac{\mu_{\text{relaxed}}}{\mu_\infty} = \left( \frac{V_{\text{relaxed}}}{V_\infty} \right)^2 = \frac{0.86-0.83\nu}{1-\nu} \quad \text{[Ghahremani, 1980]} \quad (1.2c)
\]

where \(\mu_{\text{relaxed}}\) is the shear modulus of an aggregate relaxed by grain-boundary sliding, \(\mu_\infty\) is the shear modulus corresponding to pure elastic deformation, and \(\nu\) is Poisson’s ratio. Note that these equations do not contain grain-size.

The transition from unrelaxed (high frequency) to relaxed (low frequency) states is associated with a peak in attenuation of seismic waves (Figure 1.16). The frequency at which this peak occurs is dependent on grain-size as [e.g., Nowick and Berry, 1972],

\[
\omega_{\text{EAGBS}} = \frac{\delta \cdot \mu_\infty}{d \cdot \eta_{\text{GB}}(T,P;C_W)} \quad (1.3)
\]

where \(\delta\) is grain-boundary thickness, \(d\) is grain-size, and \(\eta_{\text{GB}}\) is grain-boundary viscosity. The presence of a peak frequency for attenuation is a unique feature of velocity reduction by EAGBS and therefore a potential fingerprint of the process.

1.5.2. Experimental Observations on Anelasticity Including EAGBS

1.5.2.1. EAGBS

EAGBS was first observed for Al (metal) [Ké, 1947], where both energy dissipation and modulus reduction were measured. Pezzotti [1999] measured energy dissipation in Al\(_2\)O\(_3\) and MgO but not the modulus reduction. (Modulus reduction can be calculated from energy dissipation using the Kramers–Kronig relationship to show that the modulus reduction in these materials is on the order of a few percent.) Two papers were published on upper mantle materials: Jackson and Faul [2010] (see also Jackson et al., 2014) on olivine and Sundberg and Cooper [2010] on an olivine+orthopyroxene mixture. Both energy dissipation and modulus reduction were measured in these studies. For olivine-rich samples with grain-size 5–15µm, a peak in attenuation was observed at around 900–1000°C, associated with a substantial modulus reduction (Figure 1.17). Jackson and Faul [2010] reported ~7% reduction in shear modulus, while Sundberg and Cooper [2010] reported ~30% reduction.

EAGBS peaks are observed as a small addition to “high-temperature background” (absorption-band behavior), so the detailed nature of the anelasticity is difficult to parameterize simply. The reasons for differing modulus reduction between different laboratory studies are unknown. In addition, although theoretical models suggest that the magnitude of velocity reduction is independent of grain size, some dependence on this parameter was seen in experiments.

1.5.2.2. Influence of Water

A strong influence of water content on the characteristic frequency is suggested by experimental observations of a close link between anelasticity and creep, although direct tests of the influence of water on anelasticity have not been conclusive. A study on a natural sample by Aizawa et al. [2008] observed water to strongly enhance anelasticity, while a recent study on synthetic olivine aggregates in the Ti-dominated regime showed small effects [Cline et al., 2018].

Both theoretical consideration and experimental results suggest a close link between microcreep (transient creep) and anelastic relaxation [Karato and Spetzler, 1990; Webb et al., 1999; Tan et al., 2001; Webb and Jackson, 2003; Faul and Jackson, 2015]. Also, theoretical and experimental studies showed a close connection between transient and steady state creep [e.g., Amin et al., 1970]. Indeed McCarthy et al. [2011] suggested a direct link between steady state creep and anelasticity for an organic material (borneol). Experimental studies showed a substantial effect of water (hydrogen) to enhance steady state creep in olivine [e.g., Karato et al., 1986; Mei and Kohlstedt, 2000a,b; Karato and Jung, 2003]. Consequently, it is natural to expect that water enhances anelastic relaxation in olivine [Karato, 1995, 2003].

There are a few studies on a natural dunite (olivine-rich rock) that support this notion [e.g., Jackson et al., 1992; Aizawa et al., 2008]. Interpretation of these earlier studies is complicated, however, because a natural rock was used and so the chemical environment was not fully controlled. In contrast, Cline et al. [2018] reported results of an experimental study on the influence of water on anelasticity where synthetic olivine polycrystals were used. In
In this study they dissolved hydrogen in their samples in combination with Ti, and explored a range of hydrogen (and Ti) content as well as oxygen fugacity. They found that the influence of hydrogen content is relatively small compared to the influence of oxygen fugacity. This is a surprising result because the influence of water (hydrogen) content (water fugacity) on steady state creep of olivine is known to be substantially larger than that of oxygen fugacity [e.g., Bai et al., 1991; Karato and Jung, 2003].

The reason for this surprising result is not well understood. One possibility is that the hydrogen-related defect structure and mobility in the Ti-dominated regime are different from the intrinsic regime where the concentration of hydrogen-related defects is independent of the...
concentration of impurities (such as Ti). The Ti content in the mantle olivine is small (~20–30 ppm wt (TiO₂) = ~30–50 ppm Ti/Si [Pearson et al., 2003]), much less than the inferred water content in the asthenosphere (~100 ppm wt (H₂O) = ~1500 ppm H/Si [Hirschmann, 2006]). Consequently, hydrogen-related defects in Earth likely occur in a regime that differs from that studied by Cline et al. [2018]. We expect that the influence of hydrogen on physical properties such as anelasticity is different among different regimes. A study to assess the differing roles of hydrogen is needed to evaluate the relevance of the results by Cline et al. [2018] to Earth’s upper mantle. In addition, the study by Cline et al. [2018] is on the absorption band regime and the influence of water on EAGBS has not been studied.

1.6. DISCUSSION

1.6.1. Partial Melt Model Versus Subsolidus Models for the LAB

Difficulties of the partial melt model for the MLD are obvious because of the low temperature involved, and we will not add any discussions here. The possible role of partial melting as the cause for the LAB, however, does need further scrutiny. We shall discuss three issues: (a) velocity reduction and the melt distribution in the asthenosphere, (b) electrical conductivity, and (c) seismic-wave attenuation in the asthenosphere. The two properties (b) and (c) are sensitive to partial melting, temperature, and water content.

1.6.1.1. Seismic Wave Velocity and Melt Distribution

In section 1.4.1 (Figure 1.11) we discussed that in order to reduce seismic-wave velocity by a few percent, a substantial amount of melt (2–5%) is needed. Holtzman [2016], however, discussed that even a small melt fraction (~0.1%) could reduce seismic wave velocity by a few percent, an idea based on a theoretical model by Takei and Holtzman [2009] which suggests that even a small amount of melt might enhance diffusion creep modestly. Holtzman [2016] used this model and the link between long-term creep and anelasticity to claim that a substantial velocity reduction will occur even with a small amount of melt.

The stress distribution calculated by Takei and Holtzman [2009] has singularities suggesting that the stress distribution is not steady state and enhanced diffusion creep likely corresponds to transient diffusion creep (see Figure 1.15) [Raj and Ashby, 1971]. Such a stress singularity has small effects on dislocation creep, but the presence of substantial seismic anisotropy in the asthenosphere suggests that the dominant mechanism of deformation is dislocation creep. Consequently, the applicability of the results by Takei and Holtzman [2009] to anelasticity and velocity reduction in the asthenosphere is questionable.

How about melt accumulation near the LAB (Figure 1.12a)? Theory of melt migration under gravity suggests that compaction is efficient, being characterized by a compaction length of ~0.1 km to ~1 km for typical asthenosphere conditions [e.g., Ribe, 1985], too thin to have any effects on geophysically observable properties. Indeed magnetotelluric observations in the NoMelt region (~70 Ma Pacific Ocean) show no evidence of a peak in electrical conductivity at the LAB [Sarafian et al., 2015] (see Chapter 2). An exception is possible near trenches. Due to the stress field associated with plate bending, regional tension could be generated near the bottom of the oceanic plate. This would suck melt to cause locally melt-rich regions that Yamamoto et al. [2014] proposed might explain the presence of “petit spots” near ocean trenches. The global presence of accumulated melt at the LAB, however, is unlikely.

How about a layered structure (Figure 1.12b)? Kawakatsu et al. [2009] proposed that a sharp and large velocity drop at the LAB is caused by the presence of near horizontal melt-rich layers in the asthenosphere. They argued that even with a net melt fraction of ~0.1%, as inferred from petrology [e.g., Hirschmann, 2010], if there were melt-rich bands as suggested by the experimental results by Holtzman et al. [2003], then there would be large radial seismic anisotropy (V₉₀ > V₅₈) in the asthenosphere but not in the lithosphere. The global presence of accumulated melt at the LAB, however, is unlikely.

The layered structure model, however, has some notable difficulties. Horizontal shear deformation is expected in most regions of the asthenosphere, but a melt-rich layer will not remain horizontal according to laboratory studies [Holtzman et al., 2003]. In such a case, a large amount of melt cannot be kept in these layers because gravity will separate melt from the solid rock. In addition, if this model were to explain the observed velocity drop of ~5%, the melt fraction in the melt-rich bands must have a very specific value, 14.3% (Figure 1.18). Deviation of the melt fraction in the melt-rich bands from this value of more than 1% will lead to velocity perturbations that do not agree with seismological observations [Karato, 2014, 2018].

The Kawakatsu et al. [2009] model also implies that the magnitude of the velocity drop at the LAB is the same as the magnitude of radial anisotropy. The reported value range of velocity drop at the LAB is 5–10%, whereas the magnitude of radial anisotropy in the oceanic asthenosphere is usually ~2–5% [e.g., Montagner and Tanimoto, 1990, 1991]. For these reasons, we consider that the observed properties of the LAB are difficult to explain using a layered structure model.
in the old oceanic asthenosphere is the melt fraction if melt were homogeneously distributed—
in the old oceanic asthenosphere $\phi_0 = 0.1\%$ [Hirschmann, 2010]. In order to explain the observed velocity reduction of 5–10% (in the old oceanic asthenosphere [e.g., Kawakatsu et al., 2009; Rychert et al., 2005]) with $\phi_0 \approx 0.1\%$, the melt fraction in melt-rich layers ($\phi$) should have a very specific value (14.3 ±0.5, −1%) that is highly unlikely. (From Karato [2018].)

1.6.1.2. Electrical Conductivity

Some MT studies on the oceanic regions report a peak in electrical conductivity in the shallow asthenosphere [Evans et al., 2005; Naif et al., 2013]. These studies were close to ocean ridges but Naif et al.’s [2013] study was also close to a trench. A melt-rich layer at the top of the asthenosphere could explain the observed peak in conductivity. Another study away from the ridge, however, shows no peak in conductivity [Sarafian et al., 2015], which discourages a melt-accumulation model for the LAB at least for regions far from ocean ridges (see also Chapter 2). This is a notable point since a very large and sharp LAB is observed in old oceanic upper mantle [Rychert et al., 2005; Kawakatsu et al., 2009]. These two observations suggest that a large and sharp velocity drop at the LAB should be attributed to a process other than partial melting. An obvious alternative explanation is to invoke hydrogen-enhanced conductivity [Karato, 1990; Wang et al., 2006; Dai and Karato, 2009]. We conclude that accumulated melt could occur in limited regions near ridges but not globally.

1.6.1.3. Seismic Wave Attenuation

Both the partial-melt model and the EAGBS model have clear predictions for seismic-wave attenuation: there would be a peak in seismic wave attenuation if there is sufficient melt [Jackson et al., 2004]; if the EAGBS mechanism is the cause of the LAB, then similarly an attenuation peak near the LAB would be expected (see Figure 1.16). There have been a few observations on the oceanic LAB to suggest the presence of a peak in attenuation [e.g., Takeuchi et al., 2017]. This behavior is different from the known frequency dependence of attenuation in most minerals [e.g., Jackson, 2009] as well as the results of previous seismological studies [e.g., Anderson and Minster, 1979; Shito et al., 2004].

Takeuchi et al. [2017] proposed two explanations of this anomalous frequency dependence. One is the presence of partial melt, and another is the consequence of EAGBS. Jackson et al. [2004] showed that the presence of melt leads to a peak in attenuation. EAGBS also predicts a peak in attenuation (Figure 1.16). How can we distinguish these two?

One possible way to distinguish partial melt model from the EAGBS model is to look at the bulk attenuation, namely attenuation associated with volumetric strain. Grain-boundary sliding involves viscous motion within thin grain boundaries, mostly in response to shear stress. In contrast, partially molten materials cause substantial bulk attenuation as viscous melts migrate within the solid matrix according to some models [e.g., Budianski and O’Connell, 1980; Marko, 1980] (see also a recent experimental study by Cline and Jackson [2016], who reported some difficulties in the experimental study). In this connection, bulk attenuation measurements would provide useful constraints.

1.6.2. The Frozen-Melt Model for the MLD

Seismic wave velocities and attenuation are little affected by compositional changes at the LAB, except for changes in water (hydrogen) content. A larger degree of compositional change, however, may occur in the continents that could be a cause for the MLD. Among various models for the MLD, a frozen-melt model proposed by Rader et al. [2015] could be an important possibility (Figure 1.13).

Rader et al. [2015] proposed that, after the formation of continents, metasomatism in the deep continental upper mantle occurred globally. The volatile-rich melts formed by such metasomatism migrated upward to the depth where the temperature lies below the solidus (~1000°C). At that depth, the melt froze to form minerals enriched with volatiles and FeO that have low seismic-wave velocities.

This model predicts the following:
1. rocks that form near the MLD typically contain a large amount of low-velocity minerals such as amphibole, phlogopite, or FeO-rich pyroxenes at the global scale;
2. the ages of rocks near the MLD are younger than the ages of rocks below and above the MLD;

![Figure 1.18](image-url) Relations between the melt fraction in a melt-rich layer and the velocity reduction of the asthenosphere containing a melt-rich layer corresponding to the Kawakatsu model: $\phi_0$ is the melt fraction if melt were homogeneously distributed—in the old oceanic asthenosphere $\phi_0 = 0.1\%$ [Hirschmann, 2010]. In order to explain the observed velocity reduction of 5–10% (in the old oceanic asthenosphere [e.g., Kawakatsu et al., 2009; Rychert et al., 2005]) with $\phi_0 \approx 0.1\%$, the melt fraction in melt-rich layers ($\phi$) should have a very specific value (14.3 ±0.5, −1%) that is highly unlikely. (From Karato [2018].)
3. electrical conductivity increases substantially at the MLD (from above to below), owing to increased water (+FeO) content.

The amount of low-velocity materials needed to explain the MLD is shown in Figure 1.19. Among various hydrated minerals, the most commonly observed in the continental mantle is amphibole. If 20–30% of amphibole is present, this could explain the observed velocity reduction at the MLD. Xenoliths show no strong indication for the global presence of amphibole-rich peridotites at the MLD depths [e.g., Pearson et al., 2003], but geological observations are not perfect and it is difficult to rule this model out for this reason alone.

Regarding the global occurrence of metasomatism, clear support would be the depth variation of the age of mantle rocks: if there were metasomatism then the age should have been reset and the regions that have undergone metasomatism should have younger ages than the surrounding regions. According to the available data, however, there is no indication of global resetting of the age at the MLD depth (Figure 1.10) [e.g., Pearson, 1999; Carlson et al., 2005].

Another consequence of this model is that there will be an increase in electrical conductivity with depth at the MLD. Electrical conductivity of most materials from the volatile-rich melt is substantially higher than that of olivine. For example, the electrical conductivity of amphibole exceeds that of olivine by a factor of $10^2$ to $10^4$ [Wang et al., 2012]. In this scenario, melt had migrated through the mantle, and so melt must have been connected (dihedral angle less than 60°). Consequently, these high-conductivity minerals are likely connected and enhance the bulk electrical conductivity [e.g., McLachlin, 1987]. Currently available MT models do not show a clear increase in electrical conductivity near the MLD depth (e.g., Meqbel et al. [2014]; see also Chapter 5).

None of these points is definitive, but taken together the above factors do not favor the frozen-melt model. In the model’s defense, observations on mantle samples (mineralogy, major element chemistry, Re–Os dating) are limited. Similarly, details on the depth variation in electrical conductivity have not been resolved in all available MT studies. Therefore it is important to make advances in these two areas (studies on mantle samples and MT) to test the frozen-melt model for the MLD.

### 1.6.3. Layered Anisotropy Model for the MLD and the LAB

The model for the LAB proposed by Kawakatsu et al. [2009] invokes a layered (radial) anisotropy caused by partial melting. Difficulties with this model were discussed
in section 1.6.1. A fabric transition between the oceanic lithosphere (A-type fabric) and the asthenosphere (E-type fabric) has been suggested [Karato, 2008b; Karato et al., 2008], but a change in anisotropy between A- and E-type fabrics is subtle and not enough to explain a large velocity drop at the LAB. Furthermore, there is no evidence for a fabric transition in the continental lithosphere down to ~200 km from the study of mantle xenoliths [Ben Ismail and Mainprice, 1998].

Layered anisotropy was also invoked to explain the MLD in the continents [e.g., Yuan and Romanowicz, 2010b]. In case of the MLD, layered anisotropy is most likely due to the layering in the geometry of LPO of minerals in rocks [e.g., Karato et al., 2008]. Upper mantle xenoliths, however, show no evidence of systematic change in LPO with depth [e.g., Ben Ismail and Mainprice, 1998], and therefore if there is a layering in anisotropy it would be caused by a change in flow geometry. Since the MLD is nearly horizontal, a change in flow geometry would result from a change in azimuthal anisotropy not in radial anisotropy across the MLD. As pointed out by Karato et al. [2015], however, azimuthal anisotropy can simulate a sharp isotropic velocity inversion only if most seismic observations illuminate a small portion of available back azimuths. Layered anisotropy could explain the upper mantle discontinuity in regions with such data restrictions, but it cannot be the cause of these discontinuities at a global scale (see Figure 1.14).

1.6.4. EAGBS Model for the MLD and the LAB

The following features are predicted by the EAGBS model:

1. velocity reduction in olivine (or olivine-rich rocks) occurs where temperature reaches a critical value, ~1000 °C, that depends on the water content (and pressure) and therefore the velocity reduction by this mechanism is global;
2. velocity reduction is large (several percent);
3. velocity reduction occurs without compositional layering, and the ages of rocks do not change across the MLD in the continents if EAGBS is the cause of the MLD;
4. velocity reduction is not associated with a sudden increase in electrical conductivity if EAGBS occurs solely due to elevated temperature (there will be a conductivity jump if EAGBS is associated with an increase in water content);
5. there should be a peak in attenuation in the depth range where velocity reduction occurs.

All these features are consistent with the known geological and geophysical observations summarized previously. In contrast, alternative models, such as the frozen-melt model, are not consistent with many of the geological or geophysical observations previously summarized.

There is another observation from the cratonic mantle that provides a useful constraint on models for the velocity reduction at the MLD and the LAB. This is the much larger velocity drop at the MLD (~100 km) than at the LAB (~200 km) (Figure 1.6b) [Abt et al., 2010]. This is a notable observation because for some scenarios, such as partial melt or monotonic temperature increase models, a larger velocity drop would be expected in the deep regions where temperature is high.

In contrast, the EAGBS model provides a natural explanation for this observation for the following reasons. Figure 1.16 depicts, from right to left (for a given seismic wave frequency), transition from a shallow region to a deeper region where temperature is higher. In the shallow part, seismic frequency is higher than \( Q_{EAGBS} \) and rocks are in the unrelaxed (high velocity) state; in the deeper part where \( Q_{EAGBS} \) increases the seismic frequency becomes lower than \( Q_{EAGBS} \). At that point, a substantial (a few percent) velocity drop occurs. With increasing depth in the absorption band regime, where velocity change and \( Q \) are directly linked by equation (1.1) and therefore the magnitude of velocity change is \( \sim Q^{-1} \), and when \( Q \sim 100 \) near the LAB, the velocity drop is \( \sim 1 \% \) not a few percent.

Regarding attenuation, Takeuchi et al. [2017] provided evidence for the attenuation peak in the asthenosphere that is consistent with the EAGBS model (and also consistent with a partial melt model) for the LAB. As to the MLD, there is no direct evidence for the attenuation peak associated with the MLD. The reported rather low intrinsic \( Q_s \) in the continental lithosphere (\( Q_s = 100–300 \), Figure 1.8) could, however, well be caused by an attenuation peak near the MLD for the following reasons. If a conventional absorption-band model (without an attenuation peak) is assumed, \( Q_s \) within the continental lithosphere can be calculated from the known \( Q_s \) in the asthenosphere (\( Q_s \sim 80 \)) and the temperatures in the lithosphere. Because temperatures in the continental lithosphere are low, substantially higher \( Q_s \) (~1000 or higher) values would be expected than those reported (\( Q_s = 100–300 \); Figure 1.8). Invoking an attenuation peak (\( Q_s = 20 \)) around the MLD (\( Q_s = 20 \) corresponds to a velocity reduction of 5%) with 20 km width may solve this apparent paradox. Such an attenuation peak will have the same degree of attenuation as a constant \( Q_s = 200 \) distributed in a layer of 200 km thickness.

Another explanation is to attribute the apparent low \( Q_s \) to scattering [e.g., Fehler et al., 1992; Jin et al., 1994]. Dalton et al.’s [2009] attenuation model is for intrinsic \( Q_s \), however, after corrections for scattering and focusing. In addition, the rough agreement between a long-wavelength study [Dalton et al., 2009] and a short-wavelength study [Mitchell, 1995] suggests that scattering is not a dominant cause for low \( Q_s \).
Regarding the MT observations, there is no clear evidence for an increase in conductivity at \( \sim 100 \text{ km} \) depth in the cratonic upper mantle, but conductivity becomes high below \( \sim 200 \text{ km} \) [e.g., Merkel et al., 2014]. Similarly, an increase in conductivity is inferred across the oceanic LAB [e.g., Evans et al., 2005; Naif et al., 2013; Merkel et al., 2014]. A simple explanation is that the large increase in conductivity across the LAB is due to a change in water content (higher below the LAB [Karato, 1990; Wang et al., 2006; Dai and Karato, 2009]). The large and sharp velocity reduction for most of the oceanic LAB may also be due to the enhanced EAGBS caused by an increase in water content [Karato, 2012; Olugboji et al., 2013].

In summary, the EAGBS model provides a unified model for the MLD and the LAB in the sense that (a) it explains the global occurrence of these boundaries, (b) the temperature requirement for the MLD and the oceanic LAB is consistent with EAGBS, and (c) it does not require chemical or anisotropy layering. A major limitation of this model, however, is that experimental studies on EAGBS are still preliminary; key features such as the grain-size sensitivity of velocity reduction and the influence of water on EAGBS are poorly constrained.

### 1.7. SUMMARY AND FUTURE DIRECTIONS

The causes of both the LAB and the MLD are reviewed in the light of geological, geophysical, and mineral (and rock) physics observations. Among several models for these discontinuities, three are potentially important: compositional layering (in major element and/or mineralogy), partial melting, and EAGBS.

Geological observations relevant to these models were reviewed, and we conclude that compositional layering model is difficult to reconcile with the observations of mantle rocks including mineralogical layering and rock-age layering.

Our preferred model for the oceanic LAB is that the large and sharp velocity drop is caused mainly by thermally induced EAGBS helped by the jump in water content (Figure 1.20). There should be a peak in attenuation at the LAB associated with EAGBS. For the continental upper mantle, the MLD is likely due to thermally induced EAGBS, but it may also be influenced by chemical layering (layering in water content). The MLD should be associated with a peak in seismic attenuation if it is mainly caused by EAGBS. If EAGBS occurs only by temperature increase, there should be no large increase in electrical conductivity. In the continental (cratonic) LAB,

![Figure 1.20](image_url) Preferred models for the MLD and the LAB consistent with available geological and geophysical observations. (Modified from Karato [2012]. Reproduced with the permission of Elsevier.)
rocks are already in the relaxed state with respect to EAGBS, so that further increase in temperature and/or water content leads only to a minor decrease in velocity; an increase in water content at the LAB, however, will increase electrical conductivity substantially.

Our model for the MLD and the LAB has some implications for change in rheological properties across these boundaries. Inasmuch as the MLD is mainly caused by temperature increase, change in long-term rheological properties across the MLD will be negligible because EAGBS changes only the short-term, small strain rheological properties but not the long-term steady-state rheological properties (in this case, a large seismic wave velocity drop is not associated with a large drop in rock strength). In contrast, the LAB in the oceanic regions and presumably that in the continent is likely associated with a change in water content. Consequently, the LAB, including the LAB in the cratonic mantle where velocity reduction is small, is associated with a major reduction in the long-term rheological properties (creep strength) with depth. The link between seismic-wave velocities and long-term rheological properties
is not direct, particularly when EAGBS plays an important role in controlling the seismic-wave velocities.

Our preferred model is developed based on experimental and theoretical studies on elastic and anelastic properties of rocks in combination with constraints from geological (+ geochemical) and geophysical observations. Geological observations are based on a limited number of rock samples, however, so that our understanding is far from complete. Further studies are warranted, including the depth variation in mineralogy in mantle xenoliths and more precise determination of the depth dependence of the age of mantle xenoliths. Similarly, geophysical observations are also limited.

In the following, we summarize some of the possible future directions on the studies on the origin of the LAB and the MLD.

- Geophysics (see Figure 1.21)
  - Conduct MT studies on the continental upper mantle. If the frozen melt is the cause of the MLD, there should be a substantial increase in conductivity at the MLD. Thermally induced EAGBS will cause no increase in conductivity at the MLD.
  - Determine the sharpness of the velocity drop at the MLD. For the thermally induced EAGBS model, the MLD is diffuse. For a compositional layering model, it is sharp.
  - Determine the seismic-wave attenuation in the continental lithosphere. For the EAGBS model there is a peak in attenuation.
  - Determine the magnitude of bulk attenuation. If partial melting were a cause of attenuation, bulk attenuation would be substantial. For other mechanisms, shear attenuation dominates.

- Geology (geochemistry)
  - Conduct detailed studies on the mineralogy, isotopic compositions (ages) of mantle xenoliths as a function of depth. For a frozen-melt model, there should be marked compositional anomalies at around the MLD at a global scale. For a frozen-melt model, rocks around the MLD should show younger ages than rocks above and below it.
  - Conduct laboratory studies to test the grain-size sensitivity of velocity reduction by EAGBS. If such sensitivity is confirmed, develop a model to provide a physical explanation for it.
  - Conduct laboratory studies to understand the influence of water content on anelasticity, particularly EAGBS in the regime where hydrogen is dissolved in a similar way as in Earth's upper mantle.

A study of each of these topics involves technical challenges. In addition to making advances in these specific issues, it is also important to integrate all available observations to devise acceptable geological and geophysical models. Those studies will help us understand the significance of the MLD and the LAB in the upper mantle in connection to the geological evolution of the upper mantle.

ACKNOWLEDGMENTS

We thank Huaiyu Yuan for inviting us to make this contribution and for his patience in handling our paper. This study is partly supported by the grants from NSF. We thank Chris Cline, Gary Egbert, Don Forsyth, Ian Jackson, Cin-Ty Lee, Tolulope Olugboji, Jarka Plomerová, and Akira Takeuchi for helpful discussions. The comments by Barbara Romanowicz and two anonymous reviewers were helpful in improving the paper. Thank you all.

REFERENCES


ON THE ORIGIN OF THE UPPER MANTLE SEISMIC DISCONTINUITIES 33


