

What kind of thing is a planet?

ROHAN SCOTT BYRNE

May 5, 2025

Contents

0.1	Review and Theory	1
0.1.1	Planets from first principles	1
0.1.2	Heat and geometry: planetary basics	1
0.1.3	From planetary to exoplanetary: an expanding scope of inquiry	4
0.1.4	Geodynamics: an evolving discipline	6
0.1.5	Review	13
0.2	Discussion	13

0.1 Review and Theory

The purpose of this study is to deduce what underlying parameters most strongly control a planet's choice of tectonic mode. The fundamental question regards the existence or otherwise of a well-defined 'tectonic phase diagram' in which primarily temperature defines the proper global geodynamic regime at any given time. Phase spaces are defined by their phase transitions, and in the field of planetary geodynamics, only one such transition is well-attested and amenable to study: the emergence on Earth of plate tectonics from an earlier pre-tectonic state of disputed character. The Early Earth is therefore key to the larger question of tectonic modality.

In this review chapter, we will present a brief history of scholarship on the topic, summarise the chief lines of evidence, and discuss the leading hypotheses regarding the divergent fates of Earth and our most familiar planetary cousins. Our line of inquiry will be illustrated with some simple one-dimensional modelling. We will not discuss in detail the analytical and numerical intricacies of the problem at hand, which will instead be related as appropriate in the prefacing remarks of the relevant results chapters (Chapters 3, 4, 5, & 6); nor will we provide much background or context for our methodological approach, which is treated separately (Chapter 2 & Chapter 7).

0.1.1 Planets from first principles

Our subject is planets - but what is a planet?

According to the International Astronomical Union highly-publicised ruling of 2006 ?, "a planet is a celestial body that; 1) is in orbit around a sun, 2) has sufficient mass to have a nearly round shape, 3) has "cleared the neighborhood" around its orbit." Under this rubric, Pluto, Charon, Eris, and other icy worlds of the outermost solar system are disqualified. Indeed, the rationale for the IAU's opinion was that any broader definition would result in a categorical calamity: potentially hundreds of new worlds would achieve the mantle of 'planet', including many that at the time seemed unlikely to be of much significance or interest ?, creating needless public and academic 'annoyance' ?.

The IAU's ruling sparked an impressive global backlash, including an attempt by lawmakers to quash the revision ? and various pleas for celebrity astronomers to intercede ?. The case has even become a niche topic of sociological literature, with papers discussing its effect on children ?? and its impact on popular perceptions of the credibility of astronomy ??.

In the immediate aftermath, a petition of over 300 signatures from attendees at the IAU meeting itself protested the new classification, which had been decided by a partial convention of the union rather than a general colloquium, perhaps to avoid protest ?. Alan Stern, the director of the New Horizons mission to the outer solar system, quickly became one of the highest-profile advocates for revising or ignoring the IAU opinion ?. He and his colleagues have since developed a so-called 'geophysical definition' which is now the standard at NASA ?: "A planet is a sub-stellar mass body that has never undergone nuclear fusion and that has sufficient self-gravitation to assume a spheroidal shape adequately described by a triaxial ellipsoid regardless of its orbital parameters."

Stern's definition is a profound revision of the cosmos, but one which is well-argued and well-suited to the purposes of planetary scientists. Under Stern's mandate, not only Pluto and Eris, but also the asteroid belt objects Ceres and Vesta, as well as various 'satellite planets' - Earth's moon, Pluto's partner Charon, and most of the satellites of the outer gas giants - all would be considered equally as planets. Such an expansion puts the geophysical, atmospheric, and climatic aspects of planets front and centre, above and beyond their orbital or formational histories. It sets the stage epistemologically for a fundamental deepening of our understanding of the unique processes that govern sub-stellar objects in our universe, and - as we will discuss later (Chapter 8) - may be regarded in retrospect as a milestone event in the birth of a new field of focussed, integrated, holistic planetary science.

0.1.2 Heat and geometry: planetary basics

The geophysical definition of a planet provides some clarification on what should and what shouldn't be included as essential considerations of our topic. A planet is a gravitationally-bound ellipsoidal object, not currently or formerly made of plasma, surrounded by mostly empty space. Such a definition

has some immediate implications. The chemical and gravitational prescriptions of Stern's definition provide an upper and lower spatial scale respectively: anywhere from the radius of Mimas, at 200 km the smallest rounded body in the solar system ?, to the radius of Jupiter, about 70,000 km, above which any additional mass tends to be compressed without increasing planet size ?. It implies a temporal scale too: between the atmospheric (hours) and geological (gigayear) timescales.

Other implicit constraints are more subtle. Because a planet is qualified by its shape - a function of mass and matter - it is parsimonious to stipulate that these quantities are connate to each planet and cannot be changed. Of course, in reality, such changes do occur: consider the formation of Earth's moon from a putative impact with a Mars-sized planetoid in the Hadean ?, or the hypothesised delivery to Earth of a significant mass of volatiles by icy planetoids ?. In the system outlined here, however, such events would be considered 'catastrophic' - outside of the prescribed spatiotemporal scales of the subject, and violating the prescribed conservative quantities of the problem; in essence, such a calamity would be considered the 'death' of the original planet and the 'birth' of a new one.

There are also more sustained violations: erosion of planetary atmospheres into space, for example ??, or the capture of solar protons in the aurorae ?, or the ongoing low-level exchange of meteoroids between Earth and Mars ?. Such endemic exceptions, however, plague all systems; hence we will dub them 'pathological', requiring careful handling but not deference.

Planets, then, can be said to be spheroidal, sub-stellar objects of a fixed mass and bulk chemistry. Within these parameters, however, there are considerable degrees of freedom:

- The geometric characters of the spheroid can change - eccentricity, radius, and obliquity - with consequences for the centre of mass, the gravitational field, and the tidal coupling of the planet with any extra-planetary systems.
- The distribution of mass can change, anywhere between the extremes of an arbitrarily thin hollow shell and an arbitrarily dense core, with consequences for the moment of inertia and rate of rotation as well as the gravitational field and the gravitational binding energy. The total mass, of course, cannot change - this would be 'catastrophic' - hence the gravitational pull of a planet at a sufficient distance above the surface is fixed.
- The angular velocity of the planet can change due to internal forces and geometric transformations. Angular momentum can also change, though necessarily only by interactions with extra-planetary bodies: when these interactions are periodic, systematic, and persistent over geological time, they may be parameterised as an intrinsic forcing within the planet; otherwise, they may be considered 'pathological' (if frequent) or 'catastrophic' (if infrequent) exceptions, outside the intrinsic scope of the problem.
- The chemical makeup of the planet can change. Indeed, all the degrees of freedom within any constituent molecules of a planet are available as degrees of freedom of the planet as a whole, as are all the possibilities for the breaking of old bonds and the forging of new ones.
- The isotopic makeup of a planet can change, albeit largely without consequence for bulk chemistry - hence their enormous analytical utility as tracers and monitors for Earth sciences. Earth is overwhelmingly composed of less than 300 stable nuclides ?.
- The elemental inventory of a planet can change; here, however, we must be careful. The decay of short-lived radioactive nuclides like Uranium-235 and Thorium-232 into daughter elements are significant to the system; the decays of extremely long-lived nuclides like Uranium-238 and Calcium-48 are not significant over 'planetary' timescales as stipulated above ("Nudat 2" n.d.). Decay of non-radioactive nuclides due to the finite half-lives of subatomic particles are not significant on any reasonable timescale and are not available as degrees of freedom for planetary processes as we choose to consider them.
- The spatial distribution of chemical and elemental species can change: to wit, the differentiation of the felsic crust, mafic mantle, and metallic core due to gravitational forcing; but also, the partitioning of rare-earth elements and radiogenics into the continental crust due to purely chemical factors.

Each of these degrees of freedom entail certain endothermic or exothermic reactions, and these naturally chain one to another in complex ways. The net result, however, is the exhaustion of all possible gradients and the conversion of all potential energy into diffuse kinetic energy, i.e. heat. But

what is the total internal energy of a planet - its enthalpy?

Because there is no such thing as ‘absolute’ enthalpy, it is necessary to first define a natural point of reference for the system against which changes in the energy budget may be compared. Thankfully, the stipulations we have already made regarding the conserved quantities of planets, as well as the limitations we have provided on relevant spatiotemporal scales and the resultant pruning of available degrees of freedom, imply one fairly reliable, though somewhat counter-intuitive, reference point: not the planet’s initial condition, but rather its ‘rest state’ after a sufficient span of time.

Consider a situation where the Earth as a whole exchanges no heat with its environment. In fact, this is a surprisingly credible simplification: the global geothermal flux is estimated at only 47 terawatts ??, predominantly along the mid-ocean ridge ?. This is insignificant compared to the 100,000 terawatts processed by the solar/atmospheric/oceanic system and comparable to the over 16 terawatts consumed by human civilisation in the modern era ?; a small-enough flux that, if held constant and ignoring all internal heat production, a billion years would only see a 232 K drop in planetary average temperatures ?. Indeed, looking backward, we find Earth’s core has only lost in the order of 50 K per billion years since the planet was formed ?, confirming our intuition.

What is the ‘ultimate’ temperature of an isenthalpic planet such as we have described? This will be the temperature obtained at an arbitrarily remote time in the future - long enough for all permissible degrees of freedom to be exploited. At this time, only heat remains: a planet has no more tricks to perform. It transpires that this number is not all that difficult to derive, nor so uncertain as we might assume - at least in comparison to the inverse problem, which is fraught with controversy.

On ‘Ultimate Earth’, the radiogenic nuclides have completely decayed into heat and stable daughters, while the stable nuclides have sorted by density from top to bottom - a crude but serviceable assumption. The Earth is now a perfect sphere of the smallest possible size, and its rotation has completely ceased due to tidal interactions with other cosmic masses.

We attempted to calculate the temperature of Ultimate Earth based off order-of-magnitude estimates, published scaling laws, and reference data ??????. When calculated from present day observations, we derive a temperature of between 4,000 and 5,000 K, or about 2.92×10^{31} joules: comparable to the temperature of the Earth’s core, where most of today’s heat is stored ?. When we instead venture to calculate the temperature starting from the presumed initial condition of the Earth, we obtain an Ultimate Earth temperature an order of magnitude higher - 40,000 to 50,000 K, about 2.75×10^{32} joules. In other words, the Earth’s heat has fallen by a factor of ten over its 4.5 billion year history.

The reason we can be fairly confident about these figures is that they are dominated by gravitational potential energy and the heat of accretion, which can both be trivially derived analytically from Earth’s mass; cumulatively, these factors account for about 10^{32} joules of Ultimate Earth’s heat budget, about 93%. Some 88% of this energy is converted instantly into heat upon accretion; another 9% is released through differentiation, of which 8.7% is due to the core and 0.27% is due to the crust. Only 3% of the Earth’s original gravitational binding energy remains as potential energy which could feasibly be released as work and heat. The remaining degrees of freedom - orbital, chemical, and nuclear - add up to a little over 10^{31} joules, but together make up the majority of the remaining energy available to be released. Interestingly, the amount of heat that could theoretically be liberated by gravitational sorting of nuclides - i.e. the formation of continents - is comparable to the remaining radiogenic budget of the Earth. If as much continental volume were to be produced in the next 5 Gy as in the previous, it would represent a significant proportion of geothermal flux.

In reality, the Earth is not isenthalpic: it receives solar heat and sheds geothermal heat into space. We observe here that some 89% of the Earth’s heat has been lost to space since formation. This number is very close to the 88% of heat that would have been released geologically instantaneously at accretion. As Lord Kelvin famously calculated ?, the maximum timescale to reach a present-day geotherm from a wholly molten state is a mere tens of millions of years, and much, much shorter than that if more active thermal processing is at work. Hence it is conceivable, though unverifiable, that almost all of Earth’s accretionary heat was lost shortly after formation, with freezing of the core and radiogenic decay since accounting for most if not all of the observed geothermal fluxes at magnitudes similar to today’s.

We have laid out what is admittedly a fairly rudimentary construct to commence our thesis. However, we argue it furnishes the appropriate ‘ballpark’ figures and situates the intuition in the proper place for a truly planetary way of thinking about thermodynamics and flow. The essential lesson of the Ultimate Earth concept is that the overall thermal regime of a planet is in no hurry to reach equilibrium with space; that in fact, over the lifetime of many planets - curtailed by the circumstances of their host stars to mere billions, rather than tens of billions of years - no state resembling Ultimate Earth will ever be achieved by star-bound worlds. In short, heat is not destiny for a planet.

0.1.3 From planetary to exoplanetary: an expanding scope of inquiry

It is incredible to consider that, within the lifetime of a single PhD student, a family of nine known planets has become a universe of thousands.

The expansion in scope of planetary sciences has occurred so rapidly it is easy to lose track of what has changed. Discoveries, in our solar system and beyond, are moving faster than theory, and it seems a major revolution in our understanding of planets waits to be realised. An atmosphere of genuine hope and excitement surrounds the field - both within the academy and beyond.

Here we present a brief overview of relevant recent discoveries in our home solar system and beyond.

Local planets

Unmanned exploration of our local system is entering a renewed era of great activity, emboldened by growing public interest and increasing private cash ?. This exploration is yielding provocative evidence that challenges our understanding of planetary dynamics.

While much of note has been uncovered on the rocky planets of our inner solar system, especially Mars, Venus, and the Moon, here we will discuss more recent discoveries whose import has arguably not yet been fully recognised in the broader geodynamics literature - planetary bodies that were supposed to be inert, which have proved to be full of activity. On the moons of Jupiter and Saturn, and far out beyond the orbit of Neptune, young, dynamic surfaces are being revealed, speaking to active internal processes that may support life - and hold invaluable clues to untangling the history of Earth and the deeper questions of tectonics.

Europa Jupiter’s moon Europa has proved to be surprisingly analogous to Earth: its surface ice has been shown to exhibit plate-like tectonics ?, with familiar fault patterns ?, deformation at fault tips ?, cryovolcanic ‘igneous’ terrains ?, strikingly Earth-like ridge flexures, forebulges, and other collisional features ?, and some very recent evidence of ongoing subduction ?.

Europa is heated internally by tides ?, but beyond that, little is known of its thermal state. Some have speculated the existence of deep sea vents which could be excellent candidates for hosting life ?. Whether Europa proves fertile or not, it remains the only known example of active horizontal tectonics beyond Earth, and hence is of tremendous importance to the broader question of what makes a ‘living’ planet.

Io If Europa provides some context for the Earth of today, another of Jupiter’s moons, Io, may hold the key to its past. Io, which is comparable in size and makeup to our Moon ?, is the only known body beyond Earth proven to exhibit non-icy volcanism in the present day ?. The scale of volcanism on Io is extreme and unmatched in the solar system ?; first observed during the Voyager ? and Galileo ? missions, the New Horizons flyby in 2007 detected no fewer than eleven active ejecta plumes, some jetting higher than 350 km ?. Io’s tremendous activity is related to its great internal heat, derived from intense tidal forces ?, which may even be sufficient to generate a subsurface magma ocean ?.

Though initially suspected to be relatively low-temperature due to sulphur content ?, Io’s lavas are now thought to be mostly ultramafic and extremely hot ??, fuelling analogies with the Archaean komatiite lavas on Earth ??. For these reasons, Io - and the supposed ‘heat pipe’ tectonic mode that operates there (discussed later) - may provide a critical window into the early Earth ?????.

Enceladus Saturn’s moon Enceladus may be second only to Earth in geological activity ??, making it an indispensable reference for comparative planetology.

The dynamic nature of Enceladus first became known with the Voyager missions, when its surface was found have large, crater-free terrains ?? interpreted as evidence of cryovolcanism ?, as well as some apparently tectonic features such as grabens and fault banding ?, arguing for the existence of ongoing, active resurfacing on this small, frigid moon ? where none had been considered possible before ?. Enceladus was a chief target of the later Cassini mission, which captured high-resolution photos of the moon’s young and fault-marked surface ?, acquired magnetometric evidence of a volcanically-replenished water-derived atmosphere ?, and, most impressively, observed a large cryovolcanic plume ? composed of almost pure water ? jetting from a tectonically active region around the south pole ?.

Enceladus is an ‘ice moon’, with a silicate core overlain by a mostly watery mantle and a water ice outer crust ?. Its tiny size - less than 500 km in diameter - makes its evident activity a conundrum ?. Like Europa and Io, Enceladus’ internal heat had been presumed to be predominantly tidal ?, dissipated almost exclusively throughout the icy mantle ?; however, measured heat fluxes and dynamic modelling demand alternative sources, thought to be supplied by significant radiogenic heating from the silicate core ?. Consequently, Enceladus - like Earth - has both volumetric and basal heating components, and hence a Urey number, though no published estimates for this quantity yet exist. Also, on Enceladus, like on Earth, a feedback exists between the two heat sources: while Earth core heat plays a role in redistributing and partitioning heat-producing elements in the silicate mantle, Enceladus’ radiogenic heat may shift and enhance the tidal dissipation forces in its icy mantle - which may explain why this particular moon is full of activity, while the icier but otherwise similar Mimas is inert ?.

Enceladus’ tectonics have been variously characterised as Earth-like plate tectonics ?, Venusian episodic overturn ??, and mobile-lid convection ?. Particular analogues aside, what Enceladus argues above all is that the laws of geodynamics are not intrinsic to any particular chemistry or any particular spatiotemporal scale.

Pluto Pluto, controversially relegated to ‘dwarf’ status by the International Astronomical Union ?, has since become the latest overlooked body to exhibit surprising geological activity ?. Far from being a glorified comet, the New Horizons flyby in 2015 ? showed Pluto to be a living world ?, with dunes ?, troughs ?, six kilometre-tall mountains and plateaux ?, active glaciology ?, cryovolcanoes ? and volcanic resurfacing ???, and a thin but influential atmosphere ?: all in all, a range of features that collectively “rival Mars in richness” according to one research team ?.

Most significantly, it now seems Pluto - like the other icy bodies here discussed - may possess a subsurface water ocean of its own ?; seasonal partial freezing of that ocean might explain the extensional features on the surface ? and drive some of the observed cryovolcanism ?. However - unlike Europa, Io, or Enceladus - tidal heat cannot be a factor on Pluto. This has led to speculation that the icy moon is anomalously well-insulated by clathrates ?. In any case, the vitality of Pluto’s active tectonics despite the moon’s small size and presumably meagre thermal inventory underscores that tectonics can achieve a lot with a little over geological timescales.

Exoplanets

Since the first confirmed sightings of planets beyond our solar system in the 1990s ?? the number of known exoplanets has climbed well over 4,000 ?. Most of this extraordinary success is owed to the Kepler mission, now terminated ?. Some of these worlds have local analogues; many do not ?. Planets have been observed around stars of all sizes, ages, and temperatures ?, and hundreds of multiple-planet systems have been identified, including some with known populations comparable to our own solar system ??. Statistical analyses which cross-correlate hit-rate with observation probability have concluded that planets are at least as abundant stars, and likely much more abundant, putting their number in the order of hundreds of billions to tens of trillions in our galaxy alone ?.

Though unavoidable detection biases have skewed the dataset somewhat, it is today possible to obtain reasonable quantitative estimates of our galaxy’s planetary demography. The result has been a complete overturning of traditional planetary formation theory ?. Far from being a universe dominated

by cumbersome giants, it is now apparent that small rocky or icy planets in fact predominate ?. Furthermore, the most common class of planets appears to comprise those which are larger in radius than Earth but smaller than Neptune ?. A good example of this very popular new planet class is the system of three temperate, watery sub-Neptunes recently discovered in orbit around a nearby M-dwarf ?. Our local solar system has no representative of this population and it has been speculated that, if one did exist, it may have been destroyed ? or ejected ? at some point long ago. These exciting new worlds are often close-orbiting and are tantalising candidates for life ?.

Determining the interior structures, or even the density, of exoplanets discovered using radius-based methods can be difficult. Very large-radius bodies may be assumed to be Jupiter-like gas giants purely by the requirement that, if they were any denser, they would collapse. The density of smaller planets is more indeterminate; although there is some direct evidence from transit spectroscopy ? and ideally more in the near future ?, typically we are forced to rely on a combination of statistical inference and broad mass-radius scaling ??. Such methods suggest that planets within the highly populous 1 - 4 Earth radius interval fall into two broad camps: ‘super-Earths’ which are rocky but potentially much more massive than our planet, and the smallest of the sub-Neptunes, implicitly volatile-rich planets with only slightly greater masses but significantly larger radii ?. Between these camps exists an apparent ‘radius gap’ between 1.5 to 2 Earth radii ?; this has been taken to imply that an evolutionary tipping point exists above which runaway gas accumulation invariably turns rocky planets into icy ones. Modelling suggests the bifurcation is not a sharp one; some oversized rocky worlds and undersized gassy ones are expected ?. Nevertheless, the ‘Neptune transition’ implies that the scope of ‘Earth-like’ planetary science can practically be neatly constrained to the range of half to double Earth-radius. Fortunately, there is a great abundance of such planets: thousands, in fact, within 50 lightyears alone ??.

Perhaps because we have no local examples, super-Earths have invited particularly intense speculation. Much debate surrounds the question of whether such worlds should be expected to have plate tectonics or not. The debate is pertinent to purely Earth-based studies because it goes to the question of what plate tectonics actually requires in order to commence and continue. Indeed, the super-Earth scholarship is almost evenly split between those who claim plate tectonics is more likely on larger worlds ?????, those who claim it is less likely ?????, and those again who argue that size has an inherently ambiguous effect ????. The disagreement does not break down simply along methodological lines, with numerical modelling, analytical arguments, and geochemical considerations variously employed in support of all three contentions.

The super-Earth debate arguably generates more heat than light, but it is nonetheless illuminating in that it reveals the true uncertainties of geodynamics where we are unable to rely on the arrogance of fact, as we can on Earth. It prompts consideration of whether we would even have a notion of plate tectonics if we did not observe it in action at home, and invites speculation about what completely alien solutions may exist to the fundamental question of planets.

0.1.4 Geodynamics: an evolving discipline

Tectonics, from the Greek *tectonicus*, ‘pertaining to building’, is a body of theory which endeavours to characterise the manner in which the spatial configuration of the landmasses changes over time. It is one of the foundational theories of natural history and arguably comprises the backbone of modern geosciences. However, tectonic theory is not, per se, a ‘dynamic’ theory with the power to explain how, or indeed, why such reconfiguration occurs. It is in this distinction that the field of geodynamics arises.

Geodynamics is to the theory of plate tectonics as genetics is to the theory of evolution. It seeks to interface the enormously successful narrative power of tectonics with the explanatory and predictive power of the underlying laws of physics - that is, classical mechanics and thermodynamics. Geodynamics defined in this way is necessarily younger than tectonic theory, which is itself only two generations old; as a discipline, it is arguably yet to move beyond descriptive science and acquire true predictive and prescriptive powers.

Empirical foundations of geodynamic thought

The world that geodynamics sets out to describe is characterised by several axiomatic properties. Some have been apparent since ancient times; others only much more recently. What is important to note is how powerfully suggestive they are of a unifying underlying phenomenon when assembled side by side, as follows:

- The Earth's upper crust is partitioned chemically between felsic (continental) and mafic (oceanic) terrains ?.
- The Earth's crust is partitioned texturally between extensive smooth regions (plains and plateaux) and localised, predominantly arcuate rough regions (hills and mountains) ?.
- The Earth's surface is partitioned topographically between dominant low-elevation regions (ocean floors and continental shelves) and scattered high-elevation regions (continental cores) ?.
- The Earth's upper crust is partitioned mechanically between broad areas of low or moderate strain rate (plate interiors) and narrow, globally interlaced bands of high strain rate (plate boundaries) ?.
- The Earth's lithosphere can be shown to be clearly partitioned into zones of common velocity when the vectors are constructed around Euler poles ?.
- The Earth's surface is partitioned chronologically between temporally heterogeneous zones of intermixed extremely young and extremely old rocks (continents) and temporally gradated zones made of rocks no older than 300 million years (ocean floors) ?.
- The landmasses of the Earth are variously scattered into fragments or consolidated into supercontinents throughout deep time ?.
- The bulk Earth is partitioned chemically and gravitationally between volatiles (atmosphere and hydrosphere), silicates (crust and mantle), and heavy metals (the core) ?.
- The bulk Earth is heated basally by conduction across the core-mantle boundary as well as volumetrically (although not necessarily homogeneously) by radiogenic decay ?.
- The Earth, it is fairly well established, is round ?.

Such a bill of facts, once collected, testifies strongly to the existence of a global assortative process with both toroidal and poloidal components. Because of the apparent thermal and compositional gradients involved, the immediate sensible process thus implicated is convection, i.e. transport by thermal buoyancy. Convection implies zones of upwelling and downwelling: rheological features which can be identified geologically with ocean ridges ? and trenches ? respectively. Better still, the pattern of geothermal fluxes aligns almost exactly with what would be expected if Earth's surface plates represented the upper limbs of mantle-scale convection cells ?. All these virtues explain why, from the mid-20th century onward, the model of global plate movement driven by mantle convection has served as a unifying explanatory paradigm for all geoscience.

Challenges to the conventional model of plate tectonics

While the essential logic of convection-driven plate motion as a model for the Earth is sufficiently robust as to be beyond serious doubt, the scope of uncertainty at finer levels is daunting.

The Naive Theory of plate tectonics To illustrate how our understanding of global geodynamics has evolved, it is helpful to clearly articulate a 'Naive Theory' of plate tectonics to serve as a null hypothesis for the discussion. According to the Naive Theory:

- Earth is in a state of whole-mantle convection driven mainly by accretionary heat escaping from the core.
- Material and thermal heterogeneities, while certainly present, do not determine the global convective planform.
- Global convection is powered by the thermal gradient between the deep Earth and outer space, and should be expected to dwindle as this heat reservoir is depleted. Hence, we would expect elevated global geothermal fluxes in the past relative to the present.
- Plate motions represent the upper limb of these whole-mantle cells, with the plates themselves in some sense propelled by the overall mantle current. Implicitly, poloidal motions are diagnostic,

while toroidal motion is merely a waste of convective energy.

- Mantle-scale upwellings should be expected to map to mid-ocean ridges and downwellings to subduction zones. Implicitly, plate scale forces such as ridge push, basal traction, and slab pull should cooperate to transport the plates according to whole-mantle circulation, and ought to be of similar magnitudes.

In general, the ‘Naive Theory’ is how the solid Earth would be expected to behave if it was dominantly basally heated, subject to linear viscosity laws, and chemically homogeneous to a first-order estimation. Such a rheology is powerfully amenable to numerical analysis and has been extensively studied as a common end-member behaviour for all sorts of systems in a state of thermally-driven flow (this thesis treats with similar rheologies in Chapter 3 and Chapter 4). Indeed, the ‘Naive Theory’ serves as an excellent approximation for systems ranging from Earth’s atmosphere to Campbell’s tomato soup.

Insufficiencies of the Naive Theory However, when applied to the bulk Earth, the Naive Theory fails in nearly every respect, even as a generalisation. For one, it massively overestimates the contribution of ridge push to plate motions: slab pull is acknowledged as by far the dominant factor ?. It also understates the apparently very substantial role played by chemical heterogeneities, most obviously vis a vis continental crust and the supercontinent cycle ??, but also, increasingly, in respect to the hypothesised basal chemical heterogeneities at the core-mantle boundary ??, which some suspect to be integral to the global convective planform ??.

More generally, upwellings and downwellings do not always correspond with plate boundaries, are not in general symmetric, and can range in size from mantle-scale to lithospheric scales. With respect to downwellings, these can be plane-symmetric and aligned with (collisional) plate boundaries, as seen in subduction zones ?, but may or may not be mantle scale depending on plate strength and upper mantle temperatures ??; there is also a significant role played by delamination and dripping, which can occur on a range of scales and almost always remote from plate boundaries ?. Upwellings, too, can be both mantle-scale and smaller-scale. ‘Hotspots’, as seen beneath Hawaii and East Africa, may fall into either category, depending on the interpretation of the tomographic and geochemical evidence. When major upwelling zones coincide with plate boundaries, it is strictly in the form of plumes, as seen in Iceland and East Africa ?, with any planiform upwelling beneath the mid-ocean ridge being a strictly plate-instigated, shallow phenomenon ?. Finally, toroidal motions are far from superfluous, but appear to actively enhance convective vigour by reducing viscous dissipation at the surface ?.

Naive Theory and the Urey Paradox One sense in which the Naive Theory is surprisingly apt is in its minimisation of the role of radiogenic heating relative to secular cooling as a driver of convection in the mantle. The proportion of surface heat flux accounted for by radiogenic decay in the mantle is called the Urey Ratio ? and it was originally supposed to be close to unity: that is, the Earth releases its heat as rapidly as it is produced ?. This assumption was required to circumvent the so-called ‘Urey paradox’, in which unrealistic temperatures are obtained in the early Earth if today’s geothermal flux is mostly accounted for by secular cooling. However, multiple lines of evidence - including estimates of Earth’s original radiogenic inventory based on chondritic compositions ?, as well as state-of-the-art geoneutrino surveys ???? - now agree on a much lower Urey ratio, somewhere in the neighbourhood of one third ??.

In one sense, a low Urey ratio is favourable for the Naive Theory, as it implies that basal heating dominates volumetric heating: a state of affairs that promotes vigorous long-wavelength convection ??. However, to avoid the Urey paradox, this would require an inverse relationship between global geothermal flux and net mantle temperatures - in other words, it seems that the Earth sheds its heat more rapidly the cooler it gets ?. This counter-intuitive behaviour is hard to evoke with the kinds of isoviscous or exponentially temperature-dependent rheologies that the Naive Theory supposes.

Going beyond the Naive Theory We have employed the Naive Theory here mostly as a discursive device. However, though no theory of this form has ever been seriously advocated by qualified geodynamicists, the Naive Theory is a surprisingly close representation of common opinion regarding

solid Earth dynamics, both among the lay public and in the non-specialist planetary and geoscientific literature. This can be seen, for instance, in the generally accepted supposition that Mars and the Moon lack active tectonics due to their lesser size, and hence - presumably - their more rapid rate of cooling. The resilience of Naive-like assumptions about the Earth is likely due in some degree to the fact that, as misguided as the theory is, its elegant formulation and superficially potent explanatory qualities still recommend it over alternative models, the underlying causal mechanisms of which remain elusive.

So, what might an improved, unified geodynamic model of the Earth entail? We can draw up a decent catalogue of what such a model would require by merely inverting the assumptions of our by now well-falsified null hypothesis:

- Earth is in a state of mixed convection on multiple interconnected wavelengths, with the dominant driver being negative buoyancy of the lithosphere.
- Chemical and thermal heterogeneities are significant and have a first-order effect on global convective planform.
- Instead of convective vigour being determined by global temperatures, temperatures are in fact determined by convective vigour, which has a range of complicated dependencies. Global heat flux can go inversely with global temperature and potentially exhibit many other counter-intuitive feedbacks.
- Plate distributions powerfully condition global flow, and are themselves a function more of historical and local contingencies than of deep perturbations.
- Mantle-driven upwellings and downwellings are to some degree separate from surface-driven movements, with the two systems at times reinforcing and at times disrupting one another.

What conceivable system, governed by what conceivable laws, could provide for these behaviours? Over the past two decades, significant progress has been made from various quarters to provide a new paradigm that satisfies these requirements without stipulating causes beyond the essentials of thermally driven flow. Two aspects of this paradigm will now be discussed: the discovery of tectonic modes, and the special role of continents.

Continents: a decisive control in the tectonic system

Much of the progress of recent years is owed to the development of robust and versatile computational numerical modelling tools, which have proved to be particularly efficacious for investigating non-linear, and indeed, highly counter-intuitive systems ?. The story they tell is one of a global flow regime controlled by chaotic interactions between high Rayleigh number instabilities and pronounced material heterogeneities: in particular, the strong, cold, low-density materials that make up the continents.

Thermal effects of continental crust One major observation from recent geodynamic modelling is that a planet's global flow regime needn't necessarily be that which maximises heat flux ?. A good example is the effect of continental crust on mantle temperatures. Intuition holds that the presence of thick continental lithosphere should insulate the mantle, leading to reduced net surface heat flux ??? . This feedback is often given a role to play in the periodic assembly and breakup of supercontinents, by triggering the formation of large sub-continental mantle plumes ??.

Modelling, however, no longer supports this. Even in purely thermal models, supercontinentality does not necessarily lead to increased sub-continental temperatures ?. In dynamic models, only fairly sluggish rheologies experience any meaningful sub-continental warming from insulation ??; when geochemistry is considered, the effect of the partitioning of heat-producing elements into continents can be shown to substantially outweigh any insulation effect ?. In fact, under more sophisticated rheologies, increased vigour due to transient warming ? and geometric flow conditioning imposed by supercontinentality ? may in fact boost the global surface flux. In this way, the apparent regularity of the supercontinent cycle - the first-order feature of Earth's Phanerozoic history - could be accounted for purely by surface-driven forcings of the underlying mantle, without the need for catastrophic interventions like massive mantle plumes.

Continents as a resolution of the Urey paradox Continents may also provide a partial resolution to the Urey paradox. With some exceptions, estimations of the global heat budget over time have not explicitly taken into account the role of continents ???. Numerical modelling suggests that mantle heat flow, to a first order approximation, is unaffected by historically typical degrees of continental coverage ?. As higher continental abundances imply lesser radiogenic forcing in the mantle due to partitioning, the net result is a lower Urey ratio for much of Earth's history. Feedbacks of this sort resolve the Urey paradox without the need for untestable heterogeneous mantle models ?.

Continent-climate feedbacks Another under-appreciated role of continents in mantle convection comes indirectly by way of the atmosphere. Through albedo ?, silicate weathering ??, carbon storage ?, and topographic effects ??, continents have long been acknowledged as a potent factor in long-term global climate.

However, it is only recently that climate forcings on mantle flow have begun to be recognised. Numerical suite modelling shows that mantle flow can be suppressed as surface temperature rises ?; conversely, modelling of subduction zones suggests the opposite coupling, as global warming enhances sediment runoff, lubricating subduction and possibly accelerating mantle overturn ??. The net effect of climate may be so profound as to induce a complete switch in the style of tectonics ??; the picture only becomes more complicated when the variegated feedbacks between continental drift, evolution, and bioclimatic controls are also taken into account ?. Whether climate is a stabilising or destabilising agent in solid Earth circulation remains in question, as does the significance of tectonic-climate feedbacks to the maintenance of planetary habitability ?.

Tectonic modes: beyond plate tectonics

In the 1990s, the failure of linear mantle models to capture the major features of plate tectonics, even at extreme Rayleigh values ?, motivated the development of non-linear, stress-dependent rheologies ??. These models had the virtue of exhibiting strain-localisation, wherein the upper boundary layer separates into discrete units that can be generalised as plates ??. The new models supported many interesting discoveries; most notably for our purposes, the discovery of the first non-trivial 'tectonic mode' to be observed outside of nature, the 'episodic overturn' mode ?, now widely believed to be operative on Venus ??. The discovery of episodic overturn gave credence to the importance of 'phases' in rheological parameter space; as applied to planets, these phases became known as 'tectonic modes' and are today a major topic of study. While some are fairly trivial, like the 'stagnant' regime, others are more surprising.

Mobile lid When the convective vigour is much higher than the yield stress, a state is obtained where part or all of the lithosphere is failing at any given time. The 'mobile lid' category as it is presently defined is generous: it encompasses plate tectonics as well modes exhibiting very broad, or even uniform, lithospheric strains patterns.

Under an isoviscous rheology, a mobile tectonic mode persists in all but the most extreme cases ?; nevertheless, there are subtleties to the mobile lid that are presented in depth in Chapter 3.

Stagnant lid When the yield stress is very high compared to the convective vigour of the mantle, a very strong, thick, static lithosphere may be formed. In a continuum mechanics sense, a stagnant lid could be characterised as any flow regime in which the maximum strain rate at the upper boundary layer is close to zero. More recently, authors have distinguished between a hot stagnant lid, as seen perhaps on the very early Earth ??, and a cold stagnant lid, represented by the present-day Moon ?.

The stagnant lid mode is a typical consequence of depth- or temperature-dependent rheologies such as Arrhenius-type viscosity ?, which is the partial subject of Chapter 3.

Episodic overturn When the yield stress is nominal relative to convective vigour, a temporally unstable regime ensues in which long periods of lid stagnation are interspersed with abrupt failure events. Episodic overturn has a special place in tectonic mode theory: in models that reproduce a

brittle lithosphere ? in which a parameter implementing mechanical strength is the control variable, the episodic regime can be shown to occupy an intermediary position between stagnant and mobile lids. This observation demonstrates the feasibility of a ‘phase diagram’ type conception of planetary tectonic mode: testing the applicability of this paradigm is an ongoing and active field of research ?? . Venus is commonly considered a representative of this class based on the uniformity of its surface ages, around 300 My ?.

Episodic overturn is the main subject of Chapter 4 of this thesis.

Magma ocean In a magma ocean, the upper lithosphere is by definition in a state of global, continuous, contiguous failure: this is the definition of a fluid. Most magma ocean models have only a shallow ocean, and most are short-lived - indeed, so short as to be below the characteristic thermal timescale ?. Hence, characterising the magma ocean as a tectonic mode is somewhat fraught. The Hadean eon of the Earth is so-called because of the assumption that a sustained magma ocean existed at this time ?, and it is traditionally assumed that this is so for all the rocky planets ?, though mounting evidence of a ‘cool early Earth’ increasingly call this into question ??. The Moon and other small rocky bodies preserve the most unambiguous evidence of magma ocean tectonics, with the Moon’s highland-lowland dichotomy serving as the classic fossil example ?????.

The magma ocean, and the manner and timing of its demise, is an important consideration for models of the early Earth (Chapters 5 & 6).

Sluggish lid If viscosity is only moderately temperature-dependent, a ‘sluggish’ regime can develop ?, with large aspect-ratio convection cells leading to diffuse deformation zones in the upper boundary layer. Although the lid is technically mobile in this scenario, the motion is driven by mantle currents rather than vice versa, resulting in relatively tepid convection ? and limited or strictly sub-lithospheric recycling ?. While Earth is sometimes considered to have passed through a sluggish lid phase ?, Mars is the oft-cited bellwether for this mode, based on its extensive deformed but ancient terrains ?, isotopic analysis of Martian meteorites ?, and numerical modelling ?. A sluggish lid can be recognised by the existence of substantial toroidal flow in the upper boundary layer in the absence of significant poloidal flow.

The possibility of a sluggish lid on the early Earth is contemplated in Chapters 5 and 6 of this thesis.

Plume tectonics If the defining characteristics of plate tectonics are horizontal motion, arcuate deformation, and surface-driven convection, plume tectonics represents the opposite: vertical motion, bulbous deformation, and basally-driven convection. Plume tectonics, also called ‘vertical tectonics’, preceded plate tectonics as the popular explanation for the global distribution and structure of major landforms ?, and it remains popular with those unconvinced by a dominantly surface-driven global circulation regime ??. The picture is complicated by the fact that the two substantially independent systems ? nevertheless seem to not only co-exist but cooperate on the Earth, with plumes inducing, or being induced by, the formation of triple junctions at the surface ?, a process which apparently goes back to the Archaean ?. Plume tectonics is also connected with plate tectonics in that it is potentially the source for much of Earth’s early continental crust ?? via a process which may be ongoing today; consider the oceanic plateaux, such as the Kerguelen and Azores, which arguably represent Phanerozoic proto-continents ??. It has lately been hypothesised that plume tectonics not only preceded plate tectonics on the Earth ? but actively gave rise to it ?, by means of any one of numerous modelled mechanisms ??.

Controversies aside, plume or vertical tectonics is a vital part of the emerging tectonic mode paradigm. Beyond Earth, plume tectonics is the preferred explanation for the Tharsis terrain on Mars ?????, where it appears to have been localised, massive, and extremely long-lived ?; it is also implicated by various features on the surface of Venus including coronae and nova ??????.

The relationship between plume tectonics and plate tectonics, and the possibility and mechanisms of a plume-to-plate transition, are the main subject of Chapters 5 and 6 of this thesis.

Heat pipe The heat pipe model was first postulated to explain certain observations of Jupiter's moon Io ?, but it wasn't until more recently that it attained a rigorous thermodynamic treatment ?. Today it is recognised as a distinct tectonic mode - arguably the second truly non-trivial mode to be discovered after episodic overturn ?. In this model, a very strong, thick, cold lithosphere overlies a hot interior; long-lived plumes generate persistent volcanism on the surface, which continually re-buries the vacuum-chilled planet surface and so cools the interior by a process of pressure-driven downward cold advection ?. Heat pipe tectonics is even more efficient at disbursing internal heat than plate tectonics - Io's surface heat flux is estimated at an incredible 40 times that of Earth ?.

The heat pipe mode is one resolution to the problematic jigsaw puzzle of the early Earth heat budget ??; though it has yet to be convincingly replicated in numerical modelling, it remains a hot topic in the literature.

Null mode In the interest of completeness, the case of a planet thoroughly at equilibrium with the temperature of free space should be mooted. Such a mode, here called the 'null mode', is a perfect attractor in parameter space upon which all planets must eventually converge if left to their own devices. Dwarf planets such as Ceres likely approximate the null mode in our solar system; if the Earth were set adrift in interstellar space, it would reach such a state on the order of 100 Gy from now ?.

The future of tectonic modes The discovery of the episodic overturn and heat pipe modes makes it almost certain that more unseen modes remain to be uncovered. While limitless novelty can be generated with more and more complex material configurations, the real test of the paradigm will be to show that single-material, mathematically parsimonious laws, like the viscoplastic formulation that enables episodic behaviour, can continue to generate substantially original behaviours that can be called truly new tectonic modes.

Tectonic mode transitions and planetary memory

So far, we have encountered mode transitions as diagnostic or taxonomic classifications of viable systems for processing geothermal heat. This is a geophysical analogue of how weather and climate may be understood fundamentally as systems for processing solar heat. According to this manner of thinking, planets should be expected to 'select' the most efficient system for processing their heat that is made possible by their circumstances at any given time ?.

We have already seen that a planet's geothermal flux must not be naively expected to decrease as a function of time, or even as a function of temperature: both may indeed have the opposite effect. However, that does not invalidate the null hypothesis that the flux should be the highest possible.

Consider a schematic tectonic history of the Earth ?: first, a magma ocean blasted by accretionary heat, core separation, and short-lived radiogenics; then, as surface cooling drove the magma to depth, an Io-like heat-pipe phase; third, a sluggish plume-tectonic phase akin to Venus or perhaps Enceladus; and finally, with sufficient cooling yielding a sufficiently brittle lithosphere, the breakup of one plate into multiple plates and the dawn of plate tectonics. According to this history, geothermal flux starts high when global temperatures are high, dips through the single-plate era, then rises with the advent of subduction; at each stage, the tectonic mode is a direct and unitary consequence of the state variables of global relict and radiogenic heat. This is what we will call the 'phase diagram' model of tectonic mode transitions, wherein a planet tracks a path across a multidimensional tectonic phase space ??? just as a block of water ice may be compelled by its state variables of temperature and pressure to cross the liquid domain on its way to being vapour.

This way of viewing tectonic modes is increasingly challenged by geodynamic numerical modelling. A number of groups have recently furnished evidence that a planet's choice of tectonic mode is inherently multistable ???, indeterminate ?, remote from equilibrium ?, and heavily influenced by stochastic ?, contingent ??, historical ?, and circumstantial factors ???.

These lines of evidence suggest a new model: one in which tectonic mode is treated not as a state variable or a mere product of state variables, but rather as a process variable - a configuration that flows through the system, like heat itself ?. Under such a model, multiple tectonic modes could be seen

as coequal simultaneous possible configurations for each planet, each manifesting certain exigencies and instabilities that could spontaneously and unpredictably precipitate a transition into an alternative mode. Thus two planets identical in every way may, in sufficient time, evolve to be more and more unlike, until the ultimate logic of heating and cooling forces both to converge on stagnation and eventual, frigid death.

Though the new model is not yet well quantified, there are many facts we have already seen that should predispose us to suspect it may be true. Consider again the analogy with weather and climate. Non-linearities and chaotic behaviour are as influential there as it is here claimed they are in the solid Earth. However, there are two key differences. Firstly, planetary tectonic mode not only processes its heat source, but feeds back into it by determining the rate at which it is conveyed to space. On Earth, we see evidence of this effect most definitively in the lower-than-expected modern Urey ratio, which implies that Phanerozoic surface processes have in fact created a much more efficient channel for extracting heat from the core than existed prior. Secondly, planets have long-term memories. While the oldest climate information with direct relevance to the modern atmosphere is ~200 My old carbon, Earth's tectonic processes can be affected by circumstances billions of years old: for instance, when a buoyant Archaean craton impacts a subduction zone and forces a reconfiguration of plate stresses or even plate boundaries ?.

While ultimately it may be tempting to cut through the uncertainty and hysteresis and return to the basic intuition that the Earth is becoming thermally 'older' with time, with less net heat to go around and hence less prospect for activity, it must again be stressed that the actual rates of flux, even in our elevated present regime, are very small with respect to tectonic timescales. Even a seemingly irreversible process like the formation of cratons on Earth is conceptually not undoable: as we calculated earlier, sufficient potential energy remains inside the Earth to effect the 'rewinding' of that particular degree of freedom. In principle, it might even be possible to establish a circle - or, more properly, a very shallow spiral - of alternating tectonic modes, which, if sufficiently regular, might be construable as a distinctive tectonic mode of its own: consider episodic overturn ?. Such a cycle, despite the inevitable ratcheting of entropy, could nonetheless conceivably be stable even on timescales comparable to the present age of our universe.

0.1.5 Review

The emerging model of tectonic modes will demand novel and imaginative approaches. This study's particular methodology will be discussed forthwith (Chapter 2); for now, we must conclude with a reaffirmation of the essential thesis. The age of planets as dumb rocks mechanistically unspooling into space has come to an end. Felled by a torrent of empirical evidence at home and far beyond, that conception is now being replaced with a new one: broader, richer, more ripe with possibilities than what came before; but also challenging, subtle, and bewilderingly vast. It is to be hoped that this new paradigm will in time yield new tools; tools which, if wielded with boldness and ingenuity, may open up the tomb of Earth's ancient past, and lay bare the pathways of life.

0.2 Discussion