Subduction origin on early Earth: A hypothesis

Vicki L. Hansen
Department of Geological Sciences, University of Minnesota–Duluth, Duluth, Minnesota 55812, USA

ABSTRACT
I propose a hypothesis for the origin of subduction on early Earth that is directly coupled with mantle upwelling and the formation of mafic crust. The hypothesis invokes spatial and temporal overlap of both endogenic and exogenic processes, broad quasi-cylindrical mantle upwelling and large bolide impact, respectively, leading to subduction and spreading, two signature processes of modern plate tectonics. The spatial and temporal intersection of these processes could have occurred variably across early Earth, and thus subduction could have begun at different locations and times globally. The hypothesis postulates that the ability of a terrestrial planet to evolve plate tectonics results from a balance between the strength of its lithosphere and the size bolide it can attract and survive; both factors, to a first order, are a function of planet size.

Keywords: subduction origin, impact, Archean, bolide, plate tectonics origin.

INTRODUCTION
How and when plate tectonics began has been enthusiastically debated for decades. The debate currently combines a quest for the oldest evidence of plate tectonic processes (e.g., Cawood et al., 2006; Condie et al., 2006; Witze, 2006; Furnes et al., 2007). Modern plate tectonic processes embrace both the formation of mafic crust and the recycling of mafic crust to the mantle through subduction. Subduction is like a virus, once begun it can easily spread, yet a mechanism for inception of the subduction process has thus far proven elusive (e.g., Hilde et al., 1977; Casey and Dewey, 1984; Cloetingh et al., 1989; Mueller and Phillips, 1991). Thus, how subduction began is directly tied to the debate about the initiation of modern plate tectonic processes. As stated by Lin (2007), there must have been a time on Earth when plate tectonics did not operate and, thus, a process by which subduction began. I propose a hypothesis for the origin of subduction, coupled with the formation of mafic crust, that calls for serendipitous spatial intersection of endogenic processes (broad quasi-cylindrical mantle upwelling) and exogenic events (large bolide impacts) on early Earth. The hypothesis addresses the formation of long (>1000 km) fundamental boundaries, marked by contrasts in composition, rheology, and density, that evolve into subduction zones.

EARLY EARTH
Early Earth differed from contemporary Earth in at least two major ways. (1) The planet was pummeled by extraterrestrial bolides, large and small; and (2) early Earth had significantly more thermal energy. The first is clear from solar system evolution models and impact craters and basins on the Moon and Mars. The second may have resulted in a stagnant-lid regime, rather than a convective-lid regime (Tackley, 2000; Korenaga, 2006), hampering rather than enabling lithosphere-scale recycling.

Although there is much debate about the environmental conditions of early Earth, it is generally agreed that higher continental geothermal gradients, a result of greater radiogenic heat production and stronger basal heat flow, existed in the Archean (Taylor and McLennan, 1986; Bickle, 1978; Hoffman and Ranalli, 1988; Davies, 1992). An elevated geothermal gradient would lead to increased crustal buoyancy, increased ductility, and lower overall crustal strength; it would limit the ability of continental crust to develop sharp gradients in thickness or strength (Choukroune et al., 1995; Rey and Houseman, 2006). The broadly ductile nature of this crust would therefore inhibit the formation of abrupt compositional and/or rheological boundaries.

Fundamental, relatively sharp, focused boundaries form the essence of modern plate tectonics. Rheological and density differences play critical roles in the nature of tectonic processes of all terrestrial planets. Subduction ultimately requires plates—large tracts of crust or lithosphere, marked by differences in density and, possibly, strength. Lower plates, those to be subducted, must have both elastic strength and high density relative to the ductile upper mantle. In contrast, low-density upper plates need not have elastic strength, they can deform in either a ductile or brittle fashion. So perhaps we might ask, how could early Earth form at least local regions of high-strength crust, or relatively sharp boundaries marked by pronounced rheological and/or density contrast? Composition, temperature, strain rate, and fluids affect, to first order, the brittle (strong) versus ductile (weak) character of silicate crust; here I consider the first three variables to propose a mechanism by which early Earth might form long (>1000 km), sharp, fundamental boundaries that separate large tracts of crust with contrasting density and/or rheology. The fourth variable, fluids, enters the discussion with an Earth-Venus comparison.

HYPOTHESIS FOR THE ORIGIN OF SUBDUCTION
A series of time-step cartoons illustrates salient features of the hypothesis, beginning with differentiated (buoyant) crust rich in radiogenic heat-producing elements (Fig. 1). Part of the crust resides above a broad quasi-cylindrical mantle upwelling where the crust is thinned thermally and/or mechanically (Time 1). Mantle-scale processes modify the crust from below, resulting in two types of crust: unmodified and thinned.

While the mantle may modify the crust from below, extraterrestrial bolides can modify the crust from above. Bolides intersected the Earth as they traveled around the young Sun. Among the many objects to strike early Earth, a large bolide impacts the surface overlapping with the trace of already thinned felsic crust (Time 2). This impact event excavates a huge region of crust (>1000 km diameter), displacing much of it upward and outward, resulting in a vast region of a third type of crust that is thin, shocked, and weak (Time 3). The excavated crust, marking the impact site, could be expected to have a sharp boundary (strain gradient) against adjacent relatively unmodified (or now thickened) crust as a result of the extremely high strain rate experienced during the impact event. Although the crust would respond in a ductile fashion to low strain rate mantle flow from below, hypervelocity impact from above would induce a brittle response.
The bolide, which penetrates the crust, also triggers a mantle response that in turn affects the surface (Time 4). Local superheating of the mantle causes massive partial melting, forming ultramafic to mafic melt that rises, forming new crust; impact could further spawn the formation of a relatively long-lived mantle thermal anomaly (Jones et al., 2005; Elkins-Tanton and Hager, 2005). The region of excavated and thinned crust might be expected to respond much like thin lithosphere above a plume head, forming a ring-shaped trough that moves outward as the plume interacts with the thinned lithosphere (Griffiths and Campbell, 1991). The thermal anomaly causes continued partial melting; melt rises to feed a nascent spreading center, which forms in the region inside the trough. As the trough moves outward, new mafic (to ultramafic) crust (a fourth type of crust) continues to form at the spreading center (Time 5). The broad mantle upwelling would influence the trend of the spreading center. Spreading could be symmetric or asymmetric depending on regional stress conditions; local buttressing by relatively immobile lithosphere might favor asymmetric spreading, rather than symmetrical spreading at modern ocean ridges. As new mafic crust is displaced outward, away from the spreading center, the region of excavated and thinned felsic crust shortens and thickens. The nature of crustal thickening would depend on the inherent rheology of the excavated and thinned crust. One might reasonably predict a ductile response due to a high geothermal gradient, although the specific behavior is not critical to the hypothesis. What is important is that the excavated and thinned crust is progressively displaced outward with formation of new mafic crust. The spreading center propagates laterally in a direction parallel to the broad mantle upwelling; the growing tract of mafic crust progressively displaces the excavated and thinned crust until the new mafic crust reaches the limit of the original, impact-softened and excavated crust (Time 6).

Once the mafic crust reaches the spatial/rheological limit of the excavated crustal region, and given the right conditions (i.e., compositional and thermal depth balance against spreading rate and relative strength of mafic and felsic crusts), the early-formed mafic crust would begin to sink (Time 7), and then subduct, beneath the buoyant crust, which would become the upper plate (Time 8). The spatial/rheological limit refers to the point at which the felsic crust (modified or not) begins to resist displacement, becoming a stress concentrator (Cloetingh et al., 1989), caused by further generation of mafic crust at the spreading center. The spreading center continues to propagate along strike akin to modern ridges, and it could ultimately penetrate the felsic crust left unmodified by the impact event, but where it was thinned above the original mantle upwelling. As the system evolves, transform faults might form as dictated by geometric and kinematic criteria along portions of the mafic-felsic crust boundary. The subduction zone could propagate along its length, as did the spreading center (Time 9). With the concurrent formation of new mafic crust and subsequent subduction of older mafic crust, plate tectonics have begun in this region of the Earth.

This hypothesis calls upon serendipitous spatial overlap of both endogenic and exogenic processes. If a large bolide fell on crust far removed from a mantle upwelling, the events proposed here might not evolve into mafic crust formation and recycling via subduction, because there would be no upwelling to act as a stress guide for crustal extension. Furthermore, because the spreading center might not be able to propagate into unmodified crust, the overall areal extent of newly formed mafic crust would be limited. That is, unmodified crust might resist spreading center propagation. The proposed process could become arrested at any stage described; in this case, subduction would not begin.

The hypothesis, which assumes a constant radius Earth, accommodates, and essentially requires, concurrent formation and recycling of mafic crust. The inception of subduction requires
formation of a sufficiently large tract of mafic crust to allow for crustal cooling, and this crust must form a strong (i.e., elastic) brittle plate of significant density difference compared to the asthenosphere, but also compared to the upper plate, such that the newly formed mafic crust will be thrust below the felsic crust. The size of the impact basin is important in that an extensive tract of mafic crust must form to allow subduction to begin. The question of whether bolides of sufficient size existed is evident by examination of Earth’s neighbors, Mars and the Moon, which preserve at least partial records of that tumultuous time. Mars, about half the diameter of Earth and 11% of its mass, preserves several impact basins with diameters >1800 km (Arygle, 1800 km; Isidis, 1900 km; Hellus, 2300 km; Utopia, 3000 km). The Moon, less than one-third Earth’s diameter and 1.2% of its mass, also has huge impact basins (Orientale, 930 km; Imbrium, 1160 km; South Pole–Aitken, 2300 km). Earth, with significantly greater mass, should have sustained similarly large impact events as these smaller neighbors, forming impact basins of at least equal size. Subduction zones associated with the Lesser Antilles and South Sandwich Islands, with diameters of ~1000 km, might be taken as a minimum size for impact basins that might evolve into spreading-subduction processes.

CONSEQUENCES AND RESERVATIONS

Given that such large impact basins occur on Mars and the Moon, one might ask why plate tectonics did not begin on these bodies. The hypothesis presented here embraces a balance between a terrestrial body’s lithospheric strength and bolide size. A bolide must form a huge impact basin and penetrate the lithosphere. If a bolide is too small, or the lithosphere is too strong, plate tectonics will not occur. To a first order, lithospheric strength for terrestrial bodies is a function of planet size, or its surface area:volume ratio, A/V (Solomon and Head, 1982). Small bodies, with high A/V, have strong lithospheres. With a lower A/V, a body’s heat budget increases, and its lithosphere is weaker. The size bolide a planet can attract and sustain is also a direct function of mass, and therefore size.

Thus, plate tectonics (subduction-spreading processes) might only form on large solidsurfaced planets—bodies with relatively weak lithospheres that can sustain the impact of very large bolides. In effect, there could be a minimum size at which a planet might develop plate tectonics. The Moon is too small. Any bolide that the Moon could sustain would not initiate subduction-spreading processes because of its thick lithosphere. Mars also has a thick lithosphere, although this has not always been the case. Early in Mars’ history it had significantly more heat and, as a result, a thinner and weaker lithosphere. At that same time the solar system contained numerous large meteoroids. A single large bolide (Wilhelms and Squyres, 1984) or multiple large bolides (McGill, 1989; Frey and Schultz, 1988) may have collided with what is currently Mars’ Northern Hemisphere, forming Mars’ crustal dichotomy. The impact event(s) could have, in turn, spawned an endogenic response to develop a spreading center forming mafic crust, which was displaced asymmetrically outward and subsequently subducted along a portion of the initial impact boundary (Sleep, 1994). Thus the formation of Mars’ crustal dichotomy might have required partnership between exogenic and endogenic events (Wise et al., 1979; McGill and Dimitriou, 1990; Zuber, 2001), which evolved into plate tectonic processes (Sleep, 1994). As Mars cooled, due mainly to its high A/V but enhanced by nascent mafic crust formation, plate tectonic processes would have halted abruptly, almost as quickly as they began, preserving a truly ancient ocean-like basin across Mars’ Northern Hemisphere (Frey, 2006).

Venus, with ~80% of Earth’s mass, lacks plate tectonics. Thus, (1) Venus might mark a lower limit for the terrestrial planet mass required for plate tectonics to evolve; or (2) Venus once had plate tectonics, but fast cooling led to its demise; or (3) Venus’s dry strong crust (Mackwell et al., 1998) rendered its lithosphere too strong for plate tectonics to develop. Scenario 1 seems unlikely given the similarity in Earth and Venus size and composition. Scenario 2 is also unlikely given the role of a dense atmosphere in maintaining a hot interior (e.g., Abe and Matsui, 1988), and given that Venus’s supercritical CO₂-rich atmosphere acts more like a conducting layer than a convective layer with regard to heat transfer (Snyder, 2002). The third option highlights lithospheric rheology, a fundamental component of the bolide model. Water plays a critical role in subduction initiation (and hence plate tectonics) due to the mode of failure of silicate lithospheres (Regenauer-Lieb et al., 2001). Convergence across a dry lithosphere (such as Venus) results in focused brittle failure of the upper lithosphere, while the bottom of the lithosphere deforms in a diffuse fashion and delaminates and recycles to the mantle, leaving the surface crust relatively intact. In contrast, wet lithosphere fails across its entire mechanical thickness, resulting in the formation of a narrow fault-like zone through the lithosphere; in this case the entire lithospheric thickness recycles to the mantle (Regenauer-Lieb et al., 2001). Once initiated, subduction of the lithosphere provides a means of recycling water to the mantle, triggering a double feedback mechanism (thermoelastic and thermal-rheological) to promote plate tectonics (Regenauer-Lieb et al., 2001). Thus, Venus’s lack of plate tectonics might highlight the role of rheology across the entire thickness of a planet’s lithosphere.
ACKNOWLEDGMENTS
I thank P. Cawood, P. Schultz, and an anonymous reviewer for careful and insightful reviews, and J. Goode and J. Swenson for comments on earlier versions of the manuscript. I thank the McKnight Foundation and the National Aeronautics and Space Administration (grants NNX06AB90G and NNG05GM26G) for support.

REFERENCES CITED


Manuscript received 29 March 2007
Revised manuscript received 21 June 2007
Manuscript accepted 15 July 2007
Printed in USA