Understanding the Thermal Evolution of Earth

Martin Wolstencroft

Supervisors: J. Huw Davies and Omer F. Rana

Submitted to
The School of Earth and Ocean Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Cardiff University

December 2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Thesis advisors

J. Huw Davies and Omer F. Rana

Author

Martin Wolstencroft

Understanding the Thermal Evolution of Earth

Abstract

Studying the thermal history of Earth's mantle can provide a better understanding of Earth's evolution on a planetary scale. In this work, several mechanisms affecting the thermal evolution of Earth's mantle are investigated.

The Nusselt-Rayleigh power law relationship (Nu(Ra)) was calculated from the results of a series of models with three dimensional spherical geometry and free slip boundary conditions. Basally and internally heated convection was examined. For $Nu(Ra) = aRa^{\beta}$, β was found to be 0.294 ± 0.004 for basally heated systems and 0.337 ± 0.009 for internally heated systems. Model cases were extended to Rayleigh numbers higher than any previous study (10^9). β was not observed to reduce at high Rayleigh number, therefore, as this mechanism cannot be invoked to moderate thermal flux in the past, the influence of time dependent layering on thermal evolution was considered.

A parameter space exploration of Rayleigh number and 660 km phase change Clapeyron slope demonstrates that present day Earth could have a partially layered mantle and that full two layer convection is possible in the past at higher Rayleigh numbers. Evolution of mantle temperature was modelled, with the models cooling from

an initially layered state. As layering breaks down at high Rayleigh numbers, the mantle passes through a wide domain of partial layering before achieving whole mantle convection. The partially layered regime is characterised by a series of avalanches from the upper into the lower mantle. When an avalanche reaches the core mantle boundary it triggers a pulse of plume-like instabilities in the opposing hemisphere, producing a pulse in global surface heat flux. As the mantle cools, the avalanche-pulse events evolve towards higher frequency and lower magnitude.

If this mechanism occurs within Earth, the gradualist view of Earth's thermal evolution may need to yield to a more event-driven model. The mechanics of avalanche-pulse events could also provide an explanation for geochemical observations of periodic maxima in melt extraction from the mantle.

The modelling of Earth's mantle produces large data volumes. A distributed computing solution to the data storage problem was investigated. The system, MantleStor, is based on Peer-to-Peer technology and intended to operate over hundreds of standard workstations. A trial implementation demonstrates that MantleStor is able to safely store data in a challenging network environment. Data integrity was maintained with over 30% loss of storage machines. MantleStor is an example of an e-Science project, a discussion of e-Science and its implications is presented.

Contents

Citations to Previously Published Work

- A slightly shorter version of Chapter 2 has been accepted for publication by Physics of the Earth and Planetary Interiors: Wolstencroft, M., Davies J. H. and Davies D. R. Nusselt-Rayleigh number scaling for spherical shell Earth mantle simulation up to a Rayleigh number of 10^9.
- A 4 page version of Chapter 5 has been published: Wolstencroft, M., Rana, O. F. Davies, J. H. (2006) Distributed Storage of High-Volume Environmental Simulation Data: Mantle Modelling. IEEE/WIC/ACM International Conference on Web Intelligence. Hong Kong, IEEE Computer Society.

Any contributions made to thesis Chapters by co-authors of the above mentioned papers do not exceed the level of contribution that a supervisor would be expected to make.

Acknowledgments

I would like to thank: Liz for her support, patience, understanding and proof reading and my parents, Chris and Bärbel without whose support I could not have reached this point.

I have been very fortunate to be advised by two excellent supervisors, Huw Davies in the School of Earth and Ocean Sciences and Omer Rana in the School of Computer Science who have supported me throughout the project.

I would like to recognise the contributions of John Baumgardner, the author of TERRA and Rhodri Davies, who produced the multi-level multigrid version of TERRA. The code they have produced was central to this project. John Brodholt was instrumental bringing me into the Mineral Physics Consortium on HECToR, The UK National Supercomputer, without his efforts this thesis would be considerably thinner! Ian Merrick and Ian Thomas administered Helix and provided technical and philosophical assistance on many occasions, thank you.

From the School of Computer Science I thank Mike Daley for introducing me to Java, Roger Philp, Jon Giddy, Matthew Shields (Triana/P2PS), Ian Wang (P2PS) and the Cardiff University ARCCA team.

The members of my research group: Dave Oldham, Peter Bollada and Peter Webb are acknowledged for their part in stimulating my thoughts with their observations and discussions.

And finally, my fellow postgraduates, who have taken part in many multifarious discussions and have provided an excellent level of humour during intense times: Alan Hastie, Sarah Dare, Bryan Hatton, Marcus Badger, Anna Hey, Cat Burgess, Dave Holwell, Julia Sas, Catherine Baudon, Päivi Heiniö, Shanshan Huang, Ben Rabb, Falko Mathes, Matthew Minifie, Jenny Moss, Heather Birch, Heather Price and Tracy Aze.

This research was funded by NERC eScience grant NER/S/A/2005/13131.

For my grandparents

Für meine Großeltern

Chapter 1

Introduction

1.1 Studying Earth

One of the lasting legacies of the 20th Century space race is the image of Earth from space. This view changed perceptions of our planet and assisted in the birth of modern geoscience. In parallel, the theory of plate tectonics became generally accepted, demonstrating the significant influence of Earth's interior on its surface processes and biosphere. These realisations stimulated research into the mechanisms driving plate tectonics and the planet as a whole, the study of geodynamics.

The interior of Earth is difficult to study as direct access is impossible. Fortunately, the mid 20th Century also saw the development of affordable computing hardware (e.g. the Intel 8086 processor, 1978). This allowed computer processing and modelling to become an intrinsic part of scientific research. In the context of geodynamics, it became possible to model processes that could never be observed directly or syn-

thesised in the laboratory. It also became possible to apply complex algorithms to seismic data, enabling imaging of Earth's interior. The study of geodynamics and indeed much of modern science would not be possible without such computational power.

Science has always been about models, be they conceptual or mathematical. The accuracy of a mathematical model is dependent on how closely it represents the physical system. As models become more accurate they also become more complex and thus more computationally expensive. Model accuracy is therefore closely linked to the available computing resources. At the time of writing, the power of available supercomputers is such that it is possible to model Earth's mantle with the correct geometry at resolutions approaching 10 km. The more pressing technical problem has become one of how to analyse and visualise the output of these models in sufficient spatial and temporal detail to understand how the mantle has evolved over time. With visualisation comes a high data volume, which must be stored. Data storage has developed in parallel with the microprocessor, but modelling is as often limited by the amount of output that can be stored as it is by the power of a given supercomputer. Thus, there is a need to seek out new ways of storing model output data.

Modelling alone will achieve little if predictions are not testable. In geodynamics, the main observational methods that feed into modelling are: geochemistry, seismology and mineral physics. A model must reconcile observations from these disciplines

if it is to be considered an accurate representation of Earth.

1.2 The Structure of Earth

Earth's mantle is the region located between the crust and the core (Figure 1.1), it consists of solid silicate material but can be considered to act as a liquid on geologic time scales. It is important to note that on an atomic level, the flow mechanism is not the same as for a liquid. Mantle rock deforms as point defects migrate through the crystal lattice. Thus, the mantle is able to transfer heat by convection. On short time

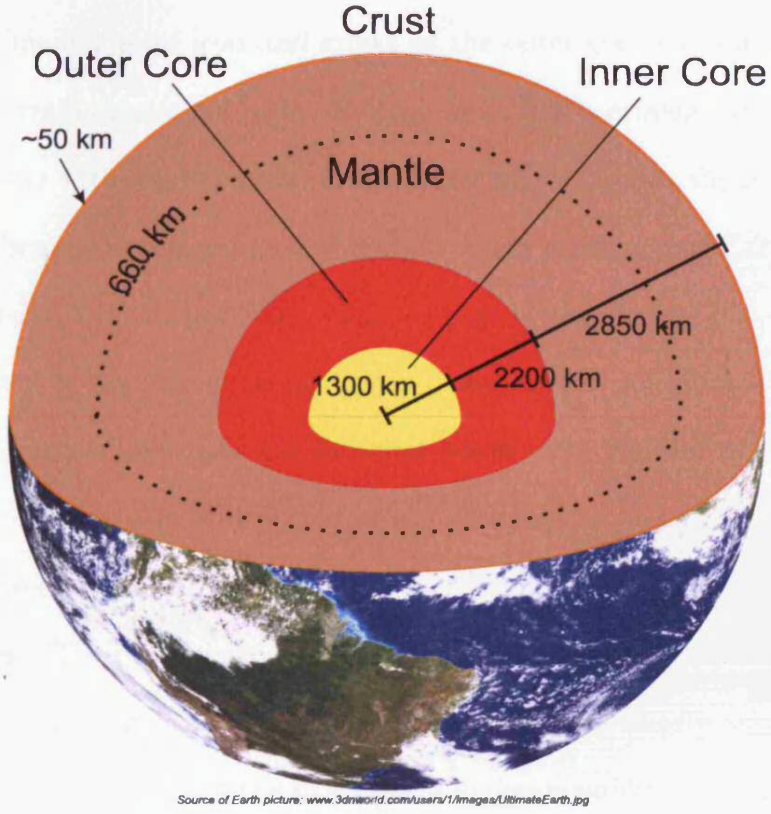


Figure 1.1: A sketch of Earth's interior showing the solid crust and mantle, the liquid outer core and the solid inner core. The 660 km depth mineral phase change is also indicated. Not to scale.

scales, the mantle is solid and transmits earthquake waves in an elastic manner. The 660 km boundary indicated in Figure 1.1, is often used to separate the mantle into upper and lower portions. The boundary represents a pressure related phase change in the primary mantle mineral olivine, an iron-magnesium silicate. There are several other phase changes that occur within the mantle, the most important of which are at 410 km and just above the core-mantle boundary.

The outer core is composed of liquid iron and a small quantity of light elements. The outer core convects vigorously, giving rise to Earth's magnetic field. The inner core is predominantly solid iron and grows as the outer core cools and freezes. The crust is Earth's rigid uppermost layer. It comprises thick continents and thinner ocean floor and together with a thin section of the upper mantle, forms the lithosphere. The lithosphere is broken into a number of plates, which move across Earth's surface at an average rate of 2-3 cm per year. The motion of the plates, like the mantle, is driven ultimately by buoyancy forces. Ocean plates sink at subduction zones because they are colder and denser than the material below. The rigidity of the plates adds complexity as it potentially allows compressional forces to be transmitted over great distances. Convection in the mantle below plates may also be capable of exerting a significant stress. The question of whether plate motion is instigated by 'slab pull' or 'ridge push' is somewhat outdated as mid-ocean ridges are clearly extensional in nature, while the brittle nature of the lithosphere makes it unlikely that the extensional forces can be transmitted over great distances without being taken up by faulting. It may be more useful to visualise ocean plates in particular as sliding down the gravity

gradient from ridge to subduction zone, in a similar sense that a river flows down to the sea. The crust is the point at which the solid Earth interacts with the hydrosphere, atmosphere and biosphere, making it one of the more important boundaries on Earth.

1.3 Thermal Evolution

In the most basic terms, the thermal evolution of Earth is the process of how the initial heat of formation and subsequent radiogenic heating interact with the mechanisms cooling the planet. Parameterised convection modelling (Mckenzie & Weiss 1975, for an early example) has been used for many years due to it's relative efficiency. The scaling relationships used are derived from both theory (e.g. Turcotte & Oxburgh 1967) and experimental work (e.g. Richter et al. 1983). The most common representation of parameterised modelling is a temperature-time curve from formation ≈4.6 Billion years ago (Ga) to present (e.g. Figure 1.2). The only available fixed point is the present day heat flux of ≈44 Terawatts (TW) (Pollack et al. 1993) and the assumption that Earth was hotter in the past, the difference being the secular cooling (dissipation of the initial heat of formation) and the reduction in radiogenic heating rate.

The relationship between mantle temperature and the vigour of convective heat transfer is related to viscosity. The viscosity of the mantle is strongly temperature dependent (Busse 1989, and references therein). The expected result of this is (1) large lateral variations in viscosity with lateral temperature variations and (2) more rapid

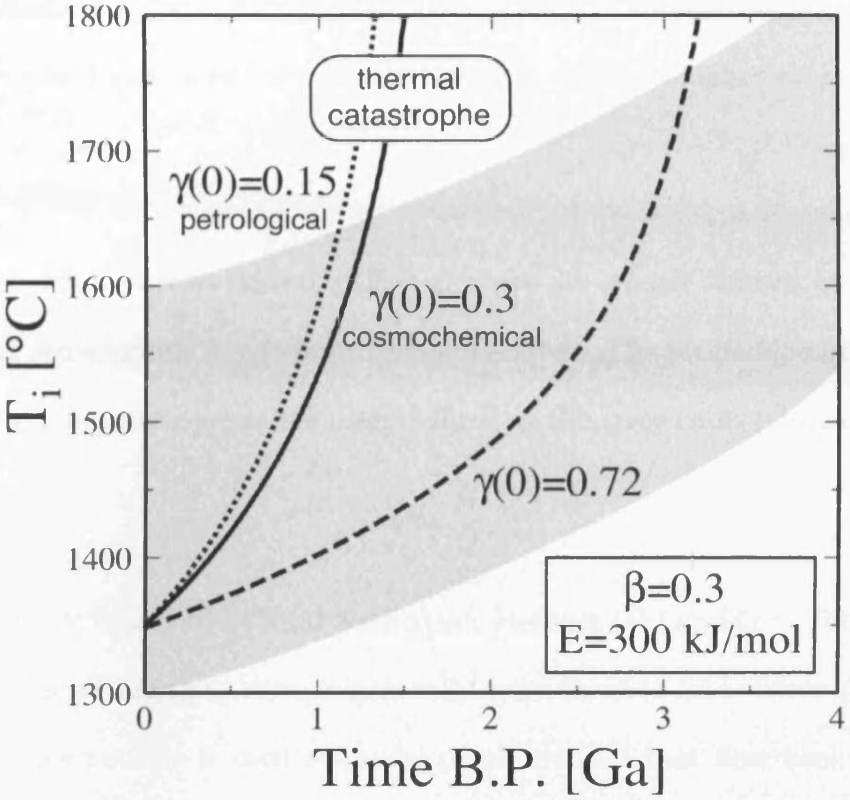


Figure 1.2: An example of a thermal evolution model. T_i is average mantle temperature, γ is the Urey ratio, E is the activation energy for temperature dependent viscosity and β is a scaling factor relating convective vigour and mantle heat flux (discussed in detail in Chapter 2). From Korenaga (2008)

convection if the mantle was hotter (lower viscosity) in the past (Tozer 1965). The former effect has been explored experimentally and shown not to significantly impact convective scaling relationships (Richter et al. 1983). It may be that the latter effect buffers mantle temperature to some degree, allowing more rapid cooling when the mantle is hotter. The difficulty with such an interpretation is the unsteady and time dependent state of mantle convection may preclude the kind of stabilisation require for such a mechanism to operate effectively. A related concept is the possibility that the core-mantle boundary heat input has remained relatively stable over time, due to

the crystallisation of the inner core and the efficiency of conduction across the CMB. The mantle could still have been hotter however, due to a higher rate of radiogenic heating in the past.

Most of the other factors affecting this process are poorly known or are difficult to reconcile. An example of a typical problem follows. The proportion of surface heat flux due to radiogenic sources has been defined as the Urey ratio (Christensen 1985):

$$\gamma = \frac{H}{Q} \tag{1.1}$$

Where γ = Urey Ratio, H = Total Radiogenic Heating (W) and Q = Total Heat Loss (W). The value of the Urey ratio is generally considered to lie between 0.15 and 0.3. However, if this number is used to model parameterised heat flow back in time, the temperature of the mantle rapidly tends to infinity before 4 Ga (e.g. Korenaga 2005, Figure 1.2). A Urey ratio of 0.7 gives a better trend but is unlikely to reflect the true composition of the mantle. In any case, after painstakingly estimating the Urey ratio of the mantle, it seems improbable that the answer is inaccurate to such a degree. This apparent paradox has acted as motivation in the search for mechanisms, which modify Earth's past heat flux. A key factor in much of this investigation is of course the unique position of Earth in the solar system, such that water exists as liquid, solid and gas at the surface and in the atmosphere. Plate tectonics is also certain to play an important role. Many of the features of the convecting mantle are also still poorly understood.

1.4 Further Problems and Paradoxes

In addition to the Urey paradox described above, there are other apparent contradictions in the deep Earth. There is uncertainty as to the magnitude of the various sources of the present Earth surface heat flux. Table 1.1 demonstrates the discrepancy between what is measured at the surface and what can be accounted for within the Earth. Whether this discrepancy represents a real problem, or could be explained by improved error estimates, is unclear.

Source/sink	*Magnitude (TW)*
Heat flux at Core-mantle boundary	2-7.5
Radiogenic heat production in mantle	12
Radiogenic heat production in crust	8
Secular cooling	9.4
Total surface heat flux	44
Apparent unaccounted sources	-7.1 to -12.6

Table 1.1: Earth's energy balance (Pollack et al. 1993).

The efficiency of the convecting mantle system exerts a powerful control over the heat flux. Given that there is a finite amount of heat within the Earth, the dissipation of this heat into space will depend on how efficiently heat can leave the system. As the thermal conductivity of silicate rock is very low, convection in the mantle ($\approx$80 % of planetary volume) is the primary means of heat transfer to the surface.

Geochemical investigations of the mantle also provide some interesting problems. Argon 40 is produced by the decay of potassium 40 with a half-life of 1.3 Gyr. As the

Argon atom is too heavy to leave Earth's atmosphere it is reasonable to assume that any ^{40}Ar produced has remained in the Earth system. The produced volume of ^{40}Ar is controlled by the amount of ^{40}K in the Earth at formation. However, observations indicate that only ≈50% of the expected ^{40}Ar can be accounted for (Allègre et al. 1996). It has been suggested that the 'missing' Ar is held in some deep reservoir in the mantle.

Non-volatile bulk Earth composition is often assumed to be similar to chondrite composition as chondrite meteorites are thought to represent the raw material out of which the solar system formed. This assumption was backed up with a wide variety of geochemical studies that showed most non volatile elements in the Earth were present in chondritic proportions. Unfortunately, high sensitivity studies such as those by Boyet & Carlson (2005) and Bennett et al. (2007) find that samples previously identified as having similar $^{142}Nd/^{144}Nd$ ratios to chondrites actually show a small but significant deviation. This indicates that either the Earth is not of chondritic composition, or that some process has removed certain elements or isolated them within the Earth, resulting in a situation where the accessible mantle is of slightly non-chondritic composition.

What the geochemical interpretations described above share, is the idea of an isolated reservoir or reservoirs within the Earth, which are only rarely sampled. Unfortunately, seismic tomography finds no evidence major for separations within the mantle at present (Figure 1.3). Either some underlying assumption about the bulk

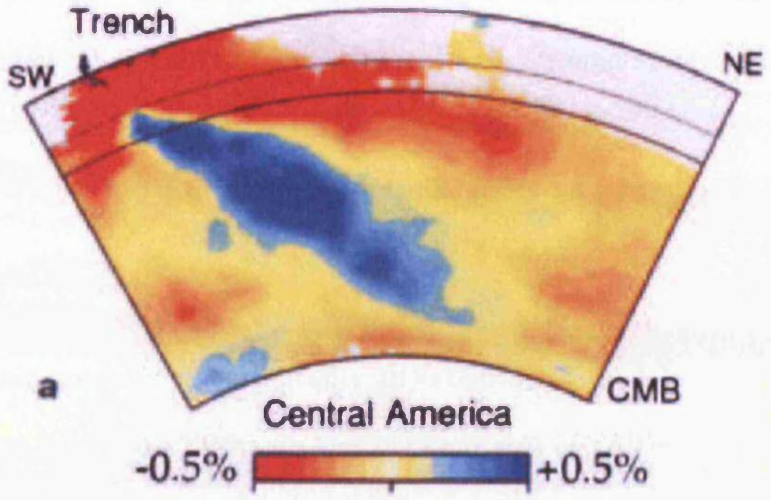


Figure 1.3: Seismic tomographic image showing a subducted slab penetrating the lower mantle. Red represents slow (hot) material and blue is fast (cold) material. From: Van der Hilst et al. (1997).

Earth composition is wrong or there is a seismically 'invisible' volume of material that has remained largely unsampled for billions of years.

Studies into the behaviour of Helium (He) provide a counterpoint to the apparent contradictions of mantle geochemistry and seismology. Measurements of He isotope ratios may be used to compare Mid Ocean Ridge Basalt (MORB) and Ocean Island Basalt (OIB). Results seem to show that OIB has a higher $^{3}He/^{4}He$ ratio in comparison with MORB. The conclusion has been drawn that OIB represents the sampling of a deeper isolated mantle reservoir while MORB samples the well mixed upper mantle. However, Parman et al. (2005) suggest that He may not behave completely incompatibly during melting and that the OIB/MORB difference can be explained by melting of depleted mantle. This is supported by the generally lower amounts of

He in OIB as opposed to MORB (Yamamoto & Burnard 2005). The He example serves as a reminder that many of the chemical and isotopic systems relied upon by various mantle theories may not be adequately understood.

1.5 Modelling

Computational modelling is essentially an extension of the laboratory experiment. In the laboratory, there are limits on the physical size of an experiment and on the magnitude of fundamental parameters such as temperature, pressure and material properties. These limits may be imposed by cost, practicality or the fact that the system may be impossible to model accurately at a small scale. The mantle falls into the latter category, particularly where geometry is concerned, because it is not possible to model a whole planet with its own gravity on the workbench. Thus, mathematical modelling has arisen to allow more realistic systems to be understood and experimented with. Simple 1-Dimensional parameterised models of the mantle require very little processing and can be done on paper, for example, the convection modelling of Turcotte & Oxburgh (1967). More sophisticated models use 2-Dimensional, 3-Dimensional or 3-D spherical geometry.

Examples of geodynamical modelling codes include: ConMan (2-D) (King et al. 1990), Citcom(S) (3-D and 3-D spherical) (Moresi & Solomatov 1995, Zhong et al. 2000), StagYY (3-D spherical geometry) (Tackley 2008) and TERRA (3-D spherical geometry) (Baumgardner 1985, Bunge & Baumgardner 1995, Yang & Baumgardner 2000). This study uses TERRA, a FORTRAN code, which has the advantage of be-

ing computationally efficient and well established, having been in use since the 1980's. TERRA is based on an efficient multigrid solver, scales well with increasing resources and can model the whole mantle at high convective vigour. Additionally, there are a number of methods within TERRA that allow the tracing of chemical heterogeneities (Stegman et al. 2003) and selective radial resolution refinement (Davies 2008). No code is perfect and TERRA has some weaknesses that must be appreciated in order to correctly interpret results. TERRA does not generally implement temperature dependent viscosity efficiently, the effects of rheology and tectonic plates (when included) must be imposed as a surface velocity condition. From a software design perspective, TERRA follows a typically 'scientific' approach; there are as many versions of TERRA as there are users and it is written in the procedural programming style. While this does not cause performance issues, the code is not as accessible as a modern object-oriented code. As alluded to above, this is a problem that many scientific codes share. A final and universal disadvantage is that all high resolution mantle modelling codes require large supercomputers, machines that are expensive to procure and maintain.

The technical aspects of TERRA relevant to this work are described in greater detail in Chapters 2, 3 and 4. Briefly, it is a finite element 3-D spherical geometry code with a computational grid derived from a regular icosahedron. The computational grid is produced by pairing the 20 triangles of the icosahedron into 10 diamonds and subdividing these diamonds repeatedly. The domain is then decomposed into multiples of 4 for allocation to individual processes. The code has been parallelised

using MPI (message passing interface) by Bunge & Baumgardner (1995). At the start of a run an initial temperature condition is provided and the code then solves fluid dynamic equations for the conservation of mass, momentum and energy. The equation of state may be compressible or incompressible. TERRA is run on cluster-style supercomputers, from the most basic 'Beowulf' machine made by linking standard PCs together, to £20 million CRAY XT4 machines. Typical computational resources required by TERRA range from 8-1024 processors and from 2-2048 GB of RAM, depending on model resolution. Output may be generated for a number of visualisation tools, Generic Mapping Tools (GMT) and MantleVis were used in this work.

1.5.1 Challenges Posed by Modelling

High model resolution is desirable as it allows Earth-like parameters to be used and permits results to be more directly compared to Earth's mantle. High resolution requires large computational resources, particularly in terms of system memory. There are methods that allow higher resolution in important regions of a model (the boundaries) without incurring the cost of high resolution over the entire computational grid (Davies 2008).

Even with such techniques, the volume of data that can be generated by a high resolution model case is significant, particularly if modelling involves ensemble runs and time-series visualisation. The volume of data produced by codes such as TERRA could easily overwhelm available storage if all potential data were output. The expensive solution to this problem would be to purchase more central storage or a large

tape archive system. An alternative (less expensive) method is to distribute storage over a number of networked machines. This is done in a closely coupled manner on clusters using file systems such as NFS or Lustre but generally operates as 'scratch' storage, not as an archive. A more loosely coupled approach might be to use many widely separated workstations to store data in their unused hard disk space, using a lightweight distributed storage application. The latter approach has been investigated in this work through the construction of MantleStor, a peer to peer (P2P) based storage system.

1.5.2 Techniques of Hard Disk Scavenging

A variety of distributed storage systems exist and tend to be used for online content delivery and search services. The main technique used in this study is that of P2P networking. P2P design principles assume unreliability, decentralised structure and low performance overhead. P2P technology has been one of the most interesting emergent technology of the Internet age as is enables large files to be distributed efficiently (e.g. BBC iPlayer). Systems such as Skype also make extensive use of P2P technology. MantleStor has been built upon a P2P API (Applications Programming Interface) called Peer to Peer Simplified (P2PS) developed by the Triana Research group at Cardiff University (Wang 2005). P2PS is conceptually similar to the more widely known JXTA (Gong 2001). P2PS provides a library of Java classes, which provide a P2P communication infrastructure between machines running a P2PS peer. Messages are passed using XML (Extensible Markup Language) Advertisements. MantleStor extends some of the features of P2PS by adding file transfer, storage, indexing and a

search mechanism.

1.5.3 Indexing

The central aspect of storing and searching data within any system is the indexing system. For MantleStor, it was decided that information relating to the model case that produced the data should be used to index it. The reason for this choice is simple, TERRA output files are named in a manner that means little to anyone other than the user who ran the code. Therefore, for the archive to be of any use, data should be tagged with information recognisable to any geodynamical modeller. Typical examples include: the time span over which the model has run, model resolution, physical input parameters and output values such as the Nusselt number, a non-dimensional parameter of heat flux. The indexing method is the key part of this storage system, distinguishing it from other storage and P2P applications. It adds a layer of data discovery that can be exploited to guide the user when considering which model runs to carry out when investigating new geodynamical questions. In short, it puts historical model outcome information in the hands of the user.

1.5.4 Summary

This work presents research into the thermal evolution of Earth through 3-D spherical geometry modelling and considers the e-Science approaches required for this class of problem. The scaling of heat flux with convective vigour, the viability of phase change layering and the effects of time dependent layering of the mantle are investigated using near Earth-like parameters. By using distributed peer to peer

techniques it should be possible to not only overcome the storage limitations on TERRA simulations, but also to create a distributed lookup, which would give users a better overview of extant model runs.

1.6 Structure of the Thesis

The thesis is composed of two major themes (a) understanding mechanisms affecting the thermal evolution of Earth and (b) the exploration of a distributed storage solution to the data deluge produced by the modelling. Chapters 2, 3 and 4 represent a progression through a series of mechanisms related to the thermal evolution of Earth, with the conclusions of each chapter motivating the next. Chapters 5 and 6 propose, test and analyse the distributed storage system devised (MantleStor); chapter 7 uses experiences gained throughout the project to examine the wider implications of computing in science (e-Science).

The thesis is assembled such that the core chapters (2-7) are written in the style of self-contained papers. As a consequence, each chapter contains an abstract, introduction, previous work, methods, discussion and conclusion. Some common aspects of the work appear in several chapters. For example, TERRA is described in chapters 2-4. In each case, the aspect of the code most relevant to that chapter is presented in detail. All references are collected in a bibliography at the end.

Chapter 2

Nusselt-Rayleigh Number Scaling for Spherical Shell Earth Mantle Simulation up to a Rayleigh Number of 10^9

2.1 Abstract

An investigation of the power law relationship between Nusselt number (Nu) and Rayleigh number (Ra) for Earth's convecting mantle is presented. The Nu(Ra) relationship was calculated from the results of a model with three dimensional spherical geometry and free slip boundary conditions. Both basally and internally heated convection has been examined. For $Nu(Ra) = aRa^{\beta}$, β *was found to be* 0.294 ± 0.004 *for basally heated systems, which is lower than the value of* $\frac{1}{3}$ *suggested by conven-*

tional boundary layer theory. The exponent $\beta = 0.337 \pm 0.009$ *for internally heated systems, when the internally heated Ra is converted to a basally heated equivalent Ra to allow comparison. The influence of the method used to calculate* β *was also considered, with particular attention paid to high Ra. As an example of the significance of* $\beta = 0.29$ *rather than* $\frac{1}{3}$*, a Ra of* 10^9 *results in a surface heat flux which is* $\approx$ *32% lower. There is no evidence that* β *reduces at high Ra; that mechanism cannot be used to moderate mantle temperature when projecting back to early Earth conditions. The differing planform of basally and internally heated models was shown to result in different scaling relationships between RMS surface velocity and Ra for the two modes of heating, in particular, a much lower surface velocity for internally heated cases relative to equivalent Ra basally heated cases.*

2.2 Introduction

The aim of geophysical research is to understand Earth and its processes over the entire history of the planet on all scales. The endeavour is hindered by incomplete knowledge of some of the basic mechanisms of planetary development. The thermal evolution of Earth is one such area. Earth is, to a large extent, a heat engine. Assumptions regarding the evolution of temperature over time appear in most, if not all, conceptual models of plate tectonic theory, geochemistry and geophysics.

A favoured method of investigating Earth's thermal evolution has been one dimensional (1D) parameterised modelling (e.g. Turcotte & Oxburgh 1967) of the mantle, which at over 80% of the Earth volume is by for the most important factor in the

cooling of Earth. A number of valuable insights are possible from such relatively simple models (Stevenson et al. 1983, Korenaga & Jordan 2002, Korenaga 2003, Sharpe & Peltier 1978, Schubert et al. 1979, Turcotte 1980, Davies 1980, Honda 1995, Nimmo et al. 2004, Choblet & Sotin 2000). Even when much more complex mantle convection simulations can be undertaken in spherical geometry (Zhong et al. 2000, Tackley et al. 1994, Bunge et al. 1997, Davies 2005), there are still value in the parameterised models since: (i) they provide a convenient way to run a large number of simulations, (ii) allow vigors of convection to be considered, which cannot be resolved by full 3D numerical models, and (iii) can allow simulations of thermal evolution over virtually all of Earth history.

A critical aspect of these parameterised models is the assumed relationship between the ratio of convective to conductive heat transfer, characterised by the non-dimensional Nusselt number (Nu):

$$Nu = \frac{q}{q_K} \tag{2.1}$$

where: q is the heat transferred by convection, while $q_K = \mathrm{k}\Delta T/\mathrm{D}$, is the amount of heat that would be conducted through a layer of thickness D with a temperature difference ΔT across it, k being the thermal conductivity.

The other important parameter is the convective vigour, characterised by the Rayleigh Number (Ra):

$$Ra = \frac{g\rho\alpha\Delta T D^3}{\kappa\mu} \tag{2.2}$$

where: g is the acceleration due to gravity, ρ is density, α is the coefficient of thermal expansion, ΔT is superadiabatic temperature drop across the shell, D is domain thickness (2900 km), κ is thermal diffusivity, and μ is dynamic viscosity. Both equations are presented in their most common forms after Davies (1999).

A series of models with sub Earth-like parameters can be used to derive the Nusselt-Rayleigh (Nu(Ra)) number relationship:

$$Nu(Ra) = aRa^{\beta} \tag{2.3}$$

where a is a constant and β, the index of the power law relation, is the main value of interest. Nu(Ra) can be used to extrapolate to present day and ancient Earth-like conditions (higher Ra). These values can be compared with surface heat flux measurements and thus used to constrain the thermal budget of Earth today. Further back in time the Nu(Ra) can be used to estimate heat fluxes at higher mantle temperatures (higher Ra).

The value of the Ra (Equation 2.2), is also often quoted to demonstrate how Earth-like a particular model is in terms of its convective vigour. Earth's mantle is considered to have a basally heated Ra in excess of 3×10^6 (Davies 1999, assuming $\mu \approx 10^{22}$ Pas) and probably in the range 10^8 (Bunge et al. 1997, Weeraratne & Manga 1998,

assuming a greater contribution from internal heating). The assumed Ra of Earth is based on the Ra one obtains when putting Earth values into the Ra equation. The accuracy of these values is open to interpretation, leading to the overall uncertainty as to the exact Ra of Earth's mantle.

Previous work to evaluate the relationship between Nu and Ra number has concentrated on 1D, 2D and 3D Cartesian box domains, with only a small selection of studies employing Earth-scale 3D spherical geometry (Bercovici et al. 1989, 1992, Iwase & Honda 1997). Although models that do not employ 3D spherical geometry can provide a valuable insight into mantle mechanisms, they may be limited by imposed boundary conditions, for example, the 'side walls' of a 3D box and artificially imposed aspect ratios.

Many studies carried out in the 1980's and early 1990's lacked sufficient computing resources to model high degrees of structural complexity, with many using significantly higher than Earth-like viscosity and limiting convection to low spherical harmonic degree or restrictive initial conditions to improve computational efficiency (Bercovici et al. 1989, 1992). In this study, the work of constraining the relationship between Nusselt number and Rayleigh number has been advanced by (i) using spherical geometry and (ii) utilising much higher resolutions and so more Earth-like values for viscosity and Ra. This allows the relationship to be constrained over a broader range of Ra and to higher Ra. This gives greater confidence in the derived parameters of the relationship, and limits the degree of extrapolation. Very long model

runs were undertaken allowing models to be started from short-scale random initial conditions. This removes the need for large-scale initial conditions since over their long evolution the simulations adopt their preferred large-scale pattern. Since more than one large scale planform can be stable at some Ra (Bercovici et al. 1989), using a random initial condition avoids the possibility of making the 'wrong' choice *a priori.*

Historically, β was derived from boundary layer theory (Busse 1989, and references therein) in a square box model and estimated to be of order $\frac{1}{3}$ (Turcotte & Oxburgh 1967). Boundary layer theory does not consider effects to do with wavelength of convection or the action of multi-scale flow (e.g. plumes), therefore the actual value of β is unlikely to be exactly $\frac{1}{3}$. Mckenzie et al. (1974), in the first 2D Cartesian modelling of mantle convection, plot the Nu(Ra) scaling but do not explicitly state β, focussing more on the velocity scaling. Christensen (1984) suggested that β may be closer to 0.1 for systems with a rigid lid at the upper boundary. However, Gurnis (1989) argued that the presence of plate tectonics on the Earth allowed much greater cooling efficiency, similar to that of a free-slip upper boundary. In his more Earth-like models β was generally around 0.3. More recently, Solomatov and colleagues (Solomatov 1995, Reese et al. 1999) have derived the Nu-Ra number relationship for the stagnant lid and sluggish lid regimes. This work follows Gurnis, focussing on the mobile surface model as being most representative for Earth. The stagnant and sluggish lid regimes might be more meaningful for other terrestrial bodies and therefore the relationship derived here should not be applied to other bodies without first deciding what regime best describes their evolution. Korenaga & Jordan (2002) and Korenaga

(2003) confirmed a value of around 0.3 for models of the onset of convection with temperature dependent viscosity. A result indicating that, although simpler models lack some of the more subtle features of the real mantle, Nu(Ra) appears to be robust. Stevenson (2003) also noted that β is rarely observed to be exactly $\frac{1}{3}$ as predicted by the simple boundary layer models. Experimental work (Castaing et al. 1989) and some low vigour modelling work (Bercovici et al. 1989, 1992) has produced results where $\beta = \frac{2}{7}$ or 0.28. This has lead to the suggestion that there is a transition in the behaviour of a thermally convecting system at Ra beyond 10^7, with increasing convective vigour becoming less efficient at increasing the thermal throughput of the system. The impact of such a transition would be to make β variable, reducing at higher Ra. As this study utilises high Ra, the presence or absence this effect has been investigated.

2.3 Simulation Method

The most important controlling parameter for this study is the Ra. The actual Ra for Earth is thought to lie in the region of 10^8. To achieve a similar Ra in 3D spherical geometry requires high resolution modelling, which in turn requires very significant computing resources. Previous work has used the approach of simulating a range of lower Ra and using the derived scaling to extrapolate the results to higher values of Ra. This study also utilises this approach, however, recent innovations in the TERRA code and the availability of very large computing resources allows us to model up to Ra = 10^8. The consequence of this is that very little extrapolation is required.

The TERRA code which is utilised here (Baumgardner 1985, Bunge & Baumgardner 1995, Yang & Baumgardner 2000, Oldham & Davies 2004) is a benchmarked, finite element fluid dynamics code used to simulate Earth's mantle in 3D spherical geometry. TERRA is extremely scalable, and at high resolution allows near Earth-like viscosities to be used with a lateral resolution varying from around 15 km at the core-mantle boundary to around 30km (or possibly better) at the surface. Lower resolution can be used for the sake of efficiency where lower Ra result in thicker thermal boundary layers, which are resolvable at lower spatial model resolution.

In order to achieve the very high Ra cases (cases B0 and I0, defined below) a new version of TERRA was used. Davies (2008) applied a multi-level multigrid approach, which allows *a priori* selective radial refinement of the model resolution. Thus it was possible to model the mantle with a resolution of $\approx$ 14 km at the surface and thereby resolve thermal boundary layers at a basally heated Ra $= 10^8$ and an internally heated Ra $= 1.44 \times 10^9$.

The dynamical problem solved by TERRA can be defined in terms of conservation of mass

$$\nabla \cdot \mathbf{v} = 0 \tag{2.4}$$

momentum

$$\frac{1}{\rho}\nabla P = \nu\nabla^2\mathbf{v} + \alpha g \Delta T_a \hat{r} \tag{2.5}$$

and energy

$$\frac{\partial T}{\partial t} + \mathbf{v}\cdot\nabla T = \kappa\nabla^2 T + \frac{J}{\rho C_p} \tag{2.6}$$

where: $\mathbf{v}$ is velocity, P is dynamic pressure, ν is kinematic viscosity, t is time, T is temperature, ΔT_a is adiabatic temperature drop, J is rate of internal heat generation per unit volume, C_p is specific heat at constant pressure and $\hat{r}$ is the inward directed unit radial vector.

A Newtonian viscosity constitutive equation was assumed (Bunge & Baumgardner 1995). In this work, we assume the Boussinesq approximation, where density differences are neglected except in the buoyancy term of the momentum equation. The Prandtl number is assumed to be infinite.

The equations are solved dimensionally in TERRA but can be non-dimensionalised using the following relations (after Davies & Stevenson 1992):

$$x' = \frac{x}{D}, \quad t' = \frac{\kappa t}{D^2}, \quad T' = \frac{T}{\Delta T} \quad \text{and} \quad P' = \frac{D^2 P}{\mu\kappa} \tag{2.7}$$

where 'primes' indicate non-dimensional terms and x is distance.

From these one obtains:

$$\mathbf{v}' = \frac{\mathbf{v}D}{\kappa} \quad \text{and} \quad Q' = \left(\frac{J}{\rho C_p}\right)\left(\frac{D^2}{\kappa\Delta T}\right) \tag{2.8}$$

where Q' is non-dimensional internal heating and $\mathbf{v}'$ is non-dimensional velocity. Dropping the primes Equations 2.4, 2.5 and 2.6 can then be written as:

$$\nabla \cdot \mathbf{v} = 0 \tag{2.9}$$

$$\nabla P = \nabla^2 \mathbf{v} + RaT\hat{r} \tag{2.10}$$

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = \nabla^2 T + Q \tag{2.11}$$

Thus it becomes evident that the Ra is the key controlling non-dimensional parameter.

The technical details of TERRA are available in the literature (Baumgardner 1985, Bunge & Baumgardner 1995, Yang & Baumgardner 2000). Briefly, the equations are solved in the following order; an initial temperature field T is set up and is utilised to solve for velocity and pressure, by means of an Uzawa pressure correction algorithm. The velocity is then used to find $\frac{\partial T}{\partial t}$, the rate of change of temperature, providing a new temperature with which to continue the calculation (Oldham 2004). These equations are solved by a second order Runge-Kutta method, which is used to advance the calculation by a timestep Δt. The computational domain (i.e. the spherical shell) is discretised by means of a regular icosahedron, where the 20 triangular sides (paired as 10 diamonds) are repeatedly divided into 4 sub-diamonds until the desired resolution is achieved. The degree of refinement is referred to as 'mt' and increases in powers of 2 i.e. 16, 32, 64, 128, 256 and so on. The grid is extended radially by placing several of these spherical shells above one another, generating a mesh of triangular prisms

(layers) with spherical ends. In this study, nr, the number of radial layers, is set to:

$$nr = \frac{mt}{2} + 1 \tag{2.12}$$

The following assumptions were used in this study: free slip upper and lower boundaries, no mineral phase changes, incompressible rheology, whole mantle convection, and constant viscosity. For basally heated cases both boundaries were isothermal, whilst for internally heated cases the upper boundary was isothermal and the lower boundary insulating.

The manner in which Ra and Nu are calculated depends on the style of heating. For a model where all heat enters through the base of the model:

$$Ra_{bheat} = \frac{\alpha \rho g \Delta T d^3}{\kappa \mu} \tag{2.13}$$

$$Where : \kappa = \frac{k}{\rho C_p} \tag{2.14}$$

and the Nusselt number is

$$Nu_{bheat} = \frac{htop \times d \times (rmax - rmin)}{4\pi k(rmax \times rmin)\Delta T} \tag{2.15}$$

As internally heated cases have no fixed lower boundary temperature, different methods for calculating Ra and Nu must be adopted. For internally heated cases Ra is

$$Ra_{iheat} = \frac{\alpha \rho^2 g H D^5}{\mu \kappa k} \tag{2.16}$$

and the Nu (after Reese et al. 2005) is

$$Nu_{iheat} = \frac{H \times \rho \times rmax^2 \times (1 - \eta^2 \times (3 - 2\eta))}{6k(Tavg - tb(1))} \tag{2.17}$$

Where: η = inner to outer shell radius ratio (0.546... for the Earth), H is the rate of internal heat generation (W/kg), $rmax$ is the outer radius of the shell, $rmin$ is the inner radius of the shell, $Tavg$ is the average temperature over all radial layers, $tb(1)$ is the temperature at the surface, and $htop$ is the surface heat flux.

The concept of Nu and Ra remains the same. Where previously one was able to use ΔT, now the rate of internal heating combined with the shell thickness and thermal conductivity is used to produce a ΔT equivalent, based on heat flux across the domain (Davies 1999). This allows Nu and Ra to be calculated for internally heated cases.

As can be seen from comparing Equations 3.1 and 3.2, the internally heated Ra will be 1-2 orders of magnitude larger than a basally heated Ra from an otherwise similar model. An internally heated Ra may be converted into its basally heated equivalent by dividing it by the corresponding Nu (after Davies 1999):

$$Ra_{bheat_i} = \frac{Ra_{iheat}}{Nu_{iheat}} \tag{2.18}$$

Ra_{bheat} and Ra_{bheat_i} are used for all graphs unless stated otherwise.

2.3.1 Input Parameters

While we have shown above that our models are controlled by the dimensionless Ra, the TERRA code actually solves the equations dimensionally. For completeness, the actual input values that remain unchanged from case to case are presented in Table 4.4. In this study, the Ra was varied by varying the dynamic viscosity of the model. This produced Ra_{bheat} in the range 10^3 to 10^8 for basally heated models and Ra_{iheat} from 10^4 to 10^9 for internally heated models. Varying the Ra in this way over 5 orders of magnitude allows calculation of a more robust value for β than has previously been possible.

The initial conditions for each case were also identical (Figure 2.1) and are based on low amplitude, short wavelength temperature variations (note: only temperature is needed as an initial condition). Such an initial condition inevitably extends the time taken for a given model to evolve but has the advantage of not imposing an initial planform to the model. The cases that were investigated in this study are outlined in Table 2.2. 8 cases were entirely basally heated and 7 cases were entirely internally heated. Viscosities ranged down to the near Earth-like value of $\approx 10^{21} Pa \cdot s$. Model runs were allowed to evolve until a statistically stable condition was achieved. This differs from true steady state in that the models are intrinsically unsteady at high Ra, a feature which is likely to be shared by the real mantle. In this study, a

Parameter	*Value*
Equation of State	Incompressible and Boussinesq
Reference Density	4.500×10^3 Kgm^{-3}
Gravitational Acceleration	10 ms^{-2}
Volume Coefficient of Thermal Expansion	2.500×10^{-5} 1/K
Thermal Conductivity	4 W/m/K
Specific Heat at Constant Volume	1.000×10^3 Jk$^{-1}kg^{-1}$
Temperature at Surface	3.000×10^2 K
Temperature at CMB (Basal Heat)	2.850×10^3 K
Thermal Boundary Condition at CMB (Internal Heat)	Insulating
Viscosity Profile	Uniform
Boundary Conditions (velocity)	Free Slip
Inner Radius of Spherical Shell	3.480×10^6 m
Outer Radius of Spherical Shell	6.370×10^6 m

Table 2.1: Common input values used for all cases. Note: As described above, since the governing equations can be expressed non-dimensionally these values are simply here for completeness.

simulation is considered to have reached 'steady state' if there is no strong long-term trend in the surface heat flux as a function of time.

An important question is whether a case has been sufficiently well resolved. As an approximate measure, the number of radial model layers within the boundary layer should be ≥ 5 (after Lowman et al. 2004), (Fig. 2.2); it is this requirement that dictates which resolution a case is most efficiently modelled at. High viscosity (low vigour) cases can be modelled using fewer radial layers as the thermal boundary layers are thicker. At lower viscosities (higher Ra), higher resolution is required to accurately resolve thinner boundary layers.

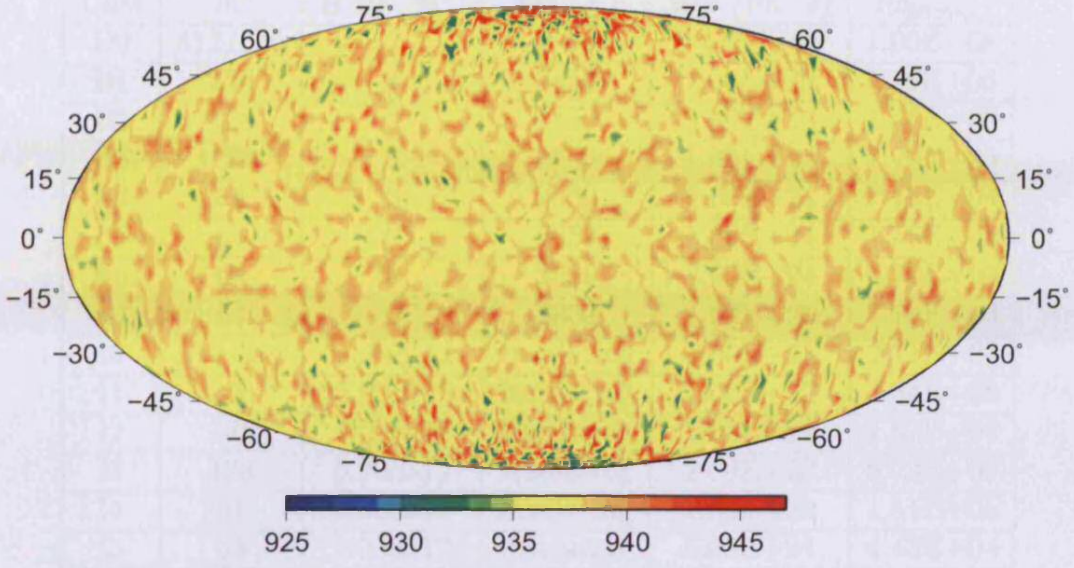


Figure 2.1: Temperature field in degrees Kelvin of the initial condition common to all modelled cases.

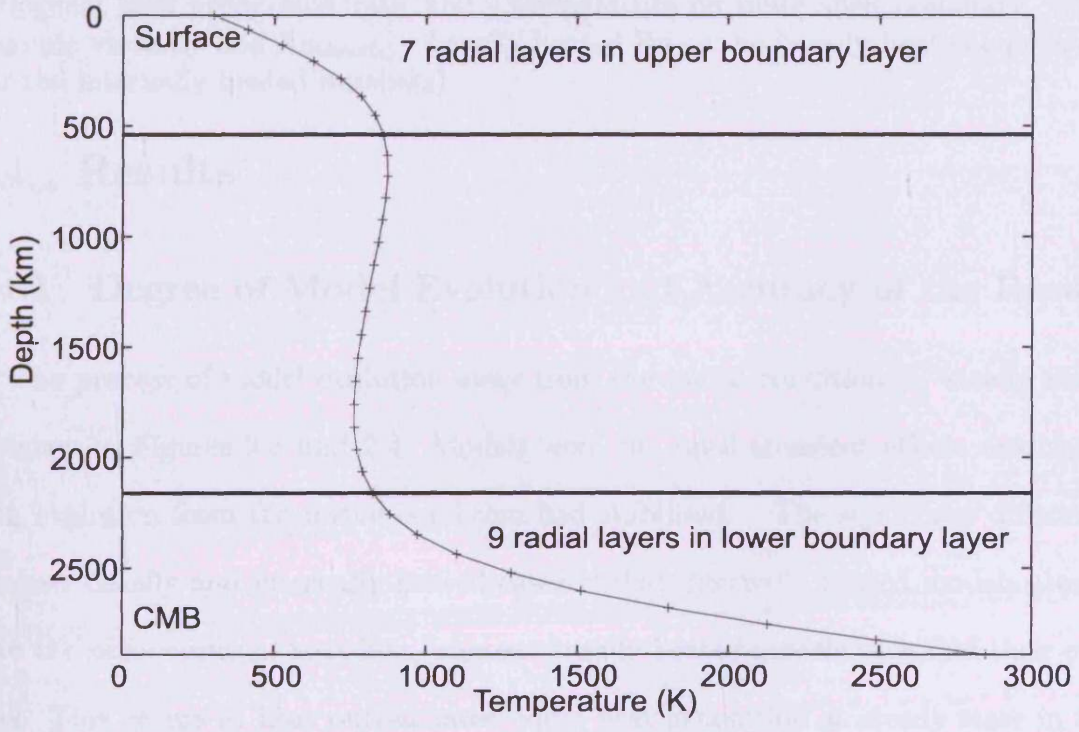


Figure 2.2: Radial temperature profile for Ra = 7.79×10^3, basally heated. Crosses represent layer boundaries. Case is well resolved.

Case	*mt*	*H* (W/m^3)	*tb2 (K)*	*visc* $(Pa \cdot s)$	$Ra_{bheat(i)}$
B0	512/256	0	2850	9.177E+20	1.00E+08
B1	256	0	2850	1.00E+22	7.79E+06
B2	128	0	2850	5.00E+22	1.56E+06
B3	128	0	2850	1.00E+23	7.79E+05
B4	64	0	2850	1.00E+24	7.79E+04
B5	64	0	2850	5.00E+24	1.56E+04
B6	64	0	2850	8.00E+24	9.74E+03
B7	64	0	2850	1.00E+25	7.79E+03
I0	512/256	5.00E-12	Insulating	1.00E+21	2.71E+07
I1	256	5.00E-12	Insulating	1.00E+22	4.84E+06
I2	128	5.00E-12	Insulating	5.00E+22	1.52E+06
I3	128	5.00E-12	Insulating	1.00E+23	8.73E+05
I4	64	5.00E-12	Insulating	1.00E+24	1.51E+05
I5	64	5.00E-12	Insulating	5.00E+24	4.43E+04
I6	64	5.00E-12	Insulating	1.00E+25	2.60E+04

Table 2.2: Values that were varied in each case. mt - model resolution, H - specific radiogenic heat production rate, tb2 - temperature on inner shell boundary, visc - dynamic viscosity and $Ra_{bheat(i)}$ - basally heated Ra or the basally heated equivalent (for the internally heated numbers).

2.4 Results

2.4.1 Degree of Model Evolution and Accuracy of the Result

The process of model evolution away from the initial condition to 'steady state' is shown in Figures 2.3 and 2.4. Models were run until transient effects associated with evolution from the initial condition had stabilised. The significant difference between basally and internally heated cases is that internally heated models always have the same eventual heat flux, whereas basally heated models each find their own level. This occurs as heat output must equal heat generation at steady state in the internally heated cases, otherwise energy is not conserved. Basally heated models by

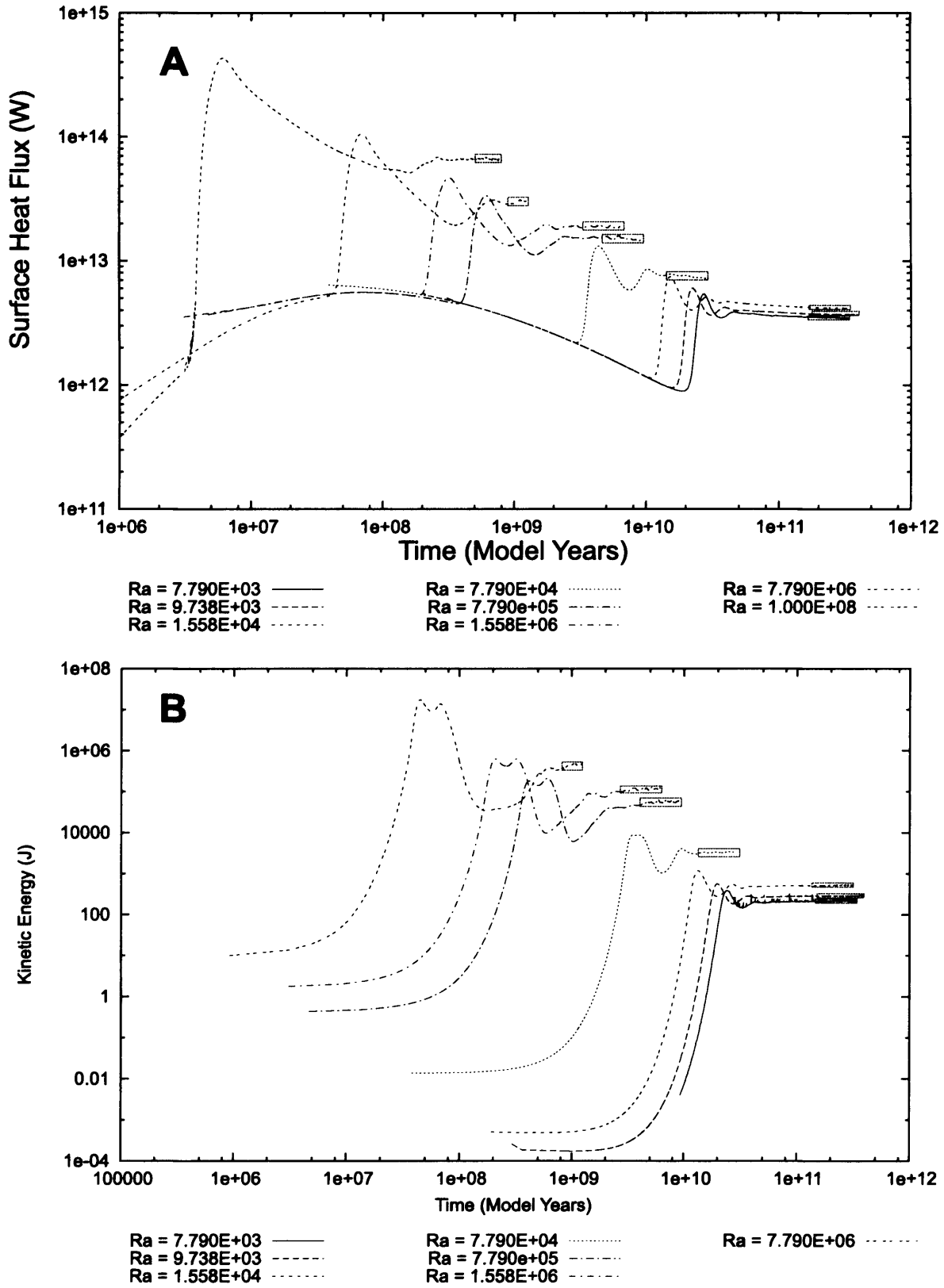


Figure 2.3: Summary of model evolution for the basally heated cases, boxes indicate the time interval over which the average Nu was calculated. A - surface heat flux, B - kinetic energy (note - does not include highest Ra run data - data unavailable) log axes.

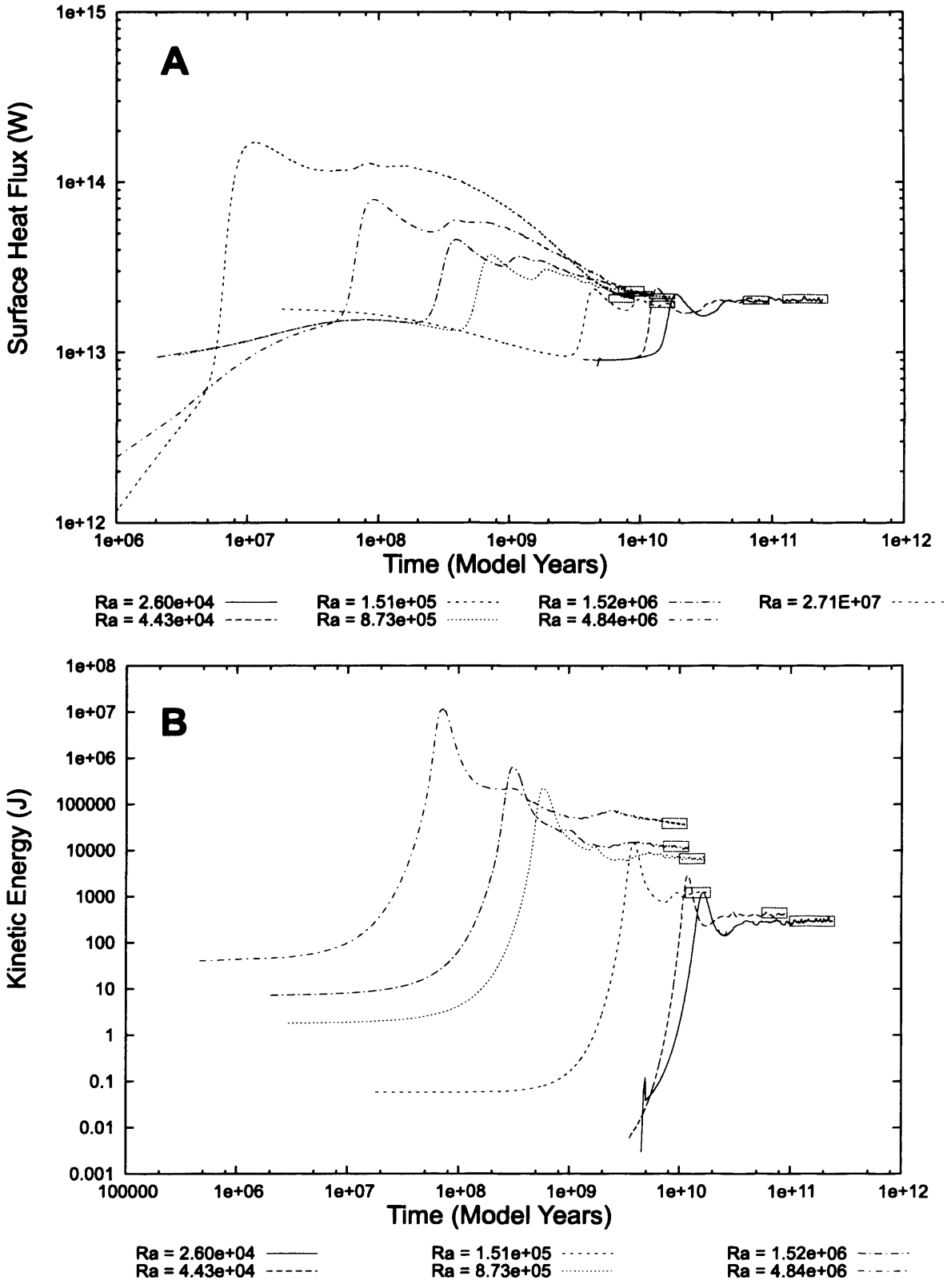


Figure 2.4: Summary of model evolution for internally heated cases, boxes indicate the time interval over which the average Nu was calculated. A - surface heat flux, B - kinetic energy (note - does not include highest Ra run data - data unavailable) log axes.

contrast need to balance their heat input at the base and loss at the surface. With the viscosity controlling the thickness of the thermal boundary layer, each basally heated case will have a different heat flux through the mantle.

The numerical error associated with the discretisation of the domain (resolution) was assessed by running high resolution cases with identical input values to lower resolution cases I2 and B3. Figure 2.5 compares the evolution of the original and higher resolution cases. The differences between the final Nu values was 0.3% for the basally heated case and 1.9% for the internally heated case. Average Nu values (over the last 1000 time steps) differed by 2.4% for basal heating and 2.3% for internal heating. This compares with a Nu time dependent variability of $\approx$3% for case B1, indicating that any error from the discretisation is of similar order to the intrinsic variability of the cases at high Ra. The overall evolution pattern and final Nu values are very similar for original and higher resolution runs, indicating that the cases are adequately resolved.

2.4.2 Nu(Ra) Relationship

A formula for Nu(Ra) was found using the output of 8 cases with 100% basal heating and 7 cases with 100% internal heating. A potential source of error is the Nu, which will be time dependent in cases that are not true steady state, a feature of most realistic models of the mantle. It can be seen from Figures 2.3 and 2.4 that the surface heat flux of a statistically stable model tends to fluctuate about an average value. To eliminate this short-term variability, the Nu was calculated by averaging

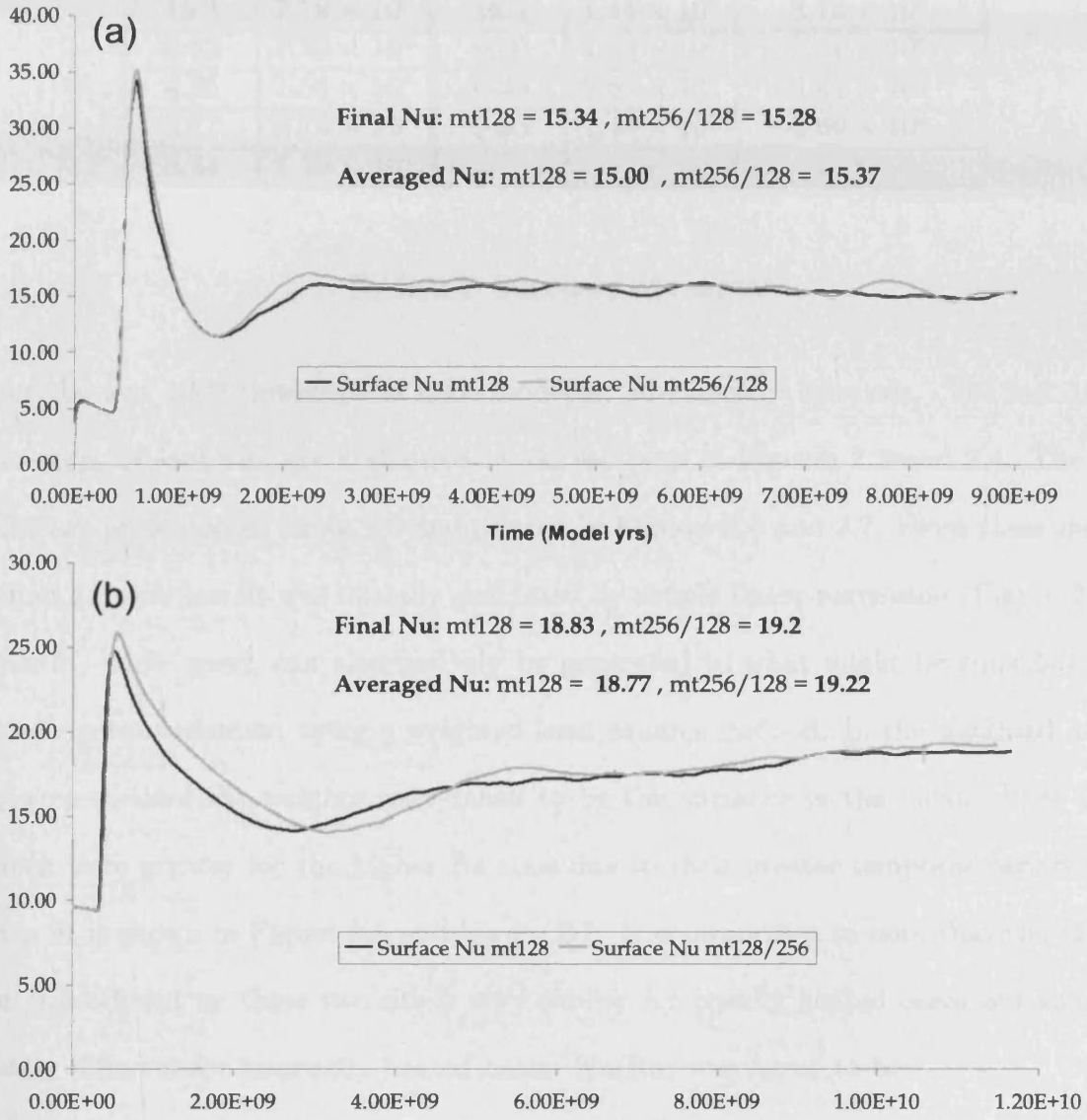


Figure 2.5: (a) Basally heated case B3 at mt128 and mt256/128. (b) Internally heated case I2 at mt128 and mt256/128.

Nu_{bheat}	Ra_{bheat}	Nu_{iheat}	Ra_{iheat}	Ra_{bheat} *(Equiv.)*
63.2	1×10^8	53.2	1.44×10^9	2.71×10^7
30.7	7.79×10^6	29.7	1.44×10^8	4.84×10^6
19.4	1.56×10^6	18.9	2.87×10^7	1.52×10^6
15.3	7.79×10^5	16.4	1.44×10^7	8.73×10^5
7.32	7.79×10^4	9.50	1.44×10^6	1.51×10^5
4.25	1.56×10^4	6.48	2.87×10^5	4.43×10^4
3.77	9.74×10^3	5.53	1.44×10^5	2.60×10^4
3.54	7.79×10^3	-	-	-

Table 2.3: Summary of results.

over the last 1000 timesteps of each model at 10 timestep intervals. The last 1000 timesteps of each run are higlighted by dotted boxs in Figures 2.3 and 2.4. The results are presented in Table 2.3 and plotted in Figures 2.6 and 2.7. From these mean values a power law fit was initially generated by simple linear regression (Figure 2.6). This fit, while good, can alternatively be generated in what might be considered a more rigorous manner, using a weighted least squares method. In the weighted least squares method the weights were taken to be the variance in the mean of the Nu, which were greater for the higher Ra runs due to their greater temporal variability. This fit is shown in Figure 2.6 and Figure 2.7. It is important to note that the value for β produced by these two fits is very similar for basally heated cases but significantly different for internally heated cases. Nu(Ra) was found to be:

$$Nu = 0.284 Ra_{bheat}^{0.294 \pm 0.004} \tag{2.19}$$

for the basally heated cases (Fig. 2.6) and:

$$Nu = 0.164 Ra_{bheat_i}^{0.337 \pm 0.009} \tag{2.20}$$

for internally heated cases (Fig. 2.7 using Ra_{bheat_i}) when simple linear regression was used to fit the data. The value for β obtained by weighted least squares fit are not used as the fit to the data is poor with that method, with a larger variance of residuals. When one considers the internally heated cases separately and plots the unchanged Rayleigh numbers (Ra_{iheat}) one gets $\beta \approx 0.24$, a value very close to the value obtained by Iwase & Honda (1997) for internally heated models.

The value of $\beta \approx 0.29$ for the basally heated case is at odds with the value of $\frac{1}{3}$ (0.33...) produced by boundary layer theory (Turcotte & Oxburgh 1967). The value of 0.337 for internally heated cases appears to agree, though we note that classical boundary layer theory was derived assuming basal heating. The error on β is however somewhat higher for the internally heated value than for the basally heated value and the weighted fit is notably different.

2.4.3 Surface velocity - Rayleigh number scaling

The difference between Nu(Ra) for basally and internally heated cases exists but is relatively small compared to the difference in the surface velocity. When the root mean square (RMS) surface velocity relationship to the Ra is plotted (Fig. 2.8), a given Ra demonstrates a lower RMS velocity for internally heated cases than for basally heated cases. The assumption implicit in TERRA is of no net rotation (rigid

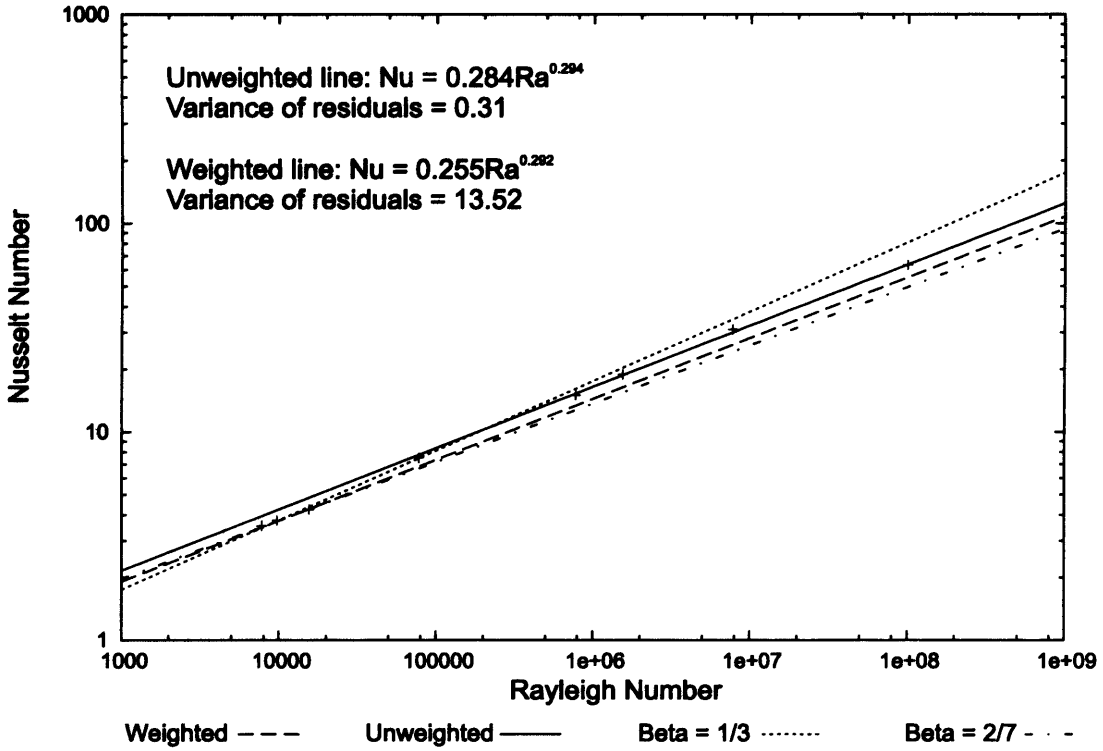


Figure 2.6: Two possible fits for output from the basally heated cases. Trends of $\frac{1}{3}$ and $\frac{2}{7}$ are also shown. A value of $\frac{1}{3}$ being the value predicted by simple boundary layer theory, and $\frac{2}{7}$ being the value suggested by Castaing et al. (1989) from their high Ra, but low Prandtl number experiments. We note that the mantle has a very high Prandtl number. The magnitude of quantifiable errors is within the dimensions of the data points.

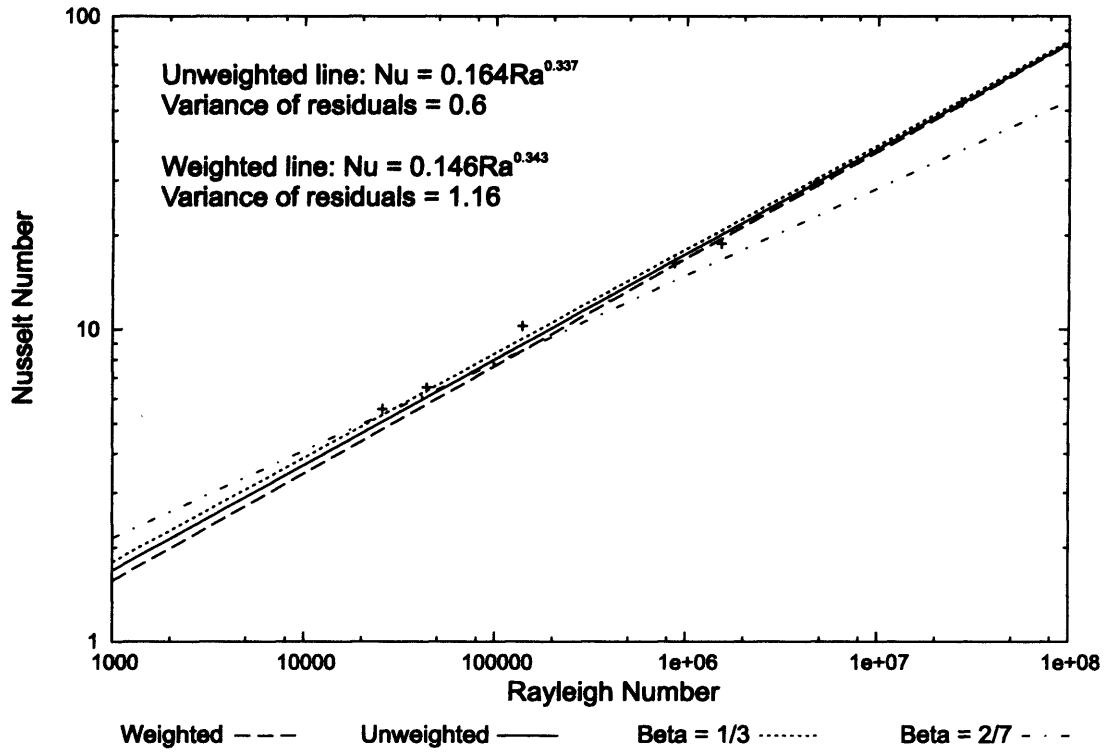


Figure 2.7: Two possible fits for output from the internally heated cases. Trends of $\frac{1}{3}$ and $\frac{2}{7}$ are also shown for comparison. As before, the magnitude of quantifiable errors is within the dimensions of the data points.

body rotation) of the surface.

The respective equations are:

$$RMSVel_{bheat} = 1 \times 10^{-13} Ra^{0.538} \tag{2.21}$$

$$RMSVel_{iheat} = 1 \times 7^{-14} Ra^{0.5} \tag{2.22}$$

In general, the results display a much wider scatter than the Nu(Ra). The cause of

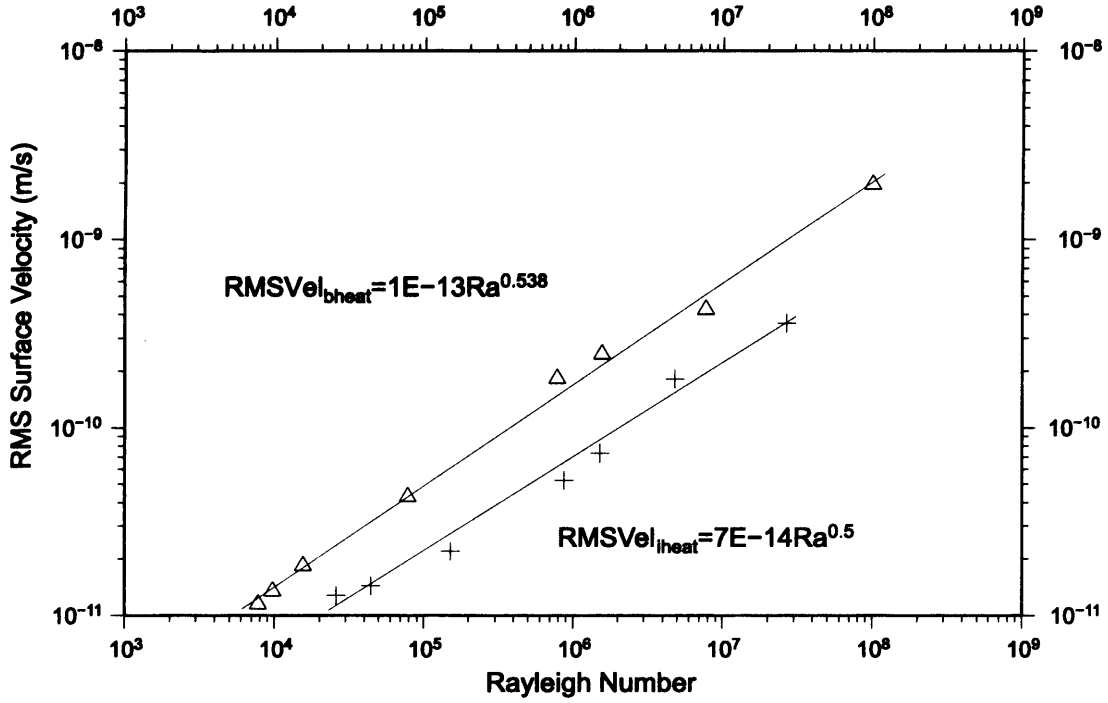


Figure 2.8: Scaling relationship of RMS surface velocity and Rayleigh number for internal and base heating (see also: Figs 2.3 and 2.4)

this is the less stable nature of the RMS surface velocity when compared to the surface heat flux (the key variable for the Nu). It is possible that some of the cases were not sufficiently evolved, however Figures 2.3 and 2.4 show that the surface velocity is much more variable than surface heat flux, even in cases that reach heat flux stability rapidly. It is therefore likely that at high vigour, velocity does not reach steady state

as readily as surface heat flux. As a result, interpretations of the surface velocity scaling relationship are likely to be less accurate than for Nu(Ra).

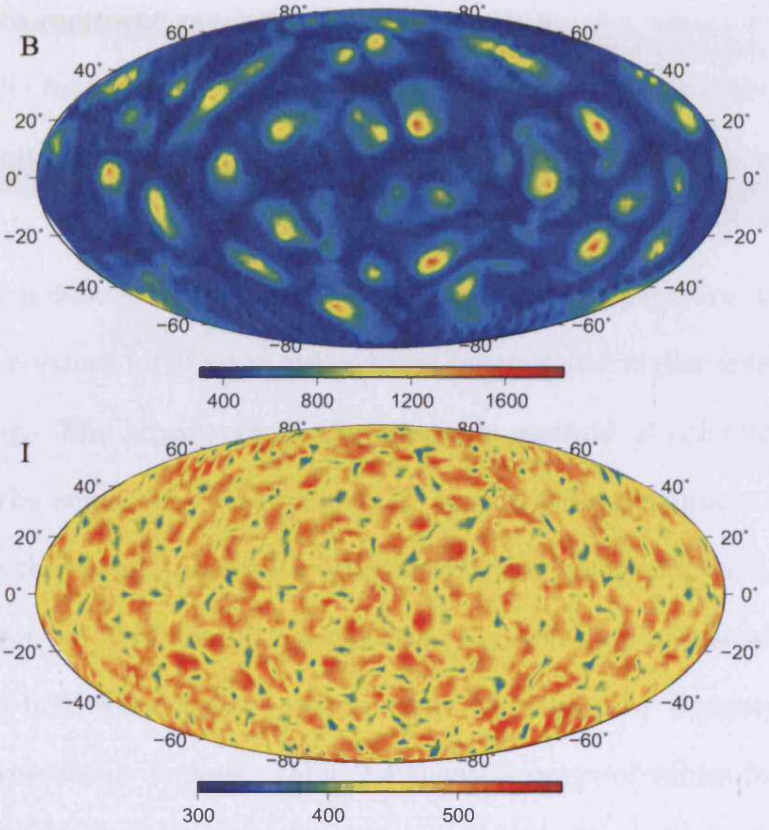


Figure 2.9: Comparison of temperature field planform at 22 km depth; basally heated model (B) and internally heated model (I), for cases with a dynamic viscosity of 1×10^{22} Pas

2.5 Previous Work

There are a number of factors, particularly relating to how the Ra and Nu are calculated, that need to be considered so previous work can be seen in the correct context. Ra and Nu are non-dimensional and are calculated from average values. For example, surface heat flux will not be the same at all points globally. Therefore, as

the Nu is calculated from model output, the nature of the model, be it 1D, 2D or 3D will have a bearing on the final value. One must also consider that not all Ra quoted in the literature represent precisely the same measure. For example, Ra for internally and basally heated models are by necessity calculated using different versions of equation 2.2 and cannot be directly compared (unless the basally heated equivalent Ra is calculated). In addition, some work involving variable viscosity by Christensen (1985) used a Ra defined using the viscosity at average temperature, which produces somewhat lower values for β when using basal heating and higher values when using internal heating. The advantage of Christensen's method of calculating the Ra is that it allows the equations to be defined for temperature dependent viscosity. The disadvantage is that comparison with other work is not straightforward. For example, Christensen obtained a value of 0.2 when using his definition of viscosity as opposed to the value of 0.28 when using the traditional definition of viscosity, the average viscosity over the entire domain. Table 2.4 shows a range of values for β calculated from a variety of model types. Work by Giannandrea & Christensen (1993) produced a value of 0.28 in large aspect ratio tank experiments. The aspect ratio of convection cells (λ) seems to have an impact in 2D models (Mckenzie et al. 1974, Olsen 1987), though it is not clear from boundary layer theory why this should be. At larger values of λ, β tends to 0.25 whereas where λ is lower, β is closer to 0.33. Therefore, it is likely that this effect was responsible for the value of 0.28 found by Giannandrea & Christensen (1993), with Ra up to 10^5. An arbitrary choice of aspect ratio is not an issue with 3D spherical geometry models, at least where whole mantle convection is considered. Models of the mantle in 3D spherical geometry (Bercovici et al. 1989,

Type	*Boundaries*	λ	*Heat*	β	*Source*
1D	Free Slip	1	B	$\frac{1}{3}$	(Turcotte & Oxburgh 1967)
2D	Rigid Lid	1	B	0.10	(Christensen 1984)
2D	Free Slip	var.	B	0.19-0.34	(Hansen & Ebel 1984)
2D	Free Slip	1	B	$\frac{1}{3}$	(Olsen 1987)
3Ds	Free Slip	-	B	0.280	(Bercovici et al. 1989)
2D	Plate v.	1	B	0.3	(Gurnis 1989)
Tank	Free Slip	large	B	0.28	(Giannandrea & Christensen 1993)
3Ds	Free Slip	-	B	0.26	(Ratcliff et al. 1996)
3Ds	Not stated	-	I	0.24 (LRa)	(Iwase & Honda 1997)
2D	Free slip	1	B	0.3	(Korenaga 2003)
3Ds	Free slip	-	**B / I**	**0.29/0.33**	**This Study**

Table 2.4: Previously calculated/measured Nu(Ra) relationships. 1D, 2D 3D, 3Ds (spherical geometry) refer to the spatial nature of the model. λ represents aspect ratio, not applicable to spherical models. L Ra indicates local Ra. B indicates basal heating, I indicates internal heating. Note: β is quoted as in original papers, the values are not necessarily comparable in their method of calculation. However, all are intended for use in parameterised thermal history models.

1992) have produced a value for β of around 0.28. However, due to technical limitations at the time, the models were only able to cover 2 orders of magnitude of Ra. This limit reduces the confidence level for the value of β calculated when one wishes to extrapolate to higher (Earth like) Ra.

The work of Korenaga (2003, 2005) examines the effect of chemically controlled viscosity in the mantle. Korenaga proposes that the dehydration of the upper mantle caused by melt generation results in the upper mantle becoming less deformable. Logically, higher mantle temperatures in the past would be expected to result in more melting and a greater stiffening effect. Thus Korenaga proposes that a hotter Archean mantle would in fact have a lower effective Rayleigh number than has been assumed.

This effect would have implications for the value of β as it would have to be adjusted to accommodate this stiffening effect. Korenaga (2003) suggests that β may be very low or even negative when this effect is taken into account. Without this effect, the value for β produced by Korenaga's models is 0.3.

2.6 Discussion

Nu(Ra) derived from boundary layer theory of thermal convection gives a β value of $\frac{1}{3}$. It is therefore important to consider why the modelling derived value for β was found to be closer to 0.29. Previous studies have found a variety of values for β, depending on their modelling and measuring methods, but have either concluded that $\frac{1}{3}$ is a point which their models 'will' trend to at higher Ra (Iwase & Honda 1997), or that the value of β is entirely different due to particular boundary conditions (Christensen 1984). Iwase & Honda (1997) used 3D spherical geometry models to obtain a value closer to $\frac{1}{4}$. They argued that at higher Rayleigh numbers their β would tend to $\frac{1}{3}$, an assumption somewhat at odds with high Rayleigh number experimental work such as Castaing et al. (1989) and indeed with this study. A significant difference in their work is that Iwase & Honda calculated their Nu and Ra locally, across the thermal boundary layer, as opposed to globally as is traditionally the case. Therefore, their value for β is based on somewhat different assumptions to the classical case and may be more sensitive to variations in planform, possibly similar to the differences in Ra-RMS velocity scaling of different heating modes seen here (Section 2.4.3) and reported previously in 2D models (Mckenzie et al. 1974). Bercovici et al. (1989, 1992)

have proposed a value for β of around 0.28. Their work, in 3D spherical geometry, demonstrates tetrahedral and cubic spherical harmonic patterns of convection at relatively low Ra. They note differing values of β for the different patterns. For example, β for their cubic cases is given as 0.28, while tetrahedral solutions give 0.26. This indicates that at low vigour the convective pattern (several of which are stable) has a significant influence over β. The models used in this study explored much larger Ra, however the values for β of Bercovici et al. (1989, 1992) should still be considered valid for low vigour convection.

In 2D Cartesian geometries, the aspect ratio and boundary conditions appear to have a significant influence (Mckenzie et al. 1974), resulting in lower values of β when aspect ratio is large and values closer to $\frac{1}{3}$ as aspect ratio approaches 1:1 (Olsen 1987). When 3D spherical geometry is used this source of variation is eliminated as the aspect ratio that convection cells can adopt is constrained by Earth-like parameters of mantle thickness and surface area ratio between the CMB and the surface. In addition, 3D models appear to be less sensitive to aspect ratio due to the extra dimension of freedom (Tackley et al. 1993, although in a slightly different context). In terms of the upper surface boundary velocity condition, this study used a free slip condition and the value for β is consistent with that of Gurnis (1989).

An additional factor that may affect β is the type of convection. Castaing et al. (1989) divide convection into soft and hard turbulent regimes, with the transition at $\mathrm{Ra} = 4 \times 10^7$. This value puts Earth's mantle in the domain of this transition. Their

experiments, which extended to extremely high vigour (Ra = 10^{12}) using liquid Helium, give $\beta = \frac{2}{7}$ (0.2857...). While this type of experiment is obviously not intended to examine mantle behaviour (the Prandtl number is over 20 orders of magnitude too large), it does suggest that real convecting systems are capable of displaying boundary layer behaviour at odds with conventional boundary layer theory, which has the tendency to reduce the value of β. The type of behaviour suggested to cause this is entrainment from the boundary layer, possibly as fingers of material. In this study, we are confident that our values are in fact closer to 0.29 than $\frac{1}{3}$. Therefore, we suggest a certain amount of non-classic boundary layer behaviour is occurring. The broad range and high magnitude of Ra modelled gives further weight to the result that $\beta = 0.29$.

This study finds no evidence of a change in β at high Ra, as illustrated in Figs. 2.3 and 2.4. This indicates that for a thermally driven system, a variable β cannot be used to moderate the long term thermal evolution of Earth when predicting backwards from the present.

A feature of Fig. 2.6 is that if the trend is extended to Nu = 1 (the onset of convection), our predicted critical Ra would be of order 10^2. The critical Ra of the mantle has been extensively investigated and is accepted to be of order 10^3. Actual values quoted are: $Ra_c = 658$ (Cartesian geometry), $Ra_c = 902$ for a spherical shell (both from Busse 1989) and 712 (Bercovici et al. 1989). Previous work shows a similar trend, although as the critical Ra is approached there is a distinct steepening of the

Nu(Ra) curve as the Ra tends to the critical value (Bercovici et al. 1992). This apparent inconsistency is explained by the fact that at Nu = 1 the system is on the point of convection and therefore the whole mantle can be considered as a single boundary layer. Boundary layer theory of convection requires upper and lower thermal boundary layers. Thus a Nu(Ra) power law is not particularly applicable in such a situation.

The power law scaling of RMS surface velocity as a function of Ra is of order $v_{RMS} \approx Ra^{0.5}$, agreeing broadly with values derived from early 2D modelling (Mckenzie et al. 1974). The scaling also shows significant differences between basally and internally heated cases (Figure 2.8). Internally heated cases consistently have lower velocities at any given Ra. As the surface heat flux of such cases is similar, the tendency must be caused by some difference in the structure of convection. Figure 2.9 demonstrates how the planform of the two heating modes differs. The characteristic lateral length of individual convection 'cells' for internally heated cases is approximately half that of basally heated cases. This has been observed in 2D models (Mckenzie et al. 1974) It is possible that the difference is due to the manner in which heat is introduced into the systems.

With internal heating, heat is generated evenly throughout the mantle. The base of the system in a 'steady state' case adopts a particular temperature with no thermal boundary layer at the core mantle boundary (CMB). The only significant thermal boundary layer is at the surface. Plume-like features that develop are much smaller and tend not to arise from the base of the system initially. Cold material sinking

from the surface to the CMB provides the dominant large-scale driving force. Thus a purely internally heated system has its own evolutionary pattern with the heat lost from the surface replenished evenly throughout the volume of the mantle. The overall effect of this internally heated mechanism is to produce shorter length scale convection.

With basal heat input, a much larger scale and generally more organised structure develops based around the upper and lower thermal boundary layers, as heat is input solely at the CMB and cooling allowed at the surface. Plumes are able to arise from boundary layer instabilities at the CMB and cold downwellings are in effect reverse cold plumes formed as the upper boundary layer becomes unstable. Therefore, convection is on a much larger scale than an equivalent internally heated case. As a result material has further to travel between upwelling and downwelling for the same convective vigour, producing higher surface velocities.

The fact the β appears to be larger for internally heated cases could also be explained by the location of heat input. A basally heated system must transfer a given 'parcel' of heat from the bottom to the top of the mantle, whereas for internal heating, heat is generated throughout the volume and the majority of heat 'parcels' have less distance to travel before reaching the surface. This process appears to scale better with increasing Ra than basal heating, leading to a higher value for β. That this β is close to the $\frac{1}{3}$ predicted by boundary layer theory, may be an effect of the aspect ratio of the convection cells, as opposed to that of the domain, which remains constant.

The generally lower aspect ratios of the internally heated convection cells appear to be closer to the 1:1 aspect ratio assumed by boundary layer theory and shown to give $\beta \approx 0.33$ in 2D modelling (Olsen 1987).

Zhong (2005) used CITCOM, a 3D Cartesian box model with 100% basal heating and using the same definition of basally heated Ra as this study, found that plume number scales at $Ra^{0.31}$. The work presented here agrees broadly with Zhong (2005) as large scale plumes, originating at the lower thermal boundary layer, are likely to be the primary transporters of heat, thus controlling surface heat flux in a 100% basally heated system. Internal heating by contrast, tends to produce a similar heat flux but over a broader area so plume number would not be expected to scale in the same manner. This mechanism could be a further candidate for the difference seen in scaling relationship between the RMS surface velocity and Ra for the two different heat sources.

The models presented here assumed free slip upper and lower boundaries. While free slip is a very good approximation of the CMB, Earth's surface boundary layer is composed of rigid plates. In addition to the plate boundary condition, the Earth also has phase transitions at 410 km, 660 km and D''. These features combined with other aspects such as: temperature and depth dependent viscosity, non uniform internal heating and heterogeneous composition may impose limits on the applicability of our value for β. We are however confident that we have characterised the scaling of Nu with Ra using a model that is more Earth-like (in terms of convective vigour and

geometry) than those used in comparable previous studies, where a simplified model has been shown to capture the essence of the more complex real Earth.

Analysis is an important and often under emphasised aspect of modelling. Many of the more realisic models are not truly steady state and may be time dependent. For example, the Nu may oscillate, even when a model is considered sufficiently evolved. The question is which value for the Nu should be used to define the power law and how such variability should be accounted for. In this study, two different power law fits were calculated, one using simple linear regression, the other by what might be considered a more rigorous weighted least squares method. The quality of the fit is illustrated by the variance of the residuals (or reduced χ^2). Interestingly, the quality of the fit for the unweighted data is excellent, while the weighted fit is poor. The weighted fit is dominated by the low Ra values, where the Nu is steady, and virtually ignores the higher Ra values, where the behaviour is not steady. A weighted fit which ignores the lowest 3 Ra cases (Figure 2.10) provides much closer agreement with the unweighted plot and agrees more closely as to a value for β (0.290 unweighted and 0.309 weighted. The variance of residuals is also less for such a weighted fit. The reason for this appears to be that the 3 lowest Ra cases with basal heating appear to lie on a slightly lower trend than the rest of the data points (see Figure 2.6). This is a consequence of the large scale cubic/tetrahedral convective patterns that low Ra cases tend to adopt (e.g. Figure 2.11), which are generally less efficient (in a heat transfer sense) than more short scale and complex patterns. This has been observed previously for low Ra spherical geometry cases (Bercovici et al. 1989, 1992). The dis-

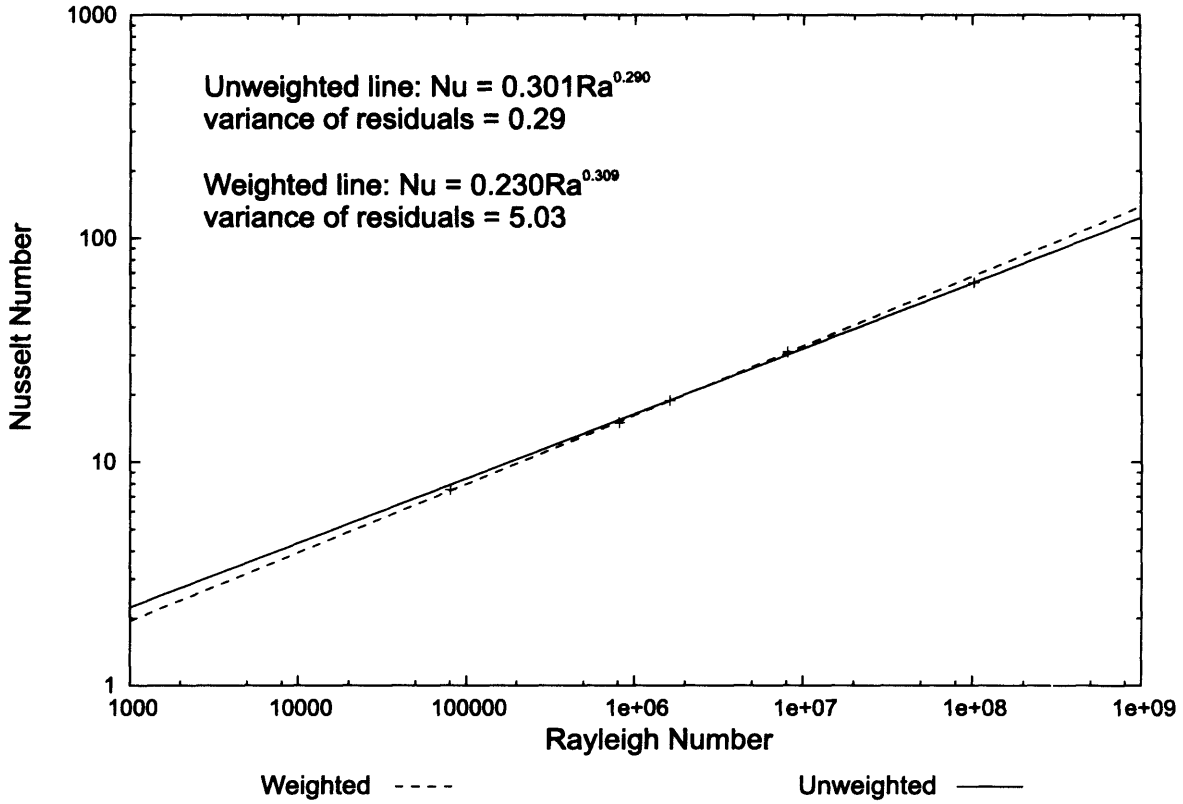


Figure 2.10: An improved fit is achieved when the lowest 3 Rayleigh number cases are excluded from the analysis. Basally heated cases only.

cussion above highlights the different values of beta that can emerge from the same initial data. This study prefers the unweighted fit of the whole data to define β.

The implication of $\beta = 0.29$ as opposed to $\frac{1}{3}$ is that the extrapolated surface heat flux in the past (at higher Ra) would be lower than previously estimated. This effect would arise from the efficiency of the convective system, before other factors, such as rheology, are taken into account.

Nu may be calculated using $Nu = aRa^{\beta}$. Our preferred Nu for $Ra_{bheat} = 10^9$ would be 116, when: a is 0.284 (Equation 2.19) and β is 0.29. The surface heat flux (SHF) may

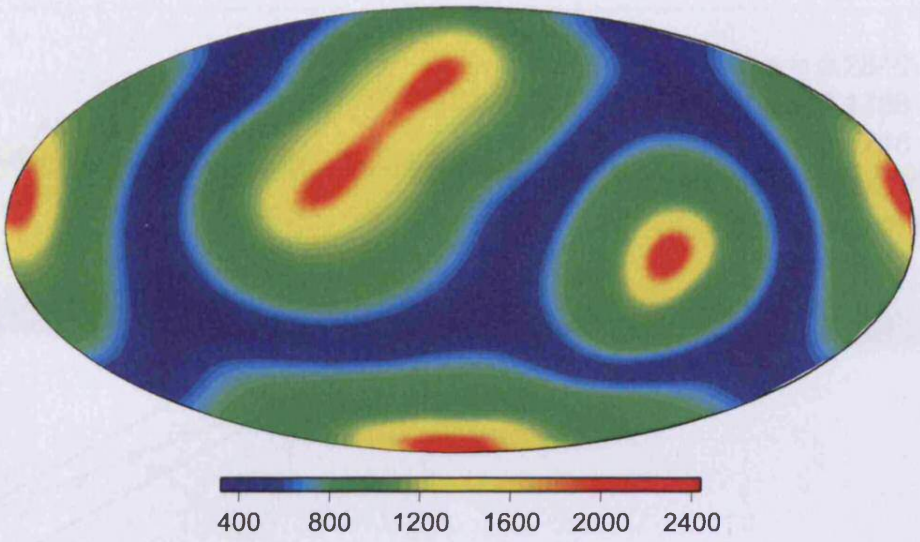


Figure 2.11: Case B7, 900 km depth. Temperature plot in Kelvin. Note tetrahedral planform

be calculated from the Nu using the following equation re-arranged and simplified from Equation 2.15:

$$SHF = Nu \times 9.832 \times 10^{11} \tag{2.23}$$

To calculate Nu for different values of β (e.g. $\frac{1}{3}$) another value for a must be chosen to prevent the two trends from being pinned at the origin. This would lead to an artificially large Nu at high Ra. In order to obtain reasonable projection, a was calculated for the mid point of the data on Figure 2.6 and also for the lowest value Ra in the data. These two values effectively pin the projected values to points within the data. These two possibilities for projecting the power law are plotted along with the original a in sketch form in Figure 2.12. The calculated Nu are shown in Table 2.5, the favoured value is show in bold and shows a 32% reduction in the Nu when $\beta = 0.29$ as opposed to $\frac{1}{3}$. These results are particularly significant when one considers that this

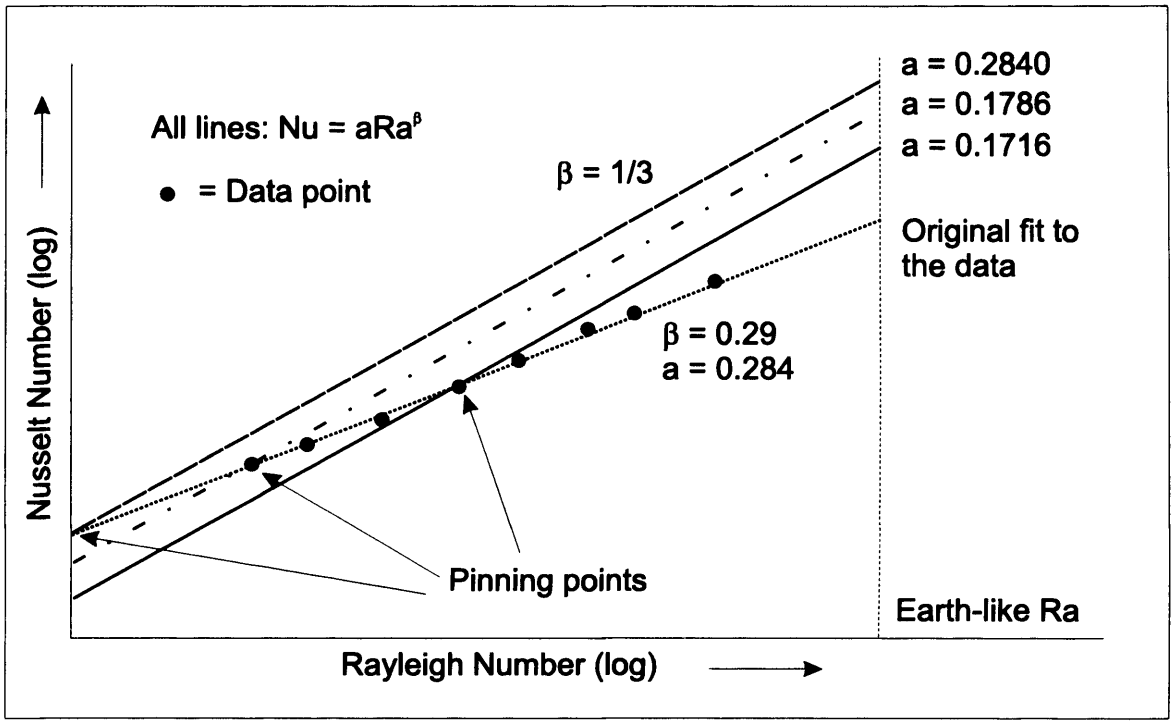


Figure 2.12: Sketch representation of how the chosen value for a effects the Nu when extrapolating to higher Ra numbers

is the most conservative choice of a when generating the $\beta = \frac{1}{3}$ relationship. When extrapolating to early Earth, where the Ra is even higher, the difference between the results predicted by the two values of β would be greater still.

a	*Nu* ($\beta = \frac{1}{3}$)	*SHF (TW)*	*Nu* ($\beta = 0.29$)	*% reduction in Nu for* $\beta = 0.29$
0.2840	214	210	116	46
0.1786	179	176	116	35
0.1716	**172**	**169**	**116**	**32**

Table 2.5: Resulting Nu for Ra=10^9 with different values of a are shown for $\beta = \frac{1}{3}$. Note that for $\beta = 0.29$, a remains 0.284. SHF is surface heat flux.

It can be seen that the Earth's ancient heat flux could be significantly lower than conventional boundary layer theory would suggest for a relatively small change in

β. This perhaps offers a mechanism that could help to partially resolve the thermal catastrophe paradox which occurs when projecting the Earth's temperature into the distant geological past as described by Korenaga (2003). As can be seen, at higher Ra the difference between Nu extrapolations becomes very significant.

Korenaga (2003, 2005) and others use β in their parameterised thermal models, the value for β being derived from a uniform viscosity model as in this study. They then however, use the scaling relationship in parameterised models where viscosity is temperature dependent. This is a slight clash of concept as the β being used is not defined for such a case. In practice, this does not appear to introduce significant error as β appears to be robust in the face of whatever parameters are in use (Korenaga 2003, Korenaga & Jordan 2002). This again highlights the need for β to be very well defined.

2.7 Conclusion

The Nu(Ra) scaling relationship has been investigated in 3D spherical geometry over a wide range of convective vigour, using both internally and basally heated models. Basally and internally heated systems were directly compared by converting the internally heated Ra to a basally heated equivalent Ra, achieved by dividing the internally heated Ra by the Nu (Equation 2.18). It was found that β is closer to 0.29 than $\frac{1}{3}$, the value suggested by simple boundary layer theory. With $\beta = 0.29$ as opposed to $\frac{1}{3}$, the Nu and surface heat flux are reduced by 32% and 55 TW respectively at Ra $= 10^9$. Diffuse heat input in internally heated models and the smaller

aspect ratio of the resulting convection cells, appears to be responsible for β being closer to $\frac{1}{3}$ ($\beta = 0.337$) for such cases. There is no evidence that β is reduced at high Ra for Earth's mantle. Other mechanisms are therefore required to resolve the thermal catastrophe paradox encountered in parameterised thermal evoloution models. The relationship between average surface velocity and Ra is different in basally and internally heated systems. This is best explained by the difference in characteristic length of convection features. Internally heated cases have a characteristic length between upwelling and downwelling regions approximately half that of basally heated cases. This highlights the influence planform may have over certain aspects of mantle behaviour.

Chapter 3

The Influence of Convective Vigour on Phase Change Induced Layering at 660 km Depth in Early Earth's Mantle

3.1 Abstract

Since its discovery, the endothermic phase change in olivine mineralogy from ringwoodite to ferropericlase and perovskite at 660 km depth (660) has been a candidate for inducing some degree of layering in Earth's mantle. A layered mantle would provide a mechanism to isolate portions of the mantle chemically. Layering could also reduce the secular cooling rate relative to whole mantle convection. Thus layering would resolve a number of paradoxes in the understanding of the deep Earth. The problem with this

interpretation is that seismology provides a convincing case for whole mantle convection at present. However, it is widely accepted that the Earth as a whole was hotter in the past and that the mantle had a higher Rayleigh number (Ra) as a result. It has been suggested in previous modelling work that for a given Clapeyron slope at 660, higher Ra increases the propensity to layering. The conceptual model for this suggests that higher Ra leads to thinner plume-like upwellings, which are then of insufficient strength to break through the restoring force provided by the negative Clapeyron slope of the phase change. The suggestion has been made that at ancient Ra, the mantle may have been layered about 660 and that this layering has at some point broken down.

This phenomenon was investigated through an ensemble run of 3D spherical mantle models. A parameter space mapping of 660 Clapeyron slope negativity (ranging from -20 to -2 MPa/K) and basal Ra (ranging from 1.56E+03 to 7.79E+06) was examined. Results suggest that the mantle may have been layered in the past. However, layering is unlikely to have endured for a significant time, requiring a further mechanism to preserve geochemical heterogeneities to the present day.

3.2 Introduction

Isotope geochemistry suggests that there is or has been a mechanism, which is able to isolate portions of the Earth's mantle over a significant period of time. To date, seismology has been unable to provide any direct evidence of either a layered mantle or a significant compositionally different region of the required size (Van der Hilst et al. 1997). The assumption is therefore that the mantle is convecting as a

single layer. This leaves many questions unanswered. For example, the observed geochemistry is not easily explained. Whole-mantle convection scaling laws (Chapter 2), when used to project past temperatures, seem to require an unreasonably hot early Earth in order to balance present heat fluxes (Korenaga 2005).

The 660 km phase change of ringwoodite to perovskite and ferropericlase (660) has a negative Clapeyron slope of between -2 and -3 MPa/K (Ito & Takahashi 1989, Ito et al. 1990, Davies 1999). This acts to provide a restoring force, which may inhibit vertical mass transfer. At present, this Clapeyron slope is not thought to be sufficiently large to prevent whole mantle convection. However, it has been shown that two factors can affect 660's propensity to layer; the steepness of the Clapeyron slope and the vigour of convection. Convective vigour is strongly linked to heat input and can be quantified by the Rayleigh number (Ra). The more negative the Clapeyron slope and the higher the Ra, the more likely layering is to occur (Machetel & Humler 2003). If one makes the reasonable assumption that the mantle was hotter in the past it is possible that layered convection may have been dominant and at some point broken down. This possibility has been suggested by a number of investigators (comment by Davies in Peltier 1996, Allègre 2002, 1997, Honda 1995). Early modelling work on mantle layering about 660 was carried out by Christensen & Yuen (1985) and re-examined by Machetel et al. (1995, and references therein). In much of this early work, particularly the 3-D studies, the computational resources limited the resolution of the models and the number of cases that could be examined.

This study investigates the influence of the 660 phase change on mantle convection in spherical geometry at high Ra. This was achieved by exploring a plausible parameter space of 660 km Clapeyron slope negativity and Ra; an ensemble of 43 model cases was carried out. Only limited extrapolation is required in Ra when projecting to the real Earth.

3.3 Previous Work

3.3.1 Why Layering?

Geochemistry offers a range of very powerful methods for investigating Earth history. Radioactive isotope systematics are most commonly used to investigate the mantle due to the identical chemical behaviour of different isotopes of the same element and the known relationships between various stable and parent-daughter isotopes.

Argon 40 is produced by the decay of potassium 40 with a half-life of 1.3 Gyr. As Argon is too heavy an atom to leave the Earth's atmosphere it is reasonable to assume that any ^{40}Ar produced has remained within the Earth system, the volume of ^{40}Ar being controlled by the amount of ^{40}K originally in the Earth. However, the amount of ^{40}Ar observed is significantly less that that which should be produced (Allègre et al. 1996). Possible explanations for this include: a significant amount of K is isolated in the core, the assumption that bulk Earth composition is the same as chondritic composition is false or there is a mechanism retaining ^{40}Ar in the mantle. The latter

explanation has been favoured and used to support the existence of a 'primitive' layer in the lower mantle. It is worth noting that the precise compatibility of Ar in mantle minerals may be in some doubt (Watson et al. 2007), potentially making Ar a much poorer measure of degassing. However, ^{4}He also appears to provide evidence for a deep isolated layer. ^{4}He is produced by the decay of thorium and uranium, at the same time as these elements produce radiogenic heat. There is a fixed relationship between the amount of heat and helium produced by these elements. It has been observed that this ratio is not maintained at the surface (O'Nions & Oxburgh 1983, Van Keken et al. 2001). This indicates that there is a mechanism retaining helium within the mantle, again often attributed to the existence of an isolated layer at depth. Additionally, the difference in helium isotopic ratios (^{4}He/^{3}He) between mid ocean ridge basalt (MORB) and ocean island basalt (OIB) has also been attributed to the existence of a deeper more primitive region that OIB samples but MORB does not (Kurz et al. 1982, Allègre et al. 1983). Neon is another element that has been used to argue for the existence of an isolated reservoir. Ne has three isotopes (^{20}Ne, ^{21}Ne and ^{22}Ne), thus atmospheric contamination can be accounted for. When plotted, the ^{21}Ne/^{22}Ne ratio shows distinctly separate distributions for MORB and OIB (Allègre 2002).

From these and other data the conclusion was drawn that the mantle could be either harbouring some kind of hidden non-interacting geochemical reservoir or convecting as two or more separate layers. We consider the latter case here. From thermal transfer theory, layered convection would result in at least one internal thermal boundary

layer, which could act as a source for plumes and could modulate certain aspects of thermal evolution.

The concept of past layered-mantle convection is an evolution of the original 2 layer model for the mantle. A revision was necessary as improved seismic observations such as those of Montelli et al. (2004, 2006) and Van der Hilst et al. (1997) show plumes rising from the core-mantle boundary and slab penetration into the lower mantle. In addition, strictly layered mantle models (Oldham & Davies 2004) indicate that a present day layered mantle would be visible to seismological studies. Other studies reinforce the conclusion that whole mantle convection is the best fit for current tomographic data (Simmons et al. 2006). The difficulty with the whole mantle convection model is that it does not seem to provide mechanisms to explain either the thermal or geochemical history. Thus a model in which layered convection has taken place in the past, undergoing breakdown before the present becomes attractive. This scenario would leave a 'memory' of layering in the isotope geochemistry. Allègre (2002) proposed that layered to non-layered transition occurred around 500 Ma. The difficulty with this hypothesis is that many of the factors involved are not well defined and constants in one study are often variables in another. For example, one study may employ variable viscosity, while another might fix the viscosity and vary another factor such as radiogenic heat production rate.

Allègre (2002) assumes that, as MORB and OIB source regions have significantly different helium and neon isotope geochemistry, there must have been a deep reser-

voir of some description. This may be somewhat more complex as it has been suggested that plumes may not be the sole source of volatiles in the upper region of the mantle (Ballentine et al. 2005). However, some mechanism is still required to provide the wide variety of end-members seen contributing to MORB and OIB. The essence of Allègre's model is that true 2 layer convection occurred in the past and then broke down. Simple layer breakdown and mixing is expected to rapidly destroy heterogeneity produced by past layering (Coltice & Schmalzl 2006), therefore a more complex system of present day 1 layer convection is required. The conceptual model (Figure 3.1) is composed of a series of partial layers with specific residence times. Three partial layers are identified by Allègre, one below 660 km, the region above 410 km and an intermediate layer (660-410 km) caused by higher viscosity and an upper layer above 410, though it is unclear why 410 should act as a barrier. Mixing times for material in these different layers is 50-360 Ma for the upper layer and 1200-1500 Ma for the intermediate high viscosity layer. This would allow for MORB source to be well mixed while OIB source can be of varied but different composition. In this model, lower mantle material is entrained through 660, gradually enriching the layers above. Allègre suggests that this sort of mixing can explain the various geochemical observations. Tackley et al. (1993) applied modelling and concluded that in 3-D geometry layering is 'leaky' layering at best. Christensen & Yuen (1985) modelled Clapeyron slopes for 660 of -6 MPa/K and also described leaky 2 layer convection, which they suggest may have been stronger in the past. Such a convection mode could allow some more buoyant slabs to reside at 660 and may play a role in the dynamics and geochemistry of continental break-up, crust formation and large igneous

provinces (Allègre & Turcotte 1986, Kellogg et al. 1999, Elliott et al. 2006). These interpretations are, for the most part, compatible with the model proposed by Allègre.

In such a system, some subducted slabs would be expected to either remain at 660 for a significant time, possibly acting as a source of shallow plumes or as an insulating layer, reducing the rate of secular cooling. A reduction in the secular cooling rate would be useful to avoid the thermal catastrophe paradox in parameterised thermal evolution modelling. However, it appears that the actual value of the Clapeyron slope for 660 is less negative than has been assumed for many studies in which layered convection has been modelled (Christensen & Yuen 1985). The actual value is thought to lie between -2 MPa/K and -3 MPa/K (Ito & Takahashi 1989, Ito et al. 1990, Davies 1999, Steinberger 2007).

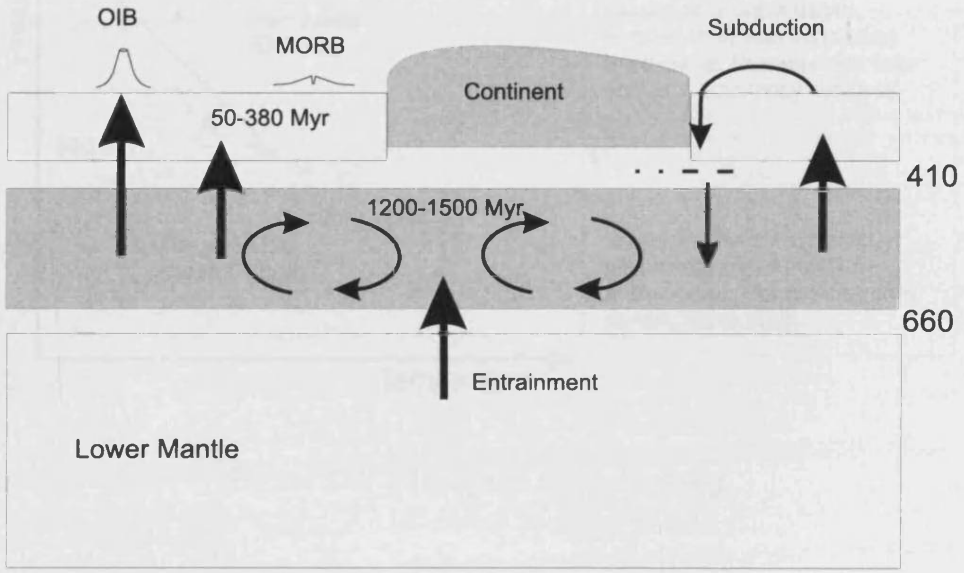


Figure 3.1: Model of a partially layered system, after Allègre (2002).

3.3.2 Mechanisms of Phase Change Layering

The potential for phase change layering, as opposed to compositional layering, has been framed as an interaction between vigour of convection (Ra) and the 660 km Clapeyron slope (Peltier 1996, Allègre 2002). A 2D parameter space consisting of a range of both these variables should define 2 areas, one where the system is layered, another where is is not.

A negative (endothermic) phase change Clapeyron slope can favour layering in the mantle by inhibiting material transfer. Figure 3.2 is a sketch phase diagram demonstrating how a negative Clapeyron slope applies a correcting force to perturbations from both above and below. When combined with a negative Clapeyron slope phase

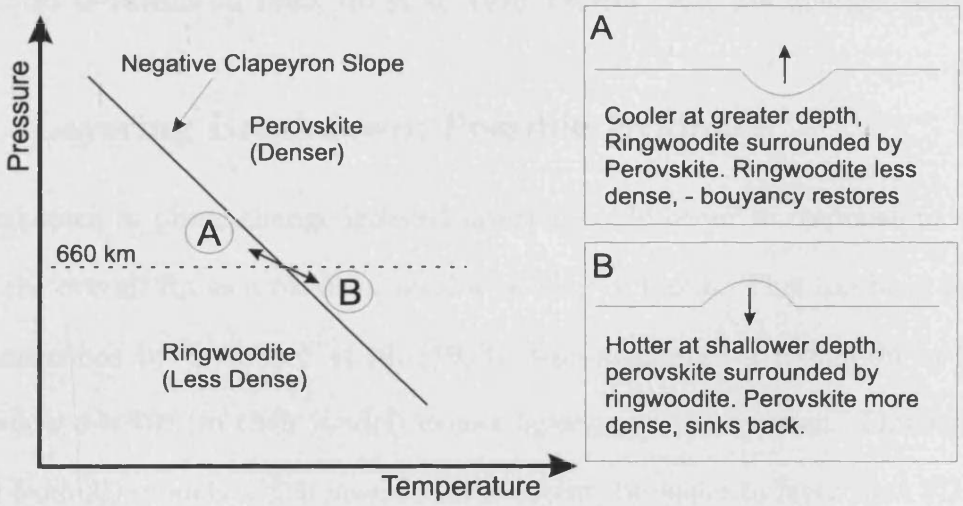


Figure 3.2: Clapeyron slope layering.

change, Ra effects layering in the following manner. At high vigour, plumes are numerous and thin such that they are unable to cause sufficient perturbations to break

through a given Clapeyron slope at 660 km. At lower vigour, plumes are much larger and are therefore able to cause much greater perturbations in the 660 discontinuity and break through, thus layering cannot be sustained. However, just because particularly forceful plumes (or subducting plates) are able to pass through 660 does not mean that the whole system will overturn. Mixing will take a certain time before upper and lower layers are homogenised. This mixing time and style is critical if the layered to non-layered model is to have any future. Critical Clapeyron slope values that can cause layering at near Earth-like Ra are thought to be between -6 MPa/K and -2.8 MPa/K (Peltier 1996, and references therin), with the less negative values corresponding to studies using basal Ra of order 10^7. These values lie tantalisingly close to the generally accepted range of -2 MPa/K to -3 MPa/K for the 660 km phase change (Ito & Takahashi 1989, Ito et al. 1990, Davies 1999, Steinberger 2007).

3.3.3 Layering Breakdown; Possible Evidence

Breakdown in phase change induced layering could occur in response to a reduction in the overall Ra as a result of secular cooling of Earth. This has been modelled in 2 dimensions by Steinbach et al. (1993), who suggested a reduction in Ra to a value below 4×10^6 (in their model) causes layering to break down. This figure was derived from 2D models which may, to some extent, be easier to layer than 3D models (Tackley et al. 1993, 1994) but have shown consistency with other 3-D models where explicitly 3-D mechanisms are not involved (Machetel et al. 1995). Once breakdown has occurred, Steinbach et al. (1993) find a very rapid mixing/homogenisation time of 90 Myrs, which reinforces the need for the kind of partial layering suggested by

Allègre. The precise degree and rate of mixing in the mantle as a whole is still a matter of debate and appears likely to remain so for some time. Theories of a 'marble cake' mantle propose that the mantle is heterogeneous on many length scales depending on residence time. There is some support for this in geochemical observations such as the suggestion of 'stripes' of recycled crustal material in the mantle sampled along mid-ocean ridges (Graham et al. 2006).

When a strictly layered mantle is modelled, there is a marked tendency for the lower layer to retain heat and become significantly hotter than the upper layer (Oldham & Davies 2004). This is due to inefficient heat transport in the absence of mass transfer, both conduction and radiative heat transport in spinel, perovskite and low spin iron being poor (Hofmeister 1999, Goncharov et al. 2006). Therefore, if true layering were to break down, a reasonably rapid (in geological terms) influx of hot material into the upper mantle would be expected.

There are indications of 'pulsed' crustal growth, interpreted from 'mantle depletion events' identified using Re-Os systematics (Pearson et al. 2007) and dated using ancient zircon grains (Rino et al. 2004). Helium isotopic evidence may also hint at periodic large melting events (Parman 2007) though the cause of these events has not been directly identified. A possible candidate is periodic layering and breakdown within the mantle, though this does not fit particularly well with the idea of a linear evolution from layered (past) to whole mantle convection (present), indicating that a more cyclic underlying process may be involved.

With regard to relatively recent Earth history, geochemical lines of evidence suggest that mantle temperatures may have been elevated in the Cretaceous (Machetel & Humler 2003), possibly associated with the Cretaceous magnetic superchron. A number of large igneous provinces (LIP) occur in the Cretaceous, including the Ontong-Java, the Manihiki and the Hikurangi. It has been suggested that these three may be part of the same LIP (Taylor 2006), providing a possible major surface expression of excess mantle temperature.

Looking further back in time, a wide variety of geochemical results indicate 'superplume' events (excess mantle temperatures and wide spread LIP activity) occurring in a cyclical pattern (Condie et al. 2001, Condie 2004, Courtillot & Olson 2007). Superplume events are a favoured mechanism to explain the apparent clustering of Precambrian crust ages and igneous features such as kimberlite pipes (Kumar et al. 2007). During the Archean, it is suggested that the ferocity of these events could have extensively resurfaced the planet (Abbott & Isley 2002). Such work seeks to link magnetic superchrons, plumes, LIPs and mass extinctions. They demonstrate how dramatic an effect such excess temperatures could have and how the pulsing of mantle temperature could be a powerful force, not only in the physical evolution of the Earth, but also in the biosphere.

The aim of this work was to use mantle convection modelling to investigate layering breakdown about the 660 phase change as a causative mechanism for the aforemen-

tioned observations.

3.4 Methods

3.4.1 TERRA Modelling Code

The TERRA mantle model has been extensively described in the literature (Baumgardner 1985, Bunge & Baumgardner 1995, Yang & Baumgardner 2000). Briefly, TERRA is a parallel 3 dimensional spherical geometry fluid dynamics based mantle circulation code. It is implemented using MPI (Message Passing Interface) on parallel supercomputers such as 'beowulf' computing clusters. The spatial resolution of the code is scalable depending on the complexity of the model required. In this work, mid-mantle lateral resolutions of ≈88 km, ≈44 km and ≈22 km were used. Radial resolutions were similar but enhanced at the upper and lower boundaries.

Lower resolutions are useful as they allow more efficient modelling at sub Earth-like convective vigour. As vigour increases, higher resolutions are required to resolve the dynamics. When a negative Clapeyron slope is being modelled and sub Earth-like Ra are being used (e.g. 10^4), the value of the slope needed to enforce layering is higher than it would be with a more realistic Ra of around 10^8. This relationship leads to a varying relationship between Ra and Clapeyron slope negativity for convective layering controlled by an endothermic phase change.

TERRA implements phase changes in the olivine mineralogy as a sheet mass anomaly

at the specific depth (410 or 660) as used by Tackley et al. (1994) and Bunge et al. (1997). In terms of how this operates; radial forces are added to the model layer in which the phase change occurs in response to perturbations in the phase change horizons. These are detected by taking the temperature at the depth and applying either an upwards or a downwards force depending on the phase at that point. For example, if a point at 660 km is hotter than the temperature for the shallower phase to be stable then a downward force is applied and vice versa. The volume change and the effect of latent heat across the phase change are not included due to the Boussinesq equation of state used.

The cases modelled use both basal and internal heating, the Ra generally quoted is the basally heated Ra, defined in Equation 3.1. The internally heated Ra is given by Equation 3.2 and is relevant to this study as a 'true' overall Ra would be a combination of the two. However, to keep this work consistent with the previous literature, a traditional basally heated Ra is used to describe each case.

$$Ra_{bheat} = \frac{\alpha \rho g \Delta T D^3}{\kappa \mu} \tag{3.1}$$

$$Ra_{iheat} = \frac{\alpha \rho^2 g H D^5}{\mu \kappa k} \tag{3.2}$$

$$Where: \kappa = \frac{k}{\rho C_p} \tag{3.3}$$

Where: α = thermal expansion, ρ = density, g = gravitational acceleration, H is the rate of internal heat generation (W/kg), Δ T = temperature change across domain,

D = domain thickness, κ = thermal diffusivity, μ = dynamic viscosity, k = Kinematic viscosity and C_p = specific heat and constant pressure.

3.4.2 Mapping out Layered to Whole Mantle Transition

The parameter space of 660 km Clapeyron slope and Ra, which was mapped out, lies between -2 MPa/K and -20 MPa/K and 1.56×10^3 and 7.79×10^6. Figure 3.3 outlines the cases modelled, higher Ra cases being modelled at higher resolution as indicated. The cases modelled were heated, like the Earth, both basally and

Ra_{bheat}	1.56E+03	5.19E+03	1.11E+04	5.19E+04	9.74E+04	5.19E+05	1.56E+06	7.79E+06
Ra_{iheat}	2.87E+04	9.57E+04	2.05E+05	9.57E+05	1.79E+06	9.57E+06	2.87E+07	1.44E+08
Cl660								
-2								41
-4	1	2	3	4	5	40	34	39
-8	6	7	8	9	10	37	36	
-12	11	12	13	14	15			
-14	26	42	43	38	27	30	29	
-15	31	32	35	33	28			
-16	16	17	18	19	20			
-20	21	22	23	24	25			

Figure 3.3: The parameter space being explored with case numbers. Note the restricted number of cases at higher resolution. Due to the high computational cost of these higher resolution runs, cases with more negative 660 Clapeyron slope values were not run as their status can be inferred from lower vigour cases. Red 88 km, yellow 44 km and green 22km resolution. Both basally heated (bheat) and internally heated (iheat) Ra are given for each case.

internally. The internal heating simulates radiogenic heating, which in the model is evenly distributed throughout the mantle. The viscosity profile of the simulated mantle was kept uniform to eliminate any layering effects of contrasting viscosities. The Ra was altered by varying the dynamic viscosity between 5×10^{25} Pas and 1×10^{22} Pas. As mentioned above, there are 2 Ra of importance when considering both internal and basal heating. The ratio between these two Ra is fixed for all cases

as the upper and lower boundaries have fixed temperature values and the internal heat generation is the same for each case. The ratio is:

$$Ra_{iheat} = 18.425 Ra_{bheat} \tag{3.4}$$

Parameter	*Value*
Internal heating	5×10^{-12} W/m^3
Clapeyron slope of 410 km phase change	1.5 MPa/K
Equation of State	Boussinesq
Reference density	4.500×10^3 Kgm^{-3}
Gravitational acceleration	10 ms^{-2}
Volume coefficient of thermal expansion	2.500×10^{-5} 1/K
Thermal conductivity	4 W/m/K
Specific heat at constant volume	1.000×10^3 Jk$^{-1}kg^{-1}$
Temperature at outer shell boundary	3.000×10^2 K
Temperature at inner shell boundary	2.850×10^3 K
Viscosity Profile	uniform
Boundary conditions (velocity)	free slip
Inner radius of spherical shell	3.480×10^6 m
Outer radius of spherical shell	6.370×10^6 m

Table 3.1: Common input parameters. Shown for completeness as the dimensional equations solved by TERRA can be shown to be controlled only by Ra (Chapter 2).

The decision as to whether a particular case is layered or not is very important and requires a number of factors to be taken into account. The first step is to plot the average absolute mass flux (ignoring sign) for each radial layer for the final timestep once the case has stabilised sufficiently. This method of determining the state of layering has been used in previous studies (Tackley et al. 1994, Peltier 1996). In theory, a layered system should have a mass flux of zero at the depth of layering, however, phase change layering involves a boundary with a certain relief (although

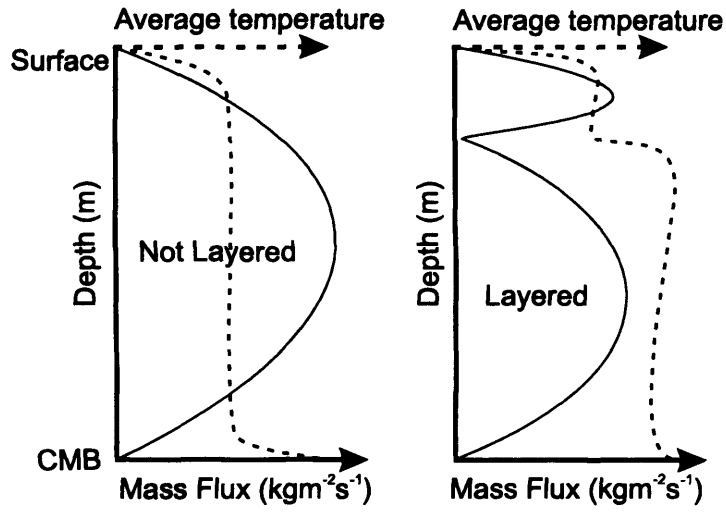


Figure 3.4: Schematic example of radial mass flux and average temperature by layer.

this is not explicitly implemented in our code). Thus mass transfer through the phase boundary may be zero but values at a specific radial depth (660 km) will still show a small mass flux. The second step is to plot the average temperature by layer, which allows a check to be made on the mass flux result. A layered system's temperature profile will show an internal thermal boundary layer either side of the horizon of layering (660). Figure 3.4 demonstrates typical profiles for both mass flux and average temperature in both a layered and an unlayered system. Finally, 3-D visualisation can be used to examine upwelling and downwelling features and their interaction across layers. This was done for selected cases and was made possible by the MantleVis parallel visualisation programme (Heath 2002).

3.5 Results

A total of 43 cases were successfully modelled. Radial mass flux and temperature profiles are collated in Figure 3.5. The time dependence of the results is illustrated

by Figure 3.6. Thus, Figure 3.5 consists of representative curves, capturing a typical profile for the case in question, rather than definitive values. It is evident that a number of different convective behaviours are demonstrated by the models. Considering the broad trend, the lower right portion of the diagram (more negative Clapeyron slopes, low Ra) is generally layered, the upper left portion (less negative Clapeyron slopes, lower Ra) is not layered. The region between these end members consists of a transitional zone where the models share characteristics of both layered and whole mantle convection.

Figure 3.7 provides a summary of the modes of behaviour, individual cases are represented as points and the different behaviours are shown as coloured areas. Projecting the observed trend to higher Ra (10^7 to 10^8), an Earth-like 660 Clapeyron slope of -2 to -3 MaPa/K plots in the transitional regime, where partial layering occurs.

Based on this possibility, Figure 3.8 shows three cases; layered, partially layered and whole mantle convection. The partially layered case is interesting in that it contains both layered and whole mantle convection features. With downwelling features this is particularly noticeable. Laterally small downwellings do not penetrate 660 while larger features push through and reach the core-mantle boundary. Upwelling features are different as features in the upper 'layer' are generally seeded off a lower upwelling, which has impacted 660 and spread out, heating the region above. Therefore, partial layering is in effect a low pass filter, allowing only upwellings or downwellings above a certain size to penetrate the phase change. This concept has been discussed pre-

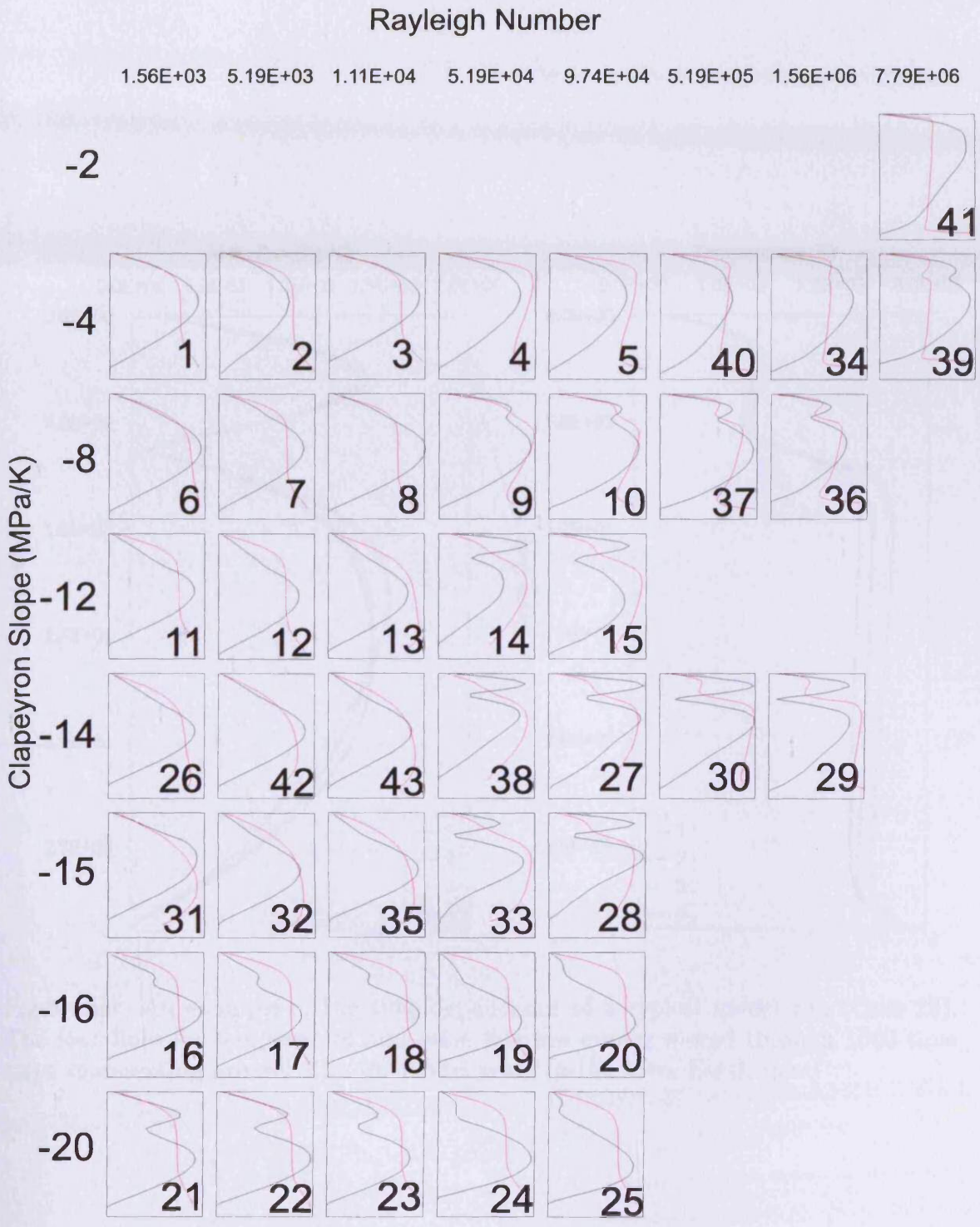


Figure 3.5: Table of radial mass flux (black lines) and radial temperature (red lines). Vertical scales match, horizontal scales are variable, therefore absolute magnitude of mass flux and temperature at a given horizon in each case is different.

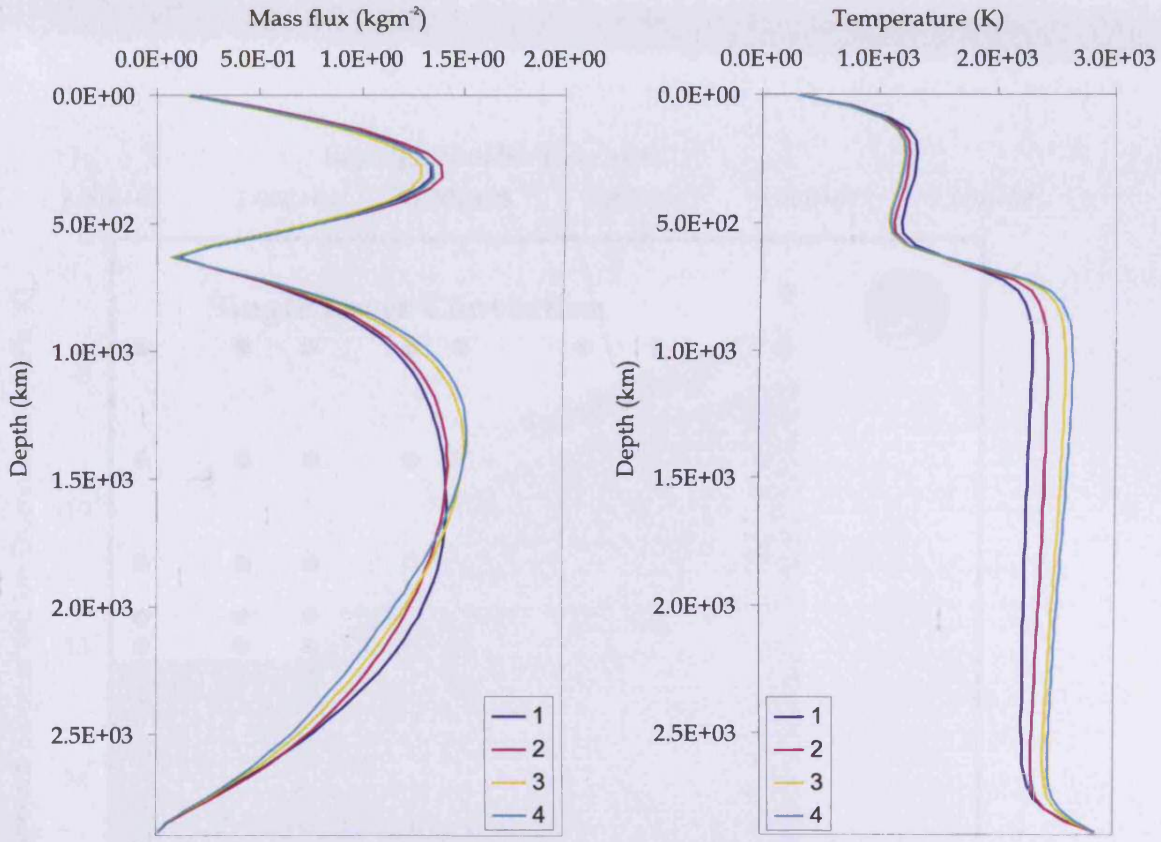


Figure 3.6: An example of the time dependence of a typical model run (Case 29). The four lines for temperature and mass flux are evenly spaced through 1000 time steps representing around 1.5×10^9 model years ($\approx$150 Myrs Earth time)

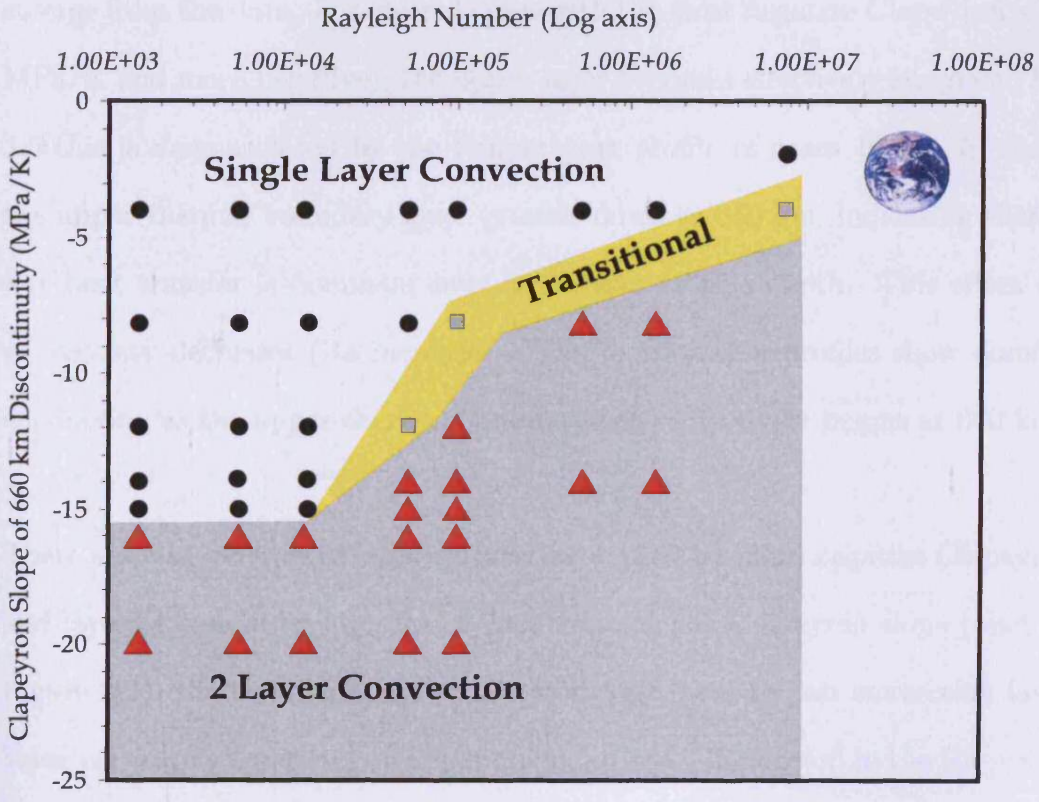


Figure 3.7: Summary of the results. Black circles - single layer convection, grey squares - transitional cases and red triangle - 2 layer convection. Present day Earth is shown at its assumed location in the parameter space.

viously by Ringwood (1994). The cumulative effect is that the lower mantle regions of partially layered models are characterised by longer wavelength features than the upper mantle regions. The effect is most noticeable in some quite high Ra cases as illustrated by Figure 3.9.

In addition to the large scale behaviour outlined above, several less obvious trends emerge from the data. For layered cases with the most negative Clapeyron slopes (-16 MPa/K and more negative), the upper layer becomes effectively stagnant. In Figure 3.5 this is demonstrated by the temperature profile of cases 16-25. In these cases, the upper thermal boundary layer extends down to 660 km, indicating that conductive heat transfer is dominant over convection to this depth. This effect increases as viscosity decreases (Ra increases). The temperature profiles show dominance of conduction as the upper thermal boundary layer effectively begins at 660 km.

There are also differences between layering caused by more negative Clapeyron slope and layering caused by high Ra. When reducing the Clapeyron slope (reading down Figure 3.5), the layering effect passes from one layer to two convecting layers to a lower convecting layer and an upper stagnant layer; illustrated in the Ra $= 9.74 \times 10^4$ column. When increasing Ra is responsible for generating layering (reading across Figure 3.5), there is no upper layer stagnation, the layering simply becomes more pronounced e.g. -14 MPa/K row in Figure 3.5.

Some of the model cases have a somewhat inverted temperature profile, where the

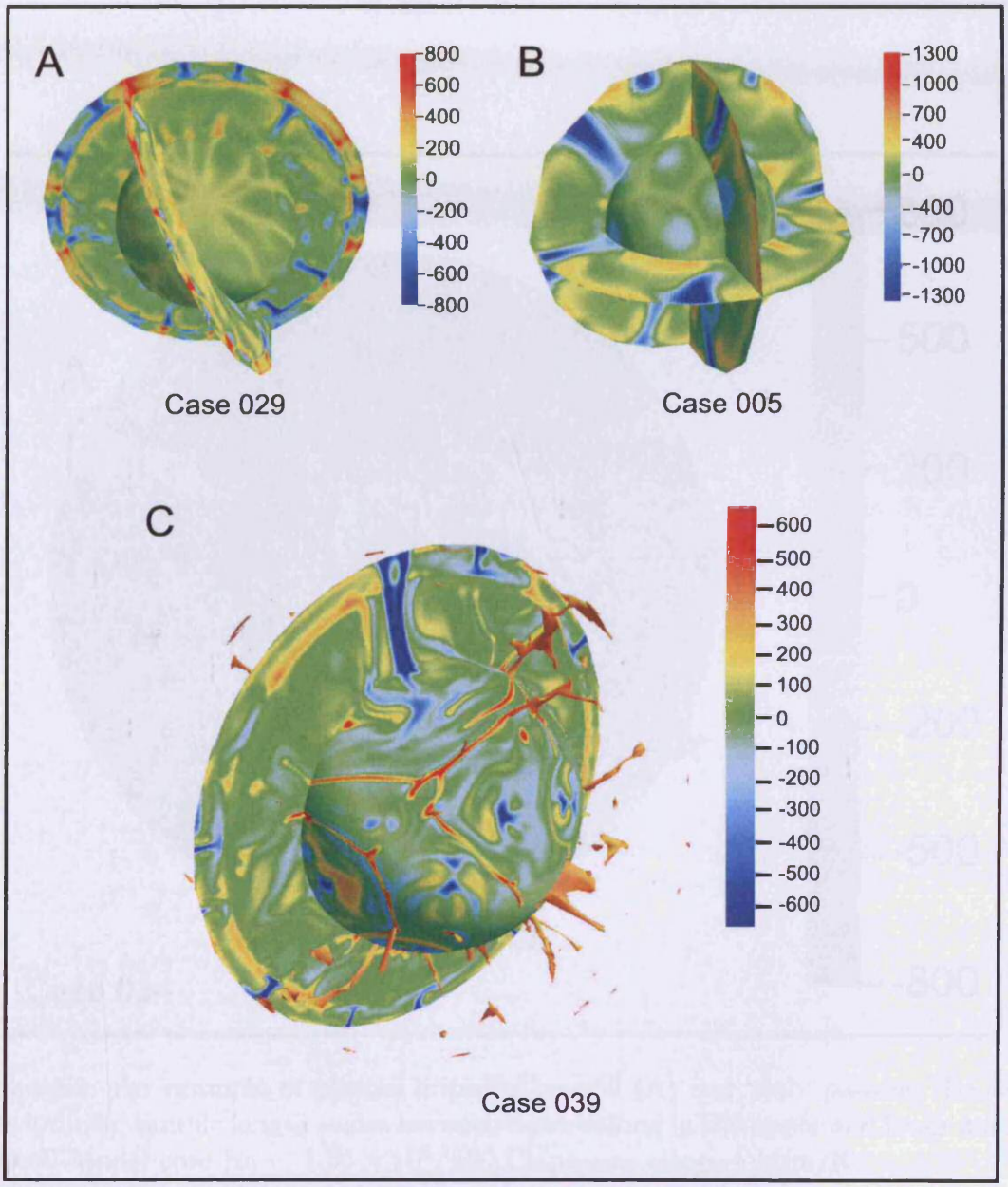


Figure 3.8: Examples of the three dominant modes of convection. (A) layered convection, (B) Whole mantle convection and (C) A partially layered mantle. All plots represent adiabatic temperature anomaly in Kelvin.

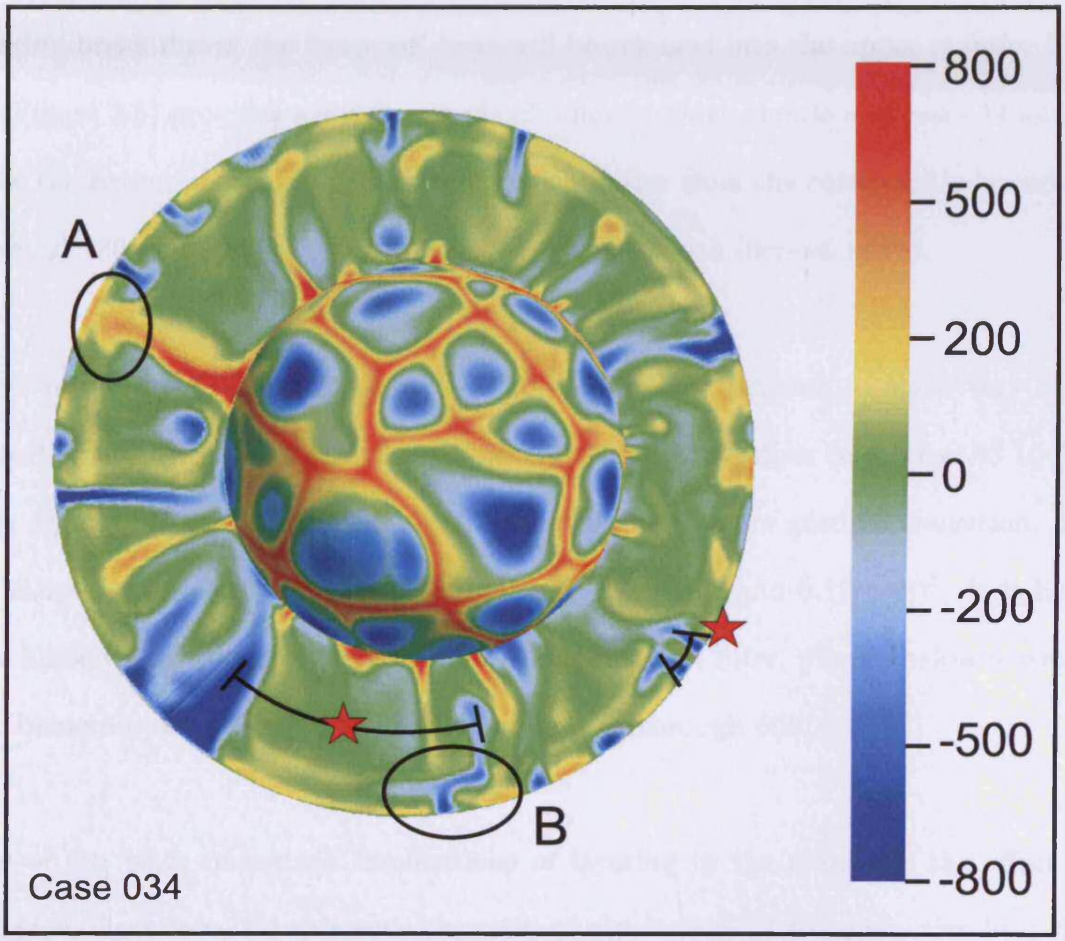


Figure 3.9: An example of plumes impeded by 660 (A) and 'slab' pausing (B). Red stars indicate sample length scales between downwelling in the upper and lower mantle regions. Model case Ra = 1.56×10^6, 660 Clapeyron slope -4 MPa/K.

uppermost lower mantle is hotter than the base of the mantle. This is a function of both the level of internal heating and the ponding of material just below 660. A higher Ra reduces this effect by allowing heat to escape the system more rapidly. The significance of a hotter region at the top of the lower layer is that, should the layering break down, the 'trapped' heat will be released into the upper mantle. Case 29 (Figure 3.5) provides a good example of a hotter lower mantle and cases 14 and 15 show the region just below 660 as being much hotter than the core-mantle boundary. Cases 27, 30 and 29 show how the effect is reduced with increasing Ra.

Some mid-range Clapeyron slopes (-12 to -15 MPa/K) appear to show very rapid transition form whole mantle convection to layered convection (e.g. case 43 to 38). Less negative (more realistic), Clapeyron slopes show a more gradual transition. The Ra range for dramatic transitions is between 1.11×10^4 and 5.19×10^4. It is likely that these Clapeyron slope values enforce a very narrow filter, plumes below a certain size being almost entirely prevented from passing through 660.

One of the more important implications of layering in the mantle is the effect on planetary heat flux. To this end, the ratio of convective (q) to conductive heat flux (q_K), known as the Nusselt number:

$$Nu = \frac{q}{q_K} \tag{3.5}$$

and Ra values for all cases have been plotted in Figure 3.10. The scatter of the data is somewhat larger than ideal, this is due to the cases not being run to allow heat flux to settle to a completely steady state, the primary goal of the modelling being the

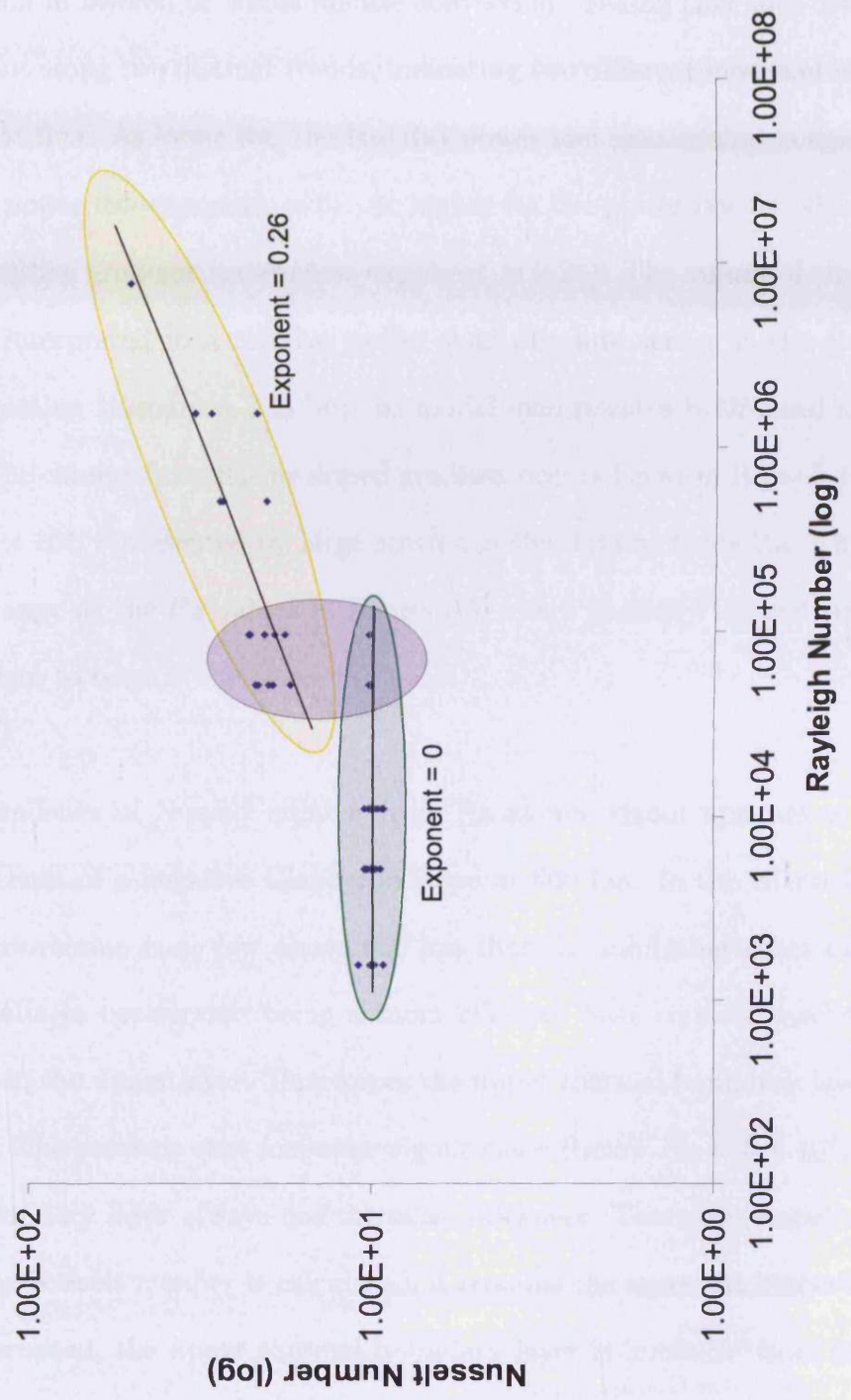


Figure 3.10: A plot of Nu against Ra for all cases modelled, the data cluster into two trends (green and yellow areas). The Ra overlap of these areas shows a relatively large variation in Nusselt numbers (purple area) and is associated with Ra of 5.19×10^4 and 9.74×10^4.

identification of layered or whole mantle convection. Taking that into consideration, the data plot along two distinct trends, indicating two different modes of behaviour as regards heat flux. At lower Ra, the Nu(Ra) power law relationship is essentially flat (Nu $\approx$ 10, power law exponent $\approx$ 0). At higher Ra the power law fit takes on a more familiar positive gradient (power law exponent $\approx$ 0.26). The values of the exponents should be interpreted in a relative rather than absolute sense, as the Ra quoted is for basal heating (Equation 3.1) but the model incorporates both basal and internal heating. The change from flat to sloped gradient occurs between Ra $= 5.19 \times 10^4$ and Ra $= 9.78 \times 10^4$, represented by large scatter in the data at those Ra. The Ra are in the same range as the Ra values in Figure 3.7 where partially layered (transitional) systems begin to occur.

The independence of Nusselt number from Ra at low vigour appears to be caused by the presence of a negative Clapeyron slope at 660 km. In the effected cases, the vigour of convection is so low above 660 km that the inhibiting effect of the phase change results in conduction being a more effective heat transfer mechanism then convection in the upper layer. This forces the upper thermal boundary layer to begin at 660 km. The result is that for lower vigour cases (below Ra $\approx 5 \times 10^4$) the upper thermal boundary layer always has the same thickness. Therefore, based on the way in which the Nusselt number is calculated, it remains the same. At higher vigour this effect is disrupted, the upper thermal boundary layer is 'unstuck' from 660 and the Nu scales in a power law relationship with the Ra. The trend is evident in Figure 3.5 and the thermal profiles of cases 6-10, 37 and 36.

To illustrate this inflection point further, Figure 3.11 demonstrates how variations in 660 Clapeyron slope negativity do not appear to affect Nusselt number until the Ra is 5.19×10^4 or larger. This demonstrates the importance of Ra in understanding a mantle system with phase changes.

A final question is whether cases that are layered remain layered for a significant period of time if their Clapeyron slope is increased or their Ra decreased. The scaled Earth equivalent time for layering to break down in an experiment when Ra was instantaneously reduced from 9.74×10^4 to 1.11×10^4 at a Clapeyron slope of -14 MPa/K (Case 27 to 43) was found to be between 9 and 22 Myrs. The temperature profile responds equally rapidly as convection quickly takes over to destroy the internal thermal boundary layer. When the 660 Clapeyron slope is changed from -20 MPa/K to -14 MPa/K at Ra = 5.19×10^3 (Case 22 to 42), breakdown occurs almost instantly, or more precisely, on a shorter time scale than the temporal resolution of the output (every 10 time steps). Quoting an equivalent Earth time for the change would be without merit. The temperature profile for the Clapeyron slope change also matches the target case very rapidly, probably because the profile was quite similar to begin with. The large caveat attached to these times for breakdown is that the Clapeyron slopes are much more negative than those of the real Earth, thus the change is expected to be more radical than at Earth-like Clapeyron slopes. What these two simple cases essentially demonstrate is that, hysteresis (an already layered system remaining layered despite changed parameters) is not present in the modelled

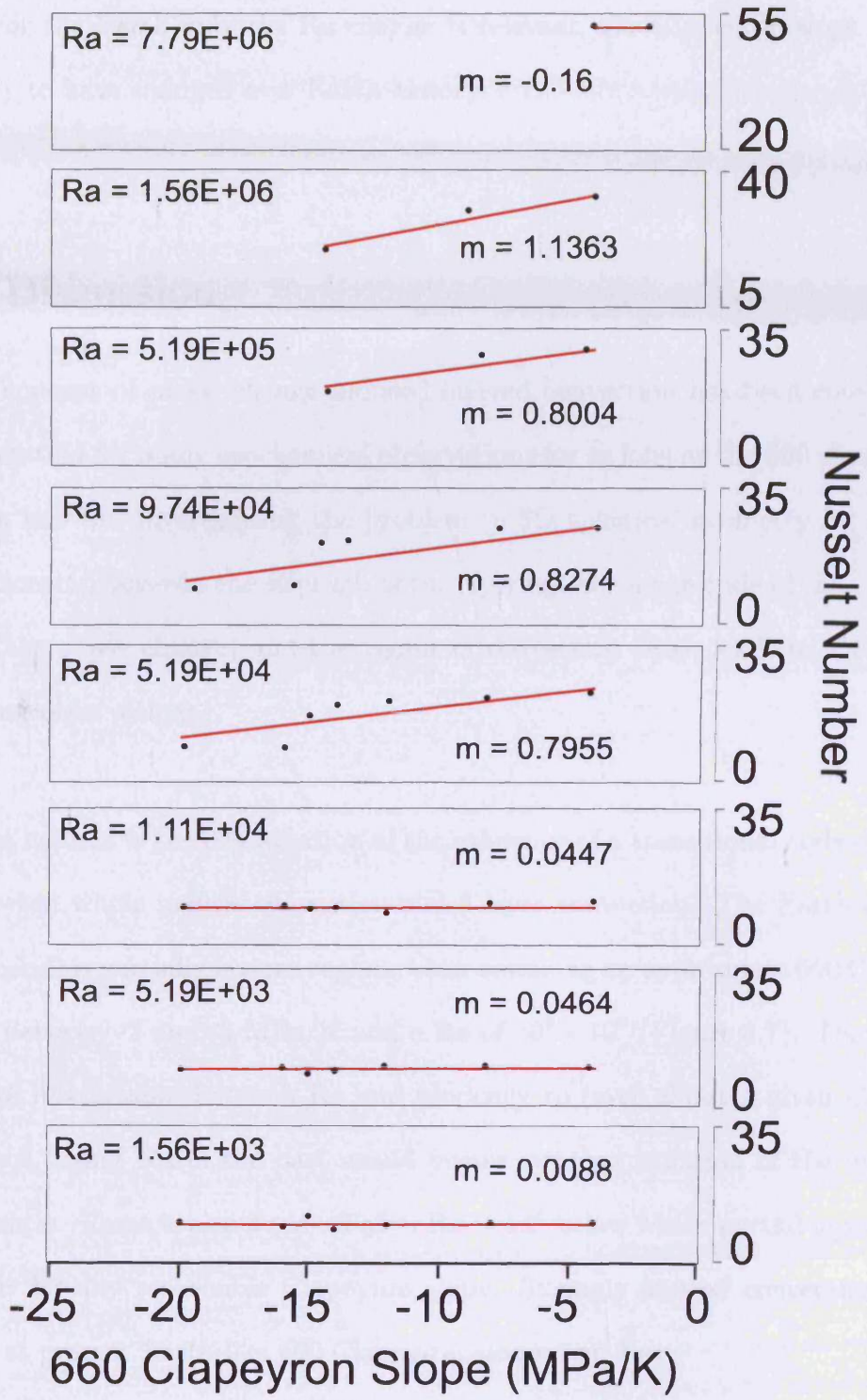


Figure 3.11: Nu scaling over the range of 660 Clapeyron slope values as Ra is increased. Lines are best fit (linear) to the data points, m is the gradient of the associated line. Note: the line for Ra = 7.79×10^6 is based on two data points and may therefore not be indicative.

cases. For the Earth only the Ra change is relevant, the Clapeyron slope of 660 is not likely to have changed over Earth history.

3.6 Discussion

The concept of phase change induced layered convection has been considered as an explanation for many geochemical observations for as long as the 660 phase change has been known. Investigating the problem in 3D spherical geometry by exploring the relationship between the strength of the layering (the magnitude of the Clapeyron slope of the phase change) and the vigour of convection (Ra), leads to a number of interesting observations.

The most evident is the confirmation of the existence of a transitional style of convection between whole mantle convection and 2 layer convection. The Earth may currently lie in this partially layered region, when assuming an appropriate 660 Clapeyron slope of between -2 and -3 MPa/K and a Ra of 10^7 - 10^8 (Figure 3.7). There is also a positive relationship between Ra and tendency to layer about a given Clapeyron slope, thus higher Ra in the past would favour stronger isolation of the upper and lower mantle. There is also a cut-off at $\approx$ Ra = 10^4 below which partial layering does not occur for any reasonable Clapeyron slope. Strongly layered convection is very unlikely at present Earth-like 660 Clapeyron slopes and Ra.

A higher Ra would be needed to cause significant layering in the past. An varia-

tion in Ra is likely to be due to a larger temperature difference across the mantle (ΔT), or a reduction in viscosity, given that the other terms used to generate Ra are considered to be more or less constant over time. An increase of ΔT by 1000K increases the Ra by $\approx$ 28% (using equation 3.1). However, it is possible that the temperature at the base of the mantle may be constrained by the fact of inner core crystrallisation and a rapidly mixing outer core, effectively fixing the CMB temperature. In that case, the effective Ra of the mantle would be higher in the past due to a higher average temperature within the mantle as a result of higher radiogenic heating and retention of a greater proportion of the original heat of formation. Additionally, the effect of higher viscosities in the lower mantle (Forte et al. 2002, not implemented in this study) might assist in stabilising any layering. Thus it is possible if not probable that at some point in the past the mantle did support layered convection.

If one assumes this to be the case, the question then becomes one of whether layered convection endured for a significant time in Earth's history and what the mode of layering breakdown was. The length of time layered convection could have lasted is tied to the thermal history of the planet, precise numbers for which are difficult to define. The mode of layer breakdown is more accessible. At Earth-like 660 Clapeyron slope and just below, the reduction of the Ra results in a smooth transition from stronger to weaker layering to whole mantle convection (upper three rows of Figure 3.5). This would suggest that any layer breakdown would progress gradually and may not lead to geological evidence such as LIPs. However, when a layered case is restarted with a Ra or 660 Clapeyron slope changed to that of a known non-layered

case, layering breaks down rapidly. For the scenario of a reduced Ra, layering broke down in $\approx$ 20 Myrs of real Earth time. As the Ra of the cases in question were significantly lower and 660 significantly more negative than the Earth, this is likely to be an upper value of the reorganisation time of the convecting mantle. Thus, once the system passes out of a Ra that enforces layering, transition to whole mantle convection appears to be rapid. This agrees with the 2D work of Steinbach et al. (1993) who argues for rapid layering breakdown, taking place within 90 Ma. To test this accurately would require model cases where the Ra is allowed to evolve (reduce) over time from a Ra of order 10^9 (this is done in Chapter 4). The implication of this result is that in order for the model of Allègre (2002) to be accurate, partial layering at the present is essential in order to preserve some form of isolated chemical reservoir.

The features of partially layered cases relevant to the evolution of the Earth are: dominance of a few large downwellings, small scale convection in the upper layer, plumes ponding below 660 before seeding smaller plumes in the upper layer and an increase in the overall complexity of the temperature field (Figure 3.8). These features suggest that in a partially layered Earth, plumes may be slowed by 660, possibly exerting some sort of magnitude control on rising plumes. Also, as only strong downwellings appear to penetrate 660, ambient downwelling may not be a particularly significant mechanism for introducing material into the deep mantle. This effect is likely to be increased when the 'downwellings' are subducted plates, which have both negative buoyancy and mechanical strength.

The heating of the region below the phase change (e.g. case 36, Figure 3.5), due to stagnation below 660, raises several interesting possibilities. Should the layering rapidly break down at any point, this heat would be released into the upper mantle. Such a process could provide a candidate mechanism for 'superplume' events and their associated LIPs. This mechanism in reverse (cold 'plumes') may be related to the 'mantle avalanches' described by Tackley et al. (1993). The development of a hot layer in the upper mantle also provides scope for dynamic features such as 'splash plumes' (Davies & Bunge 2006), which involve downwelling (subducting) cold material impacting a hot layer in the mantle. The displacement of this layer could then give rise to secondary plumes. Secondary plumes of this type were not observed in the cases modelled here, however the upper boundary was not driven by plate motion history as in Davies & Bunge (2006).

The effect that introducing a Clapeyron slope had on the Nusselt number was somewhat unexpected. The manner in which the Nu is 'held' at ≈ 10 below $\mathrm{Ra} = 1.11 \times 10^4$ is not observed in model cases where phase changes are not included (e.g. in Chapter 2). The cause of this effect lies in the behaviour of low vigour cases. The presence of a negative Clapeyron slope at 660 km in these cases disrupts the convective pattern such that convection is less efficient than conduction above 660. The result is a fixed thickness upper thermal boundary layer when the Ra is below a critical, which lies between 5.19×10^4 and $\mathrm{Ra} = 9.78 \times 10^4$. Extremely negative Clapeyron slope cases show a similar 'stagnation' of the upper layer, though the effect is relatively independent of Ra. In the Earth it is of relatively minor interest as Earth-like Ra is several

orders of magnitude higher than the point below which this occurs.

In light of the preceding results, a selection of implications of these observations for the geochemistry and thermal behaviour of planetary convecting systems are considered. A first order question worth considering is, given the negative Clapeyron slope of 660, how much shallower would the phase change occur in a hotter mantle. The answer is surprisingly small, for a 1000 K increase in temperature the phase change would occur some 40 km shallower. The 660 phase change is therefore likely to have occurred at a similar depth for much of Earth's history. Within the regime of solid state mantle convection the dynamical effects of the depth of the 660 phase change can be considered fixed over time.

If 660 causes partial layering in the Earth with a filtering effect on both upwelling and downwelling material, it should be possible to observe some consequence of this. Possible evidence would constitute; descending slab material held at 660 for a certain time and plumes being 'slowed' or ponded at 660.

Should this occur in the Earth, as well as supporting the conceptual model of Allègre (2002), it could enforce some sort of periodicity on plume mechanics. It could also play a controlling role in plume size, with a stronger layering effect in the past forcing plumes to accumulate for longer below 660 before having sufficient strength to break through, Figure 3.9 demonstrates this in one of the model cases. If this was the case plume size should reduce over time as the mantle cools and 660 becomes less effective.

Seismic studies conclude that at least some descending slabs seem to stagnate above 660 km (Van der Hilst et al. 1991, Fukao et al. 1992, Tseng & Chen 2004, Tajima & Nakagawa 2006), suggesting that the dynamical effect of the phase change on slab material may be observable. Alternately, the observations might represent an effect of higher viscosity in the lower mantle. Plumes are more complex bodies to image, indeed there is still somewhat polarised discussion as to whether classical mantle plumes in fact exist (Anderson 2005, for a plume sceptic review). That question will not be addressed here, plumes originating at the core-mantle boundary are present in all models that use sensible parameters for the Earth. It is possible to force models such that plumes are very weak (see Figure 3.12) but these require extremely large deviation away from generally accepted parameters for viscosity (Ra) and 660 Clapeyron slope. However, the fact that there is discussion indicates that seismic tomographic imaging of suspected plume regions is, at the time of writing, still not entirely conclusive (Nataf 2000). Montelli et al. (2006), using seismic methods, suggest that at least two types of plumes exist, those that can be imaged to a depth of 660 km (transition zone) and those that can be imaged into the deep mantle. The unanswered question is whether the shallow plumes are generated by heat from relatively weak lower mantle plumes (such as those described by Labrosse 2002), which are unable to penetrate 660. These weak plumes would be much more difficult to image as their velocity anomaly would be low (Montelli et al. 2004). If this was the case, then plumes that reach from the core-mantle boundary to the surface would be strong plumes with the required force to break through 660. To summarise, some

slabs do appear to be affected by 660 and 660 does appear to play a role in plume transit through the mantle. There is therefore scope within the published data for partial layering to exist at present.

A problem that arises is that a generally linear trend of cooling and reducing propensity to layer can not explain observations of repeated pulsing of heat in the mantle (Pearson et al. 2007). A pulsing mantle system would indicate that surface heat flux is time dependent and that some form of recharge-release mechanism operates in the deep Earth. Short-term pulse events would probably involve larger thermal variation than the overall background trends, leading to large uncertainty when one attempts to reconstruct the thermal history of the Earth with the large temporal spacing of many geochemical samples such as LIPs.

Cases modelled at unrealistically negative Clapeyron slopes are useful as a 'scaled down' Earth Ra requires a more negative than actual 660 Clapeyron slope to produce an effect equivalent to running at real Ra. There is however a limit to how applicable these cases are to the Earth. Such negative Clapeyron slope cases may be useful in an extra-terrestrial setting. A number of bodies in the solar system have a high water content e.g. Uranus, Neptune, Europa and large comets. Water has some 11 proven phases with several further suspected phases. The ice II to ice V transition is endothermic, has a Clapeyron slope of $\approx$ -5 MPa/K and occurs in the range of 0.5 GPa (based in phase diagrams Sanz et al. (2004), Johari & O (2007)). Higher pressure phase transitions such as ice VIII to ice X may have more negative Clapeyron

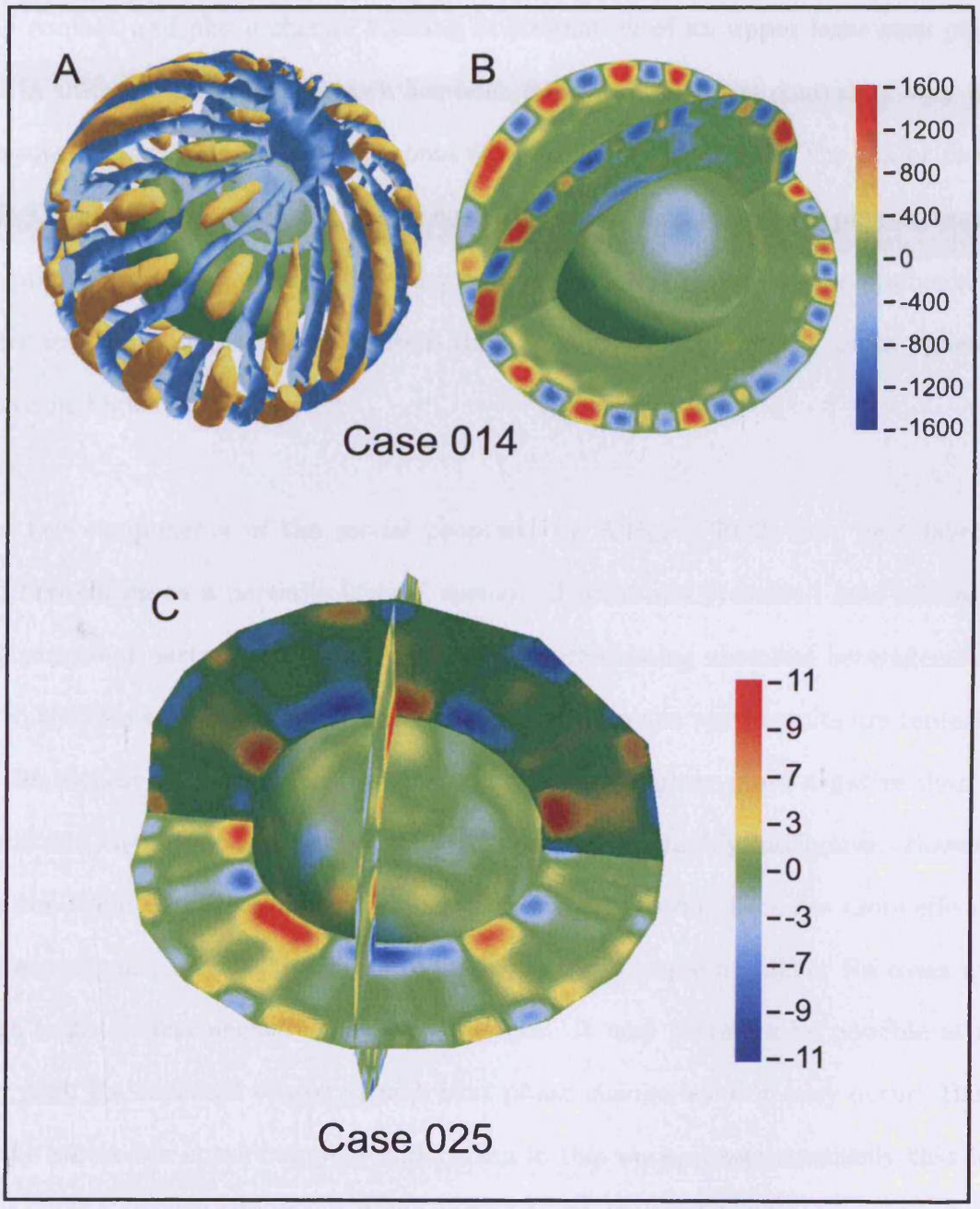


Figure 3.12: An example of how unusual (far from Earth-like) parameters produce interesting results. Parameters: Case 014, Clapeyron slope = -12 MPa/K, Ra = 5.19×10^4. Case 025: Clapeyron slope =-20 MPa/K, Ra =9.74×10^4

slopes. Thus, if an ice body is large enough and has a sufficient thermal gradient it may convect and phase change layering or stagnation of an upper layer may play a role in their evolution. Some work has been done on this in the context of very high pressure phases of water and ammonia (Carvazzoni et al. 1999). The major caveat with the application of this study to such systems is that the model physics assume an infinite Prandtl number, very reasonable for rock but it may not be applicable to water ice. An example of how unusual the dynamics of such systems might appear is shown in Figure 3.12.

The key components of the model proposed by Allègre (2002) are; past layering and breakdown to a partially layered system. The results presented here affirm the importance of partial layering to preserve any pre-existing chemical heterogeneities. As to the existence of strong layering at some point in the past, results are tentative. At the highest vigour modelled in this study with a slightly more negative than realistic 660 Clapeyron slope of -4 MPa/K, layering was highly ineffective. However, at more negative Clapeyron slopes (e.g. -8 MPa/K) layering becomes more effective at lower vigour, possibly providing a 'scaled down' glimpse at higher Ra cases with more realistic (less negative) Clapeyron slopes. It may therefore be possible at the very high Ra expected of early Earth that phase change layering may occur. Based on the parameter space mapping undertaken in this work, it seems unlikely that layering persisted until 500 Ma as Allègre requires for his model. The reason for this is that, based on the work presented here, the Ra would have reduced beyond the value required to maintain layering well before 500 Ma. Allègre developed his hypothesis

using 2-D mantle models, which appear to be able to maintain layering to much lower Ra that 3-D spherical geometry models, leading to a much more recent date for layer breakdown. Based on the results presented here some form of partial layering, which influences the behaviour of plume-like and slab-like features, is more likely to have been active in the post Precambrian. Indeed some indication of slab 'ponding' at 660 is seen in seismic images at the present.

As an alternative solution, other geochemical models are able to dispose of the concept of layering in the upper mantle entirely. Boyet & Carlson (2005, 2006) suggest that, based on $^{142}Nd/^{144}Nd$ values, the Earth underwent segregation within the first few million years after formation and that this material has never been sampled. Conceptually, this idea supports work by Labrosse et al. (2007) who suggest that a significant volume of incompatible elements were isolated in a melt layer above the core-mantle boundary, the remains of which are responsible for ultra low velocity zones observed by seismology. Though these models solve several geochemical issues, there is still the matter of the dynamics of the 660 phase change and its influence on the thermal evolution of Earth. It may be that the mantle system represents a combination of many of these effects, each influencing the overall result. Only by building a full set of these mechanisms into our models will we be able to assess the relative influences of the various mechanisms.

3.7 Conclusion

The relationship between 660 km phase change Clapeyron slope and Rayleigh number has been mapped out over a large range of both parameters.

The model of Allègre (2002) is only partially supported by the results, partial layering is possible at present mantle conditions. By contrast, the concept of strong layering surviving to 500 Ma is not supported by Earth's likely time evolution across the mapped parameter space of Cl660 and Ra. From the results presented here (e.g. Figure 3.7), the range of Ra resulting in partial layering at approximately Earth-like Cl660 values is very broad. If Earth is at best partially layered at present, then it is very likely that it was partially layered at 500 Ma. Although this chapter does not directly investigate the time of transition from layered to partially layered convection, it is likely to have occurred in the Archean eon. Therefore, other mechanisms are required to explain the observed geochemistry.

The adjustment time for the mantle to move from layered to whole mantle convection appears to be of order 20 Ma for the model cases presented here. Such a rapid breakdown would result in a significant input of heat into the upper mantle.

Partial layering of some degree is possible at present and seismological observations indicate slab pausing/deflection at 660 and a possible effect on plume transit across 660.

Chapter 4

Evolving Convective Regimes in Earth's Mantle

4.1 Abstract

The possibility of time dependent layering in Earth's mantle provides a mechanism, which could introduce periodicity into Earth's thermal evolution. In the 3D spherical geometry modelling work presented here, the 660 km phase boundary was used to induce layering to varying degrees. The interaction of a given magnitude of Clapeyron slope for the 660 phase change and a varying vigour of convection (Rayleigh number) was investigated. The time evolution of models that pass through various stages of full layering, partial layering and whole mantle convection is analysed and discussed. Significant periodic time dependence emerges in the surface heat flux as model cases move into a partially layered regime.

Previous studies have observed cold upper mantle 'avalanches' (Tackley et al. 1994) into the lower mantle. This work confirms the existence of these avalanches and finds evidence for a symmetrically opposite event, a pulse of hot material entering the upper mantle from below, driven ultimately from the surface by avalanching. Such a process has been predicted in the literature (Davies 1995) but, as far as we know, never produced in 3D spherical geometry modelling.

Should conditions within Earth's mantle be suitable for such processes to occur, the thermal history of the planet may show significant variation that is both larger and more rapid than the secular cooling trend, requiring modification of thermal history models. Periodic heat pulses have the potential to promote increased melt extraction at certain time periods, either by enhancing normal melting or by triggering large igneous provinces. This provides a mechanism that could produce 'He events' observed geochemically (Parman 2007).

4.2 Introduction

Earth's convecting mantle exerts a powerful control over the large scale geological history of the planet. Much of the debate between geochemistry and seismology has centred on the convective regime of the mantle. Geochemistry provides evidence of separate reservoirs in the mantle (Allègre et al. 1996, Boyet & Carlson 2005, 2006), while seismology proposes that the mantle convects as one (Van der Hilst et al. 1997, Montelli et al. 2004). The problem is however, not straightforward. The resolution of seismic tomography is limited by uneven data distribution and ultimately by the

long wavelength of seismic waves, requiring extensive processing to resolve relatively large features. Mantle geochemistry by contrast, is limited by the fine resolution and location specific nature of sampling (accessible samples may not be representative). Additionally, mantle geochemistry is time integrated, while seismology provides instant snapshots of the present state of the mantle. Therefore, apparent contradictions between seismic tomography and mantle geochemistry may represent a clash of methodology, where each approach is describing a different part of the same system. Computational modelling provides a powerful tool to reconcile seismology and mantle geochemistry.

A number of authors have suggested that the mantle may be partially layered about the 660 km phase change (660) (for example, Ringwood 1994, Condie 1998, Allègre 2002). Chapter 3 demonstrates that partial layering can occur within the range of 660 Clapeyron slope (Cl660) and Rayleigh number (Ra) values expected during Earth history where the basally heated Ra is:

$$Ra_{bh} = \frac{\alpha \rho g \Delta T D^3}{\kappa \mu} \tag{4.1}$$

where: ρ is density, g is the acceleration due to gravity, α is the coefficient of thermal expansion, ΔT is superadiabatic temperature drop across the shell, D is domain thickness (2900 km), κ is thermal diffusivity and μ is dynamic viscosity.

It has been shown that at high Ra, the layering effect of 660 is enhanced (Christensen & Yuen 1985, Butler & Peltier 2000). Should early Earth's mantle have been

sufficiently hotter than the present (higher Ra), it may have been effectively layered for a period of time, transitioning to whole mantle convection as the mantle cooled (decreasing Ra). The assumption of higher mantle temperatures in the Archean is supported by the analysis of melt inclusions in komatiites (Berry et al. 2008). Allègre (2002) suggested the concept of layering in the past, with an unremarkable transition to a type of partial layering at around 500 Ma. Previous modelling (Chapter 3) provides evidence that fully layered convection is only possible in the very early Earth and is unlikely to have persisted until 500 Ma. The main argument against layering about 660 proposes that Cl660 may not be sufficiently negative. Estimates vary but generally suggest a value between -1 and -2 MPa/K (Katsura et al. 2003, Fei et al. 2004), though there are more negative estimates (-2.8 MPa/K, Chopelas et al. 1994). Other factors such as: the higher viscosity of the lower mantle and the latent heat effect (Christensen 1998) are also likely to increase propensity to layer. There are some suggestions that plate size maintains long wavelength flow, reducing the propensity to layer (Korenaga 2008). This would apply to descending plates but not necessarily to plumes or ambient mantle flow. There are indications that the 660 km and 410 km phase transition may behave differently for cold material sinking than for hot material rising (Tauzin et al. 2008). Variable degrees of layered convection in the past could reconcile the apparent signature of old geochemical reservoirs with seismically observed whole mantle convection at present (Butler & Peltier 2002).

The problem with assuming transitions between different layered states is whether Earth's Ra has reduced significantly over time. This is a difficult question to answer

as it would require knowledge of Earth's initial temperature, therefore the assumption was made that Earth's Ra has indeed reduced significantly since 4.5 Ga.

The mechanism by which a layered mantle transitions to a partially layered or single layer mantle could have significant implications for the evolution of Earth. When an internally and basally heated system such as Earth's mantle is layered, the lower layer cannot lose heat effectively, leading to a 'heat up' in the lower layer (Oldham & Davies 2004). If this heat is suddenly injected into the upper portion of the mantle during a layer breakdown, large amounts of melting could take place. In a partially layered system such heat pulses might occur repeatedly, providing a mechanism for very large LIPs and/or continental building. The concept of periodic breakdown and recharge driven by partial layering has been suggested in the literature (Machetel & Weber 1991, Davies 1995, Honda 1995).

This study examines the change in Ra required to cause a transition from layered to whole mantle convection at a variety of Cl660. The time evolution of models in various states of layering is examined. The mechanisms of transition between different convective regimes are studied by allowing cooling to take place from an initially layered system. The implications for the thermal, geochemical and tectonic evolution of Earth of such a breakdown are discussed.

4.3 Simulation Methods

Earth's mantle was simulated in 3D spherical geometry using the TERRA model. TERRA is an established fluid dynamics based code for modelling the mantle (Baumgardner 1985, Bunge & Baumgardner 1995, Yang & Baumgardner 2000, Oldham & Davies 2004, Davies & Bunge 2006, Davies 2008). This study assumed the mantle to be incompressible and employed the Boussinesq approximation. The aspects of the code most relevant to this study are the implementation of phase boundaries and mantle cooling.

Phase boundaries were implemented as a sheet mass anomaly after Tackley et al. (1993, 1994) and Bunge et al. (1997). The sheet mass approximation assumes a globally continuous phase boundary, valid for phase transitions such as 660 and 410, which have been globally mapped (Shearer & Masters 1992, Flanagan & Shearer 1999). In this study, only 660 is modelled, neither are effects such as a higher viscosity lower mantle. This may somewhat reduce the effect of choosing slightly more negative values of Cl660 than may be the case for Earth. A potential weakness of the sheet mass approach is that the phase boundary is effectively assumed to be planar. However, the topography of 660 is considered to be of order 10's of km (Tauzin et al. 2008), of the same order as the resolution of the model. The phase boundary would fit within one or two model layers, providing insufficient resolution to resolve topography. Therefore, the sheet mass assumption is sufficient until such time that model resolution is <5 km.

Some models were permitted to evolve, with their effective Ra reducing over time.

Such models were given a 'run in' initial condition, where the short-scale random initial condition was allowed to stabilise before any evolution was permitted. In order to model the cooling of the mantle over time, the code allows the core to cool with the CMB temperature recalculated after every time step using the relationship:

$$MC_p \frac{dT}{dt} = kA \frac{dT}{dr} \int_{r=r_c} \tag{4.2}$$

where: M is the mass of the core, C_p is specific heat capacity of the core, T is the CMB temperature, t is time, k CMB thermal conductivity, A is the surface area of the core and r is radius of the CMB.

All heat flow across the core-mantle boundary is assumed to be the result of core cooling. The core is assumed to be homogeneous and to cool at the same rate throughout. The lack of an inner core leads the modelled core to cool slightly faster as latent heat is not released from the crystallisation of the inner core. Gravitational energy released by the separation of the lighter element is also ignored. Given the magnitude of the other approximations made when modelling the core-mantle system, these omissions are trivial. Values for the core were taken from Nakagawa & Tackley (2004) and comprise:

Parameter	*Value*
Density	1.23×10^4 kgm^{-3}
Specific heat	8×10^2 Jkg^{-1}K^{-1}

Table 4.1: Core parameters

The amount of radioactive heating may be set to decay over time and is based on the half lives of ^{238}U, ^{235}U , ^{232}Th and ^{40}K, the main heat generating radioactive isotopes in the mantle. Radioactive decay is exponential and is calculated using:

$$Q_e = CH \exp^{\lambda}(4.55 \times 10^9 - t) \tag{4.3}$$

for each element with the results summed, where: Q_e is radioactive heat generation for a given element, C is the concentration of the element, H is the heat production, λ is the decay constant and t is time from model start. The age of Earth is assumed to be 4.55×10^9.

A crust factor (F_c) is used to account for radioactive elements being heavily concentrated in the continental crust. The F_c assumes no continental crust at the start (crust factor = 1) but gradually increases as the simulation progresses until a present day factor of 0.66 is achieved using:

$$F_c = \frac{-0.34t}{4.55 \times 10^9} + 1 \tag{4.4}$$

The simulation starts at 4.55 Ga, with the nominal end point as the present. In practice this end point is often not reached due to the variable relationship between model time and Earth time. Values for the elements from Davies (1999) are shown in Table 4.2.

Element	*Concentration (C) (gg^{-1})*	*Heat production (H) (Wkg^{-1})*	λ^* *(s^{-1})*
^{238}U	0.02×10^{-6}	94.35×10^{-6}	1.1013×10^{-10}
^{235}U	0.02×10^{-6}	4.050×10^{-6}	9.8485×10^{-10}
^{232}Th	0.08×10^{-6}	26.60×10^{-6}	0.4948×10^{-10}
^{40}K	80.0×10^{-6}	3.500×10^{-9}	5.5440×10^{-10}

Table 4.2: Radioactive element heat production and decay values. *λ is the decay constant, the inverse of the half life.

The rates of decay of the elements remain in the same relationship with each other, the overall cooling rate is scaled relative to the surface velocity of the initial condition using:

$$R_{sc} = \frac{V_e}{V_m} \tag{4.5}$$

where: R_{sc} is the scaling of the decay of radioactive heating, V_m is model surface velocity and V_e is average Earth surface velocity (5 cm/yr). The rate of radioactive heating decay is multiplied by R_{sc} to produce an appropriate rate of decay for the model case in question. The total amount of radiogenic heating was scaled to match that of the initial 'run in' condition, maintaining the consistency of the model as it begins to cool.

As well as scaling the cooling, the surface velocity can be used to convert model time into equivalent Earth time using:

$$t_e = \frac{t_m V_m}{V_e} \tag{4.6}$$

where: t_e is Earth time and t_m is model time. It is valuable to know 'Earth time' for a model as it allows time scales emerging from modelling to be discussed relative to real observables.

4.3.1 Cases Modelled

The case parameters were chosen using a diagram obtained from parameter space mapping of Cl660 and Ra ranges (Figure 4.1 and Chapter 3). The parameters modelled are shown in Table 4.3. All models were isoviscous, with the reference dynamic viscosity being used to set the Ra. The general unchanged physical parameters are shown in Table 4.4. In terms of resource use; 18 cases were run, 7 of which (high Ra) required 256 processors and $\approx$ 64 GB of RAM each. The remainder (low Ra) required 8 processors and 2 GB of RAM each. All models were run on the Hector national supercomputer, a dual core AMD Opteron based CRAY XT4 machine. The amount of data produced was of order 900 GB, a significant motivation for the work of Chapters 5 and 6, which seek to implement technologies to enable the more efficient storage of such datasets.

4.4 Results

4.4.1 Defining the Critical Ra for Transition

In order to study the nature of the transition from a layered regime to whole mantle convection, the Ra over which this occurs for a given Cl660 must be known. The broad overview of this relationship was mapped out in Chapter 3, here a smaller,

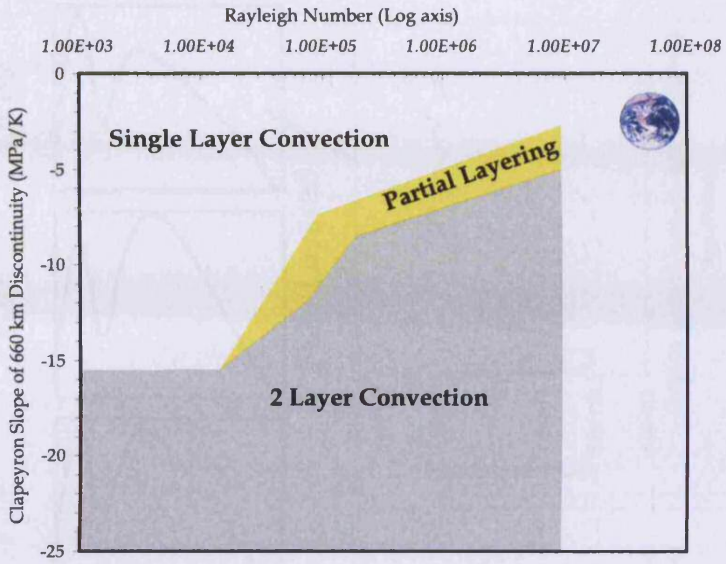


Figure 4.1: Convective mode for Cl660 and Ra.

more focussed, data set is presented. As in Chapter 3, layering is identified by plotting the global average total (both up and down) radial mass flux by layer through the model (after Tackley et al. 1994, Peltier 1996). The plots represent instantaneous values at a given point in the simulation, they are not time averaged. Figure 4.2 demonstrates the relationship at very negative Cl660 (-12 MPa/K) and lower Ra. At such low vigour the transition between layered and whole mantle convection occurs rapidly. Time series data of the surface heat flux are presented for selected cases. There are no large scale differences between the surface heat flux evolution of the different cases, although there is more small scale variability in the cases considered to be layered. The range of Ra over which transition from layered to whole mantle convection is small at low Ra, values are shown in Table 4.5.

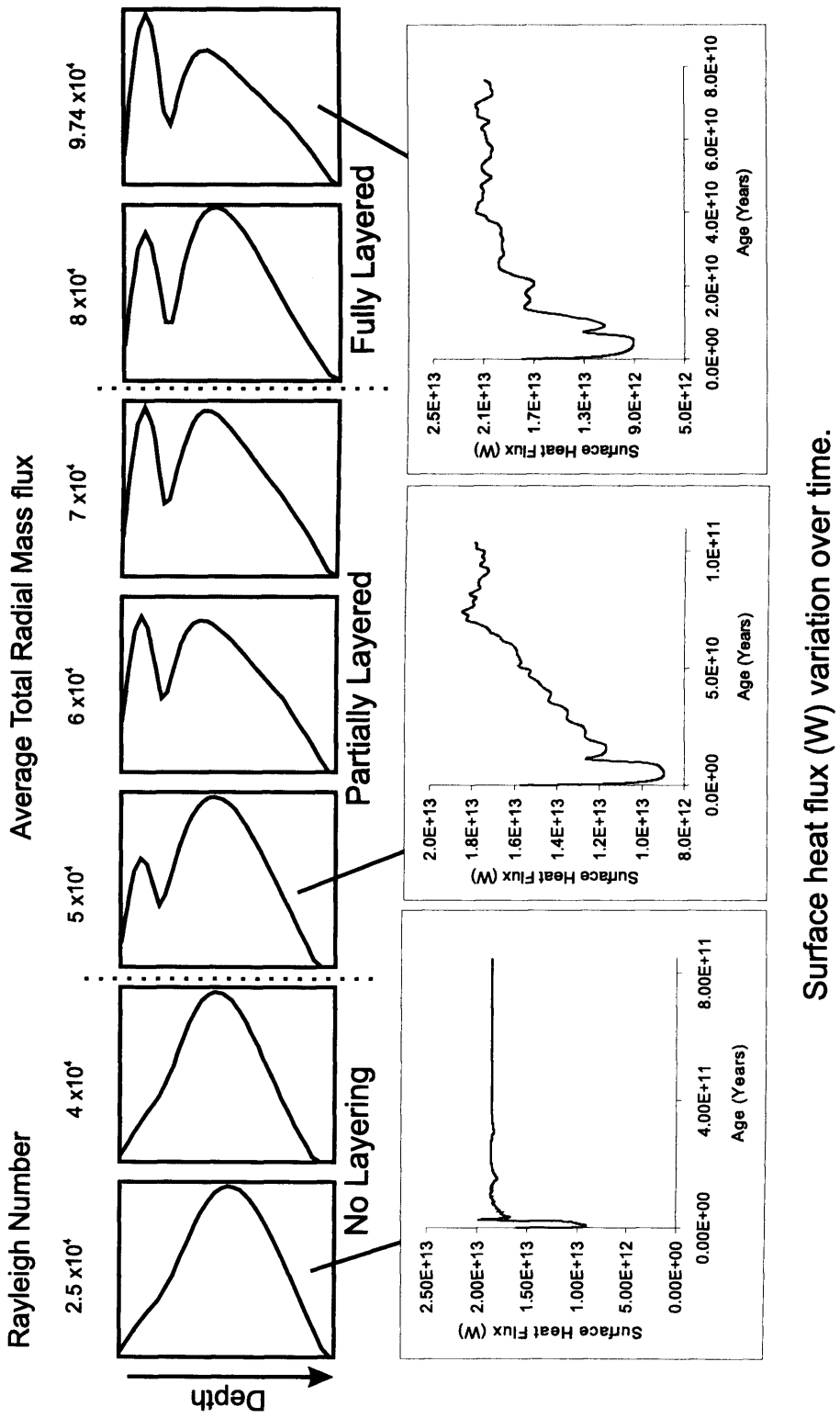


Figure 4.2: Cases demonstrating the effect of a range of low Ra at a Cl660 of -12 MPa/K.

Case	*Cl660 (MPa/K)*	*Ra or Ra range*	*Evolving**	R_{sc}
401	-8	8.66×10^6	No	
402	-8	8.00×10^6	No	
403	-8	7.76×10^6	No	
404	-8	6.76×10^6	No	
405	-8	5.00×10^6	No	
406	-8	6.00×10^6	No	
407	-8	7.76×10^6 - 6.76×10^6	Yes	10
408	-4	8.48×10^7	No	
411	-12	2.50×10^4	No	
412	-12	4.00×10^4	No	
413	-12	6.00×10^4	No	
414	-12	7.00×10^4	No	
415	-12	8.00×10^4	No	
416	-12	5.00×10^4	No	
417	-12	6.00×10^4 - 5.00×10^4	Yes	100
418	-12	5.00×10^4 - 4.00×10^4	Yes	100
015	-12	9.74×10^4	No	

Table 4.3: Summary of cases modelled. *The temperature at the core mantle boundary is allowed to reduce in response to core cooling and radioactive heat generation is allowed to decay. Details in section 4.3.

Figure 4.3 demonstrates the more complex progression of layering regimes at higher Ra and less negative Cl660 (-8 MPa/K). There are 3 phases; whole mantle layering, partial layering and full layering. The surface heat flux data show the behavioural differences between the three phases. Whole mantle convection and full layering are stable, after the initial 'run in' phase where the model develops away from the initial condition. Partially layered cases by contrast, are highly unstable, delivering periodic heat spikes to the surface once the case has developed away from the initial condition. Ra values are shown in Table 4.6.

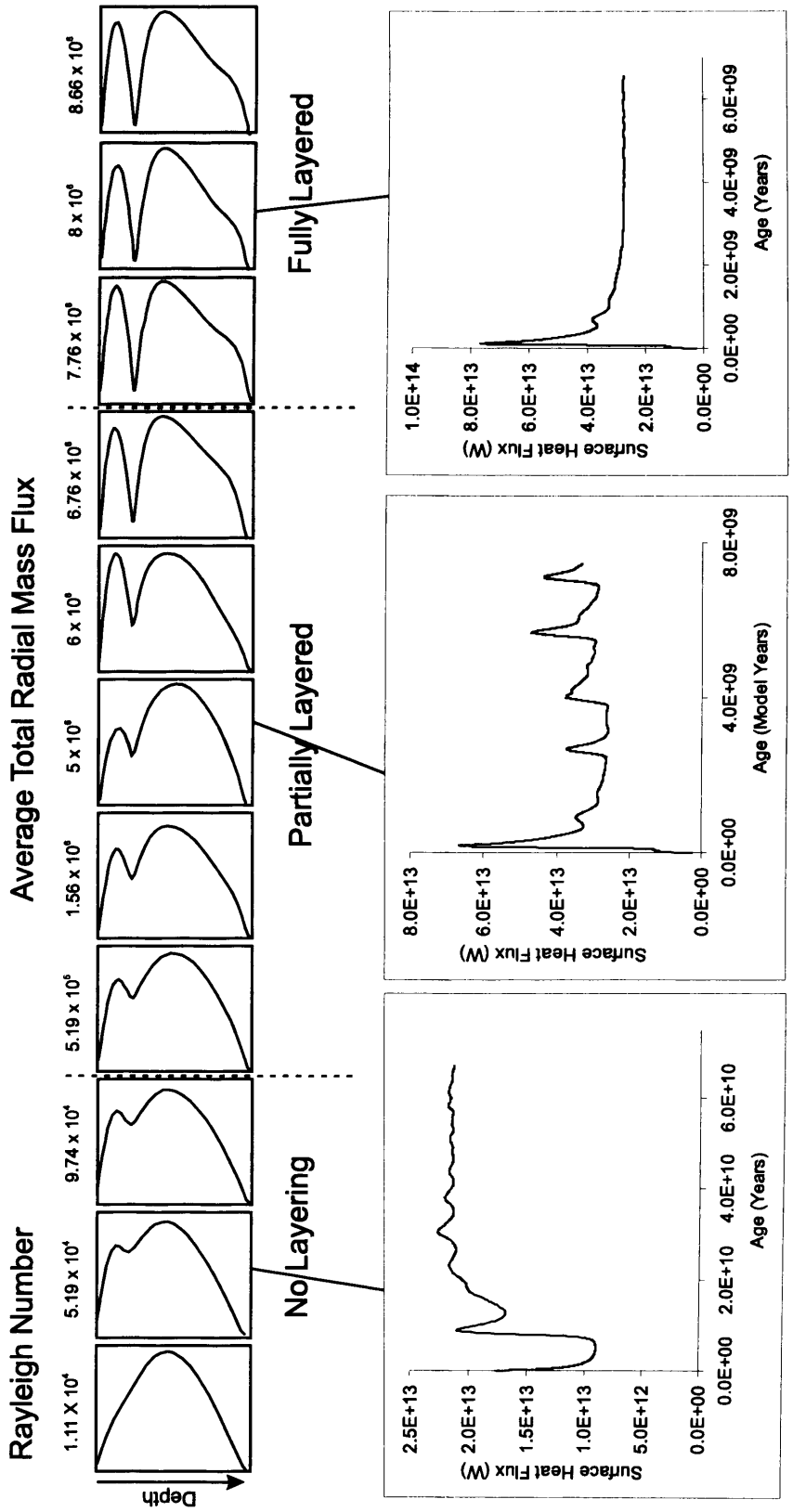


Figure 4.3: Cases demonstrating the effect of a range of higher Ra at a Cl660 of -8 MPa/K.

Parameter	*Value*
Equation of State	Incompressible and Boussinesq
Reference density	4.5×10^3 kgm^{-3}
Gravitational acceleration	10 ms^{-2}
Volume coefficient of thermal expansion	2.5×10^{-5} K^{-1}
Thermal conductivity	4 Wm^{-1}K^{-1}
Specific heat at constant volume	1.0×10^3 JK^{-1}kg^{-1}
Temperature at outer shell boundary	3.0×10^2 K
Temperature at inner shell boundary*	2.85×10^3 K
Radioactive heat generation**	5×10^{-12} Wm^{-3}
Boundary conditions (velocity)	free slip
Inner radius of spherical shell	3.480×10^6 m
Outer radius of spherical shell	6.370×10^6 m

Table 4.4: Common input values used for all cases. *Permitted to reduce in cooling cases. **Reduced in cooling cases.

Transition	*Ra*
Whole mantle convection to partial layering	4×10^4 - 5×10^4
Partial layering to layered convection	$\approx 7 \times 10^4$

Table 4.5: Ra values for the various transitions at Cl660 = -12 MPa/K and low Ra.

4.4.2 Manner of Transition

Figure 4.4 illustrates surface heat flux over time for selected cases with Cl660 = -8 MPa/K. Periodic heat pulses occur as cases enter the partially layered regime. The interval between these pulses can be converted into real Earth time (Equation 4.6). Table 4.7 summarises the converted intervals between pulses in 3 cases where models were allowed to run for a significant period of time, recording multiple pulses. Intervals reduce as Ra reduces and are of order 100-200 Ma. The interval is linked to the Ra, higher Ra resulting in longer intervals with a scaling of: Interval (yrs) = cRa$^{1.2}$ (where c is in the range 0.5-1.5). Figure 4.5 shows a partially layered case with

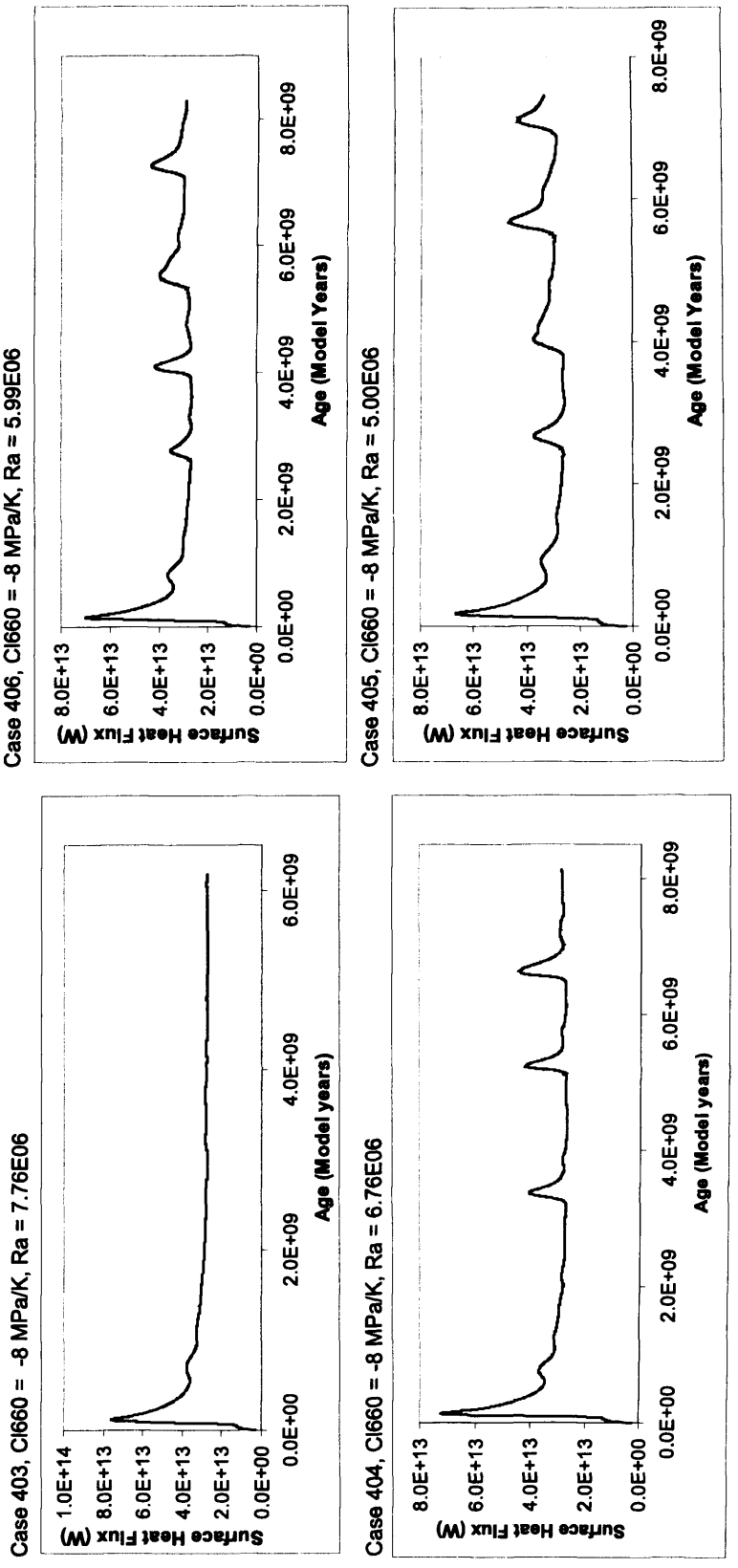


Figure 4.4: How pulsing emerges as Ra reduces. Note periodic nature of pulses.

Transition	*Ra*
Whole mantle convection to partial layering	9.74×10^4 - 5.19×10^5
Partial layering to layered convection	6.76×10^6 - 7.76×10^6

Table 4.6: Ra for the transitions at Cl660 = -8 MPa/K and high Ra.

more Earth-like Ra and Cl660, which also exhibits heat pulses. Time slices of the radial mass flux demonstrate how each pulse is preceded by a significant reduction in the degree of layering in the system.

Case	*Ra*	*PI (Model Myr)*	*S Vel. (cm/yr)*	*SI (Myr)*
404	6.77×10^6	1860	0.6342	236
		1400	0.6342	178
Average				**207**
406	5.99×10^6	1310	0.5118	134
		1480	0.5118	151
		1700	0.5118	174
Average				**153**
405	5.00×10^6	1370	0.4650	127
		1640	0.4650	153
		1430	0.4650	133
Average				**138**

Table 4.7: Scaled thermal peak intervals, assumed surface velocity for Earth: 5 cm/yr. Source of pulse interval values, Figure 4.4. PI - Pulse Interval, S Vel. -Surface Velocity and SI - Scaled Interval.

The process which leads to a surface heat flux pulse is visualised in detail in Figure 4.6 and as snapshots in Figures 4.5 and 4.8. An animation of the avalanche-pulse mechanism can be found in Appendix B. The heat pulse is driven by a single strong downwelling, often described as a mantle avalanche (Tackley et al. 1993). The avalanche initiates as a point downwelling from 660, which rapidly grows, sucking in most of the

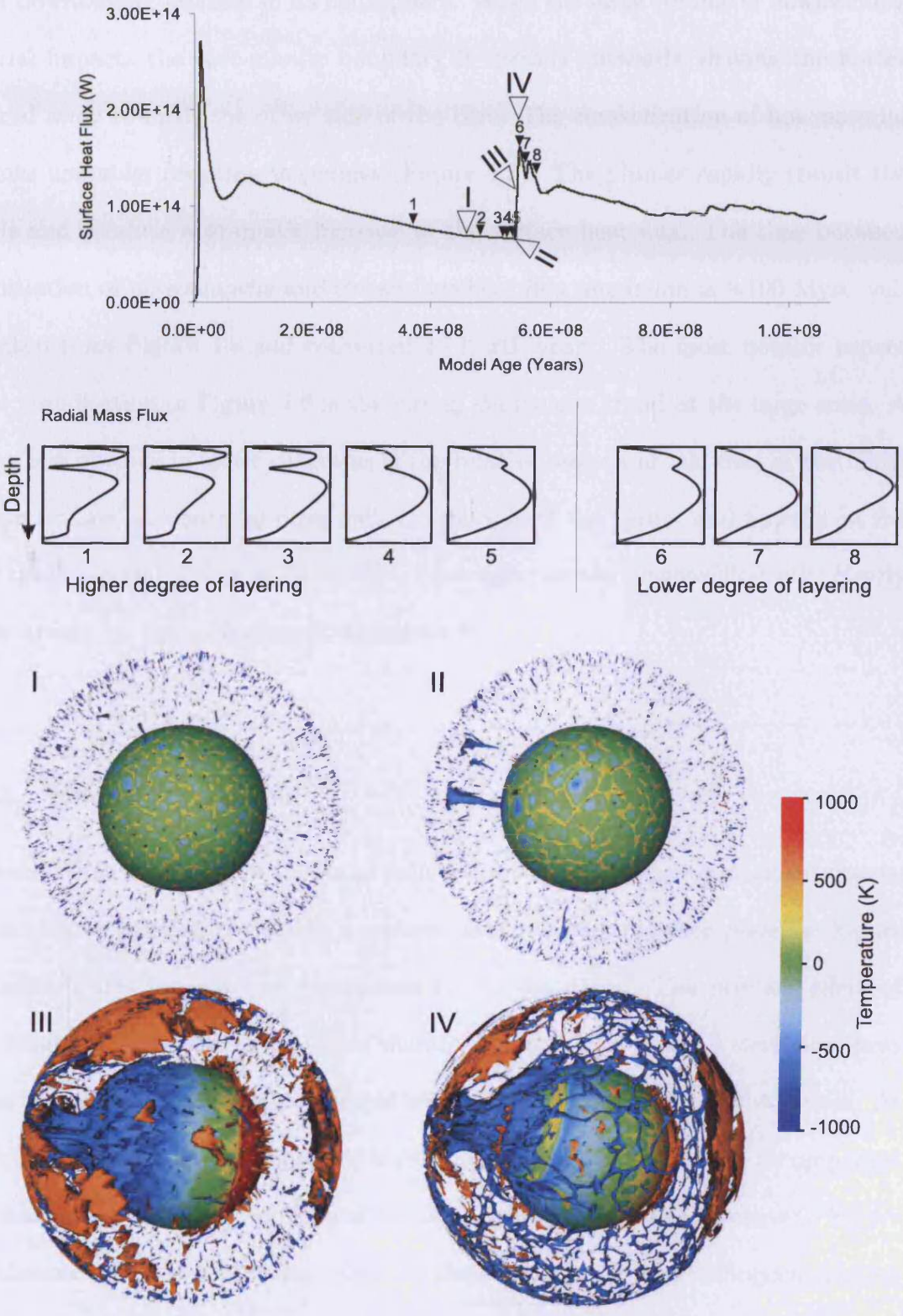


Figure 4.5: A demonstration of pulsing in a partially layered system. Case 408, more Earth-like, Cl660 = -4 MPa/K, Ra = 8.48×10^7. Note the long 'tail' following the spike in the surface heat flux graph.

major downwelling features in its hemisphere. When the large volume of downwelling material impacts the core-mantle boundary it spreads outwards, driving the hotter material away towards the other side of the core. The concentration of hot material becomes unstable, resulting in plumes (Figure 4.5). The plumes rapidly transit the mantle and produce a dramatic increase in the surface heat flux. The time between the initiation of an avalanche and the surface heat flux maximum is ≈100 Myrs, values taken from Figure 4.8 and converted to Earth years. The most notable aspect of the visualisation in Figure 4.6 is the strong degree-one trend at the large scale. A feature less obvious in static diagrams is the relative motion of material at the CMB and the surface. As material downwells on one side of the planet and upwells on the other there is a global flow at the surface converging on the downwelling side; clearly demonstrated by the animation in Appendix B.

4.4.3 Evolving models

Models with cooling allow the Ra to reduce as a run progresses, transitions between different convective regimes within a continuous model case are then possible. Figure 4.7 demonstrates the effect of transitions in low Ra cases. The primary effect of the transition from layered to whole mantle convection is a small deviation from the overall cooling trend at the point of transition, highlighted by dotted boxes. At higher Ra (Figure 4.8) the impact of transition into partial layering is accompanied by sustained and significant pulsing of the surface heat flux. The grey curve in Figure 4.8 indicates the surface heat flux trend for the same case without radiogenic heating

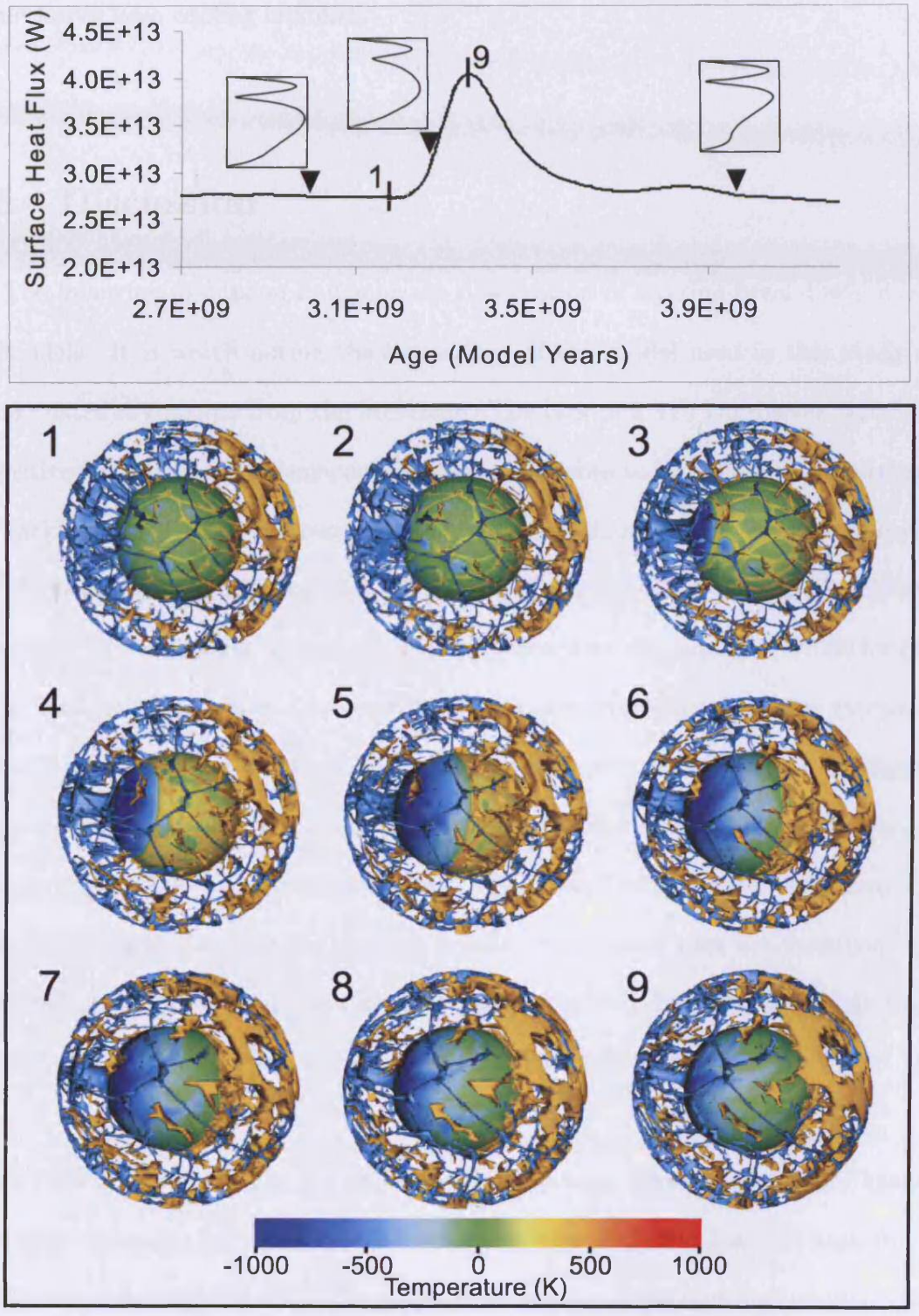


Figure 4.6: Pulse mechanism visualised, mantle temperatures are shown as a surface at the core-mantle boundary and as +400 and -400 K isosurfaces. (1) shows the state of the mantle prior to initiation of the heat pulse, (9) indicates the state of the mantle at maximum surface heat flux. Case 404, Cl660 = -8 MPa/K, Ra = 6.76×10^6

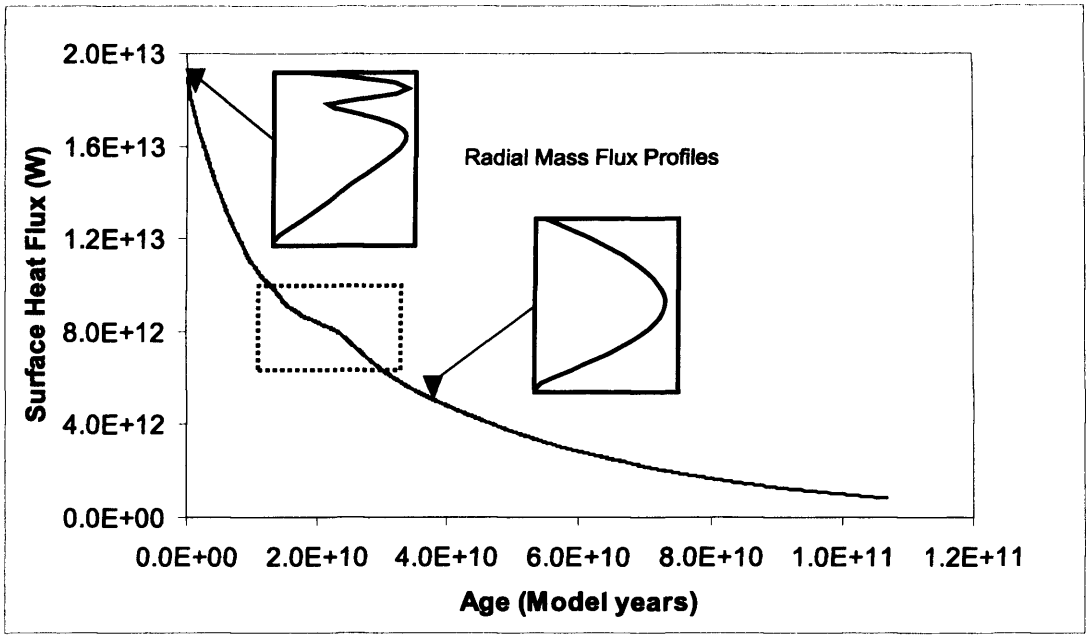
reduction or core cooling enabled.

4.5 Discussion

The following discussion builds on the observation of layering breakdown in mantle models. It is worth noting the limitations of the model used in this study and some related arguments from the literature. The lack of a 410 km discontinuity with a positive Clapeyron slope may cause mantle avalanches to be enhanced (as suggested by Tackley et al. 1994), however there may be less than straightforward differences in behaviour between 410 and 660 (Tauzin et al. 2008), which this study wished to eliminate. The reasoning behind enhancing the absolute magnitude of Cl660 for modelling was presented above (section 4.3), it is reassuring that the pulse mechanism was still observed in cases with high Ra and a Cl660 of -4 MPa/K, the least negative value used in this study. Korenaga (2008) proposed that plate tectonics may impose a large length scale on mantle convection, effectively forcing material through 660. This study was not designed to test this possibility as plates were not modelled. This is not necessarily a weakness, as this study seeks primarily to understand early Earth, a time when the exact mode of plate tectonics is largely unknown.

With regard to the critical Ra for transition between convective regime, there is a notable difference between behaviour displayed by models at low and high Ra. At low Ra and very negative Cl660 the range of Ra where partial layering occurs is small, with a rapid change from whole mantle convection to layered mantle convection. By

(a) Case 417, Cl660 = -12 MPa/K, Ra = 6.00E+04 - 5.00E+04



(b) Case 418, Cl660 = -12 MPa/K, Ra = 5.00E+04 - 4.00E+04

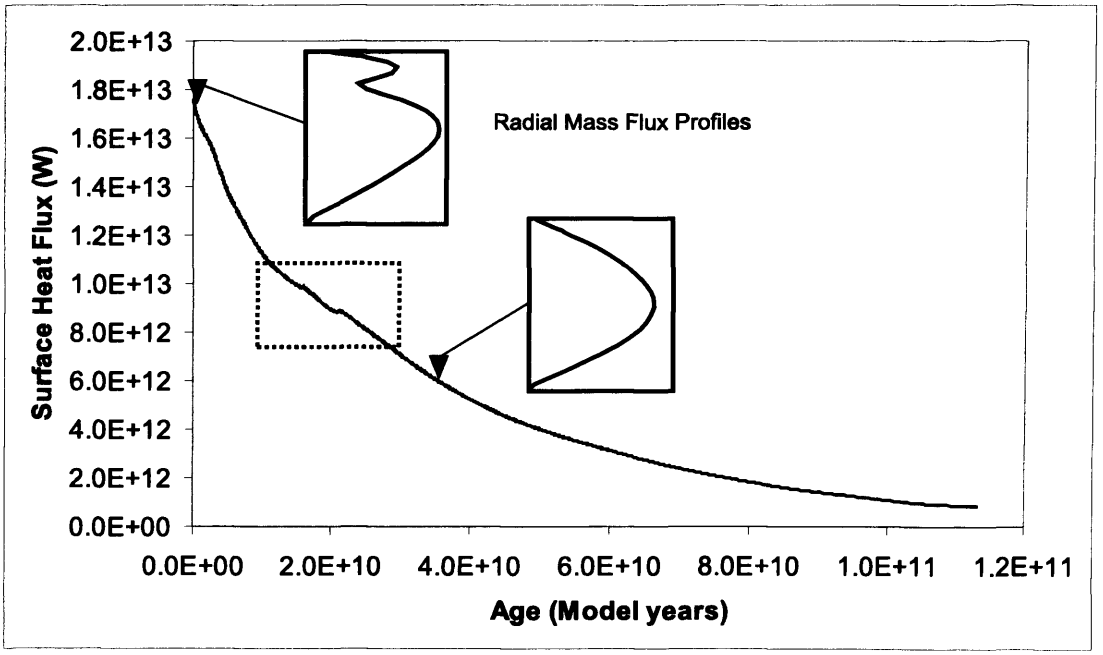


Figure 4.7: Low vigour model cases with cooling. (a) Focussing on transition from layered convection to whole mantle convection. (b) showing transition from partially layered convection to whole mantle convection. Zero time refers to the start of cooling, the cases were 'run in' before this to allow them to develop away from the original initial condition. Dotted boxes highlight the points of transition.

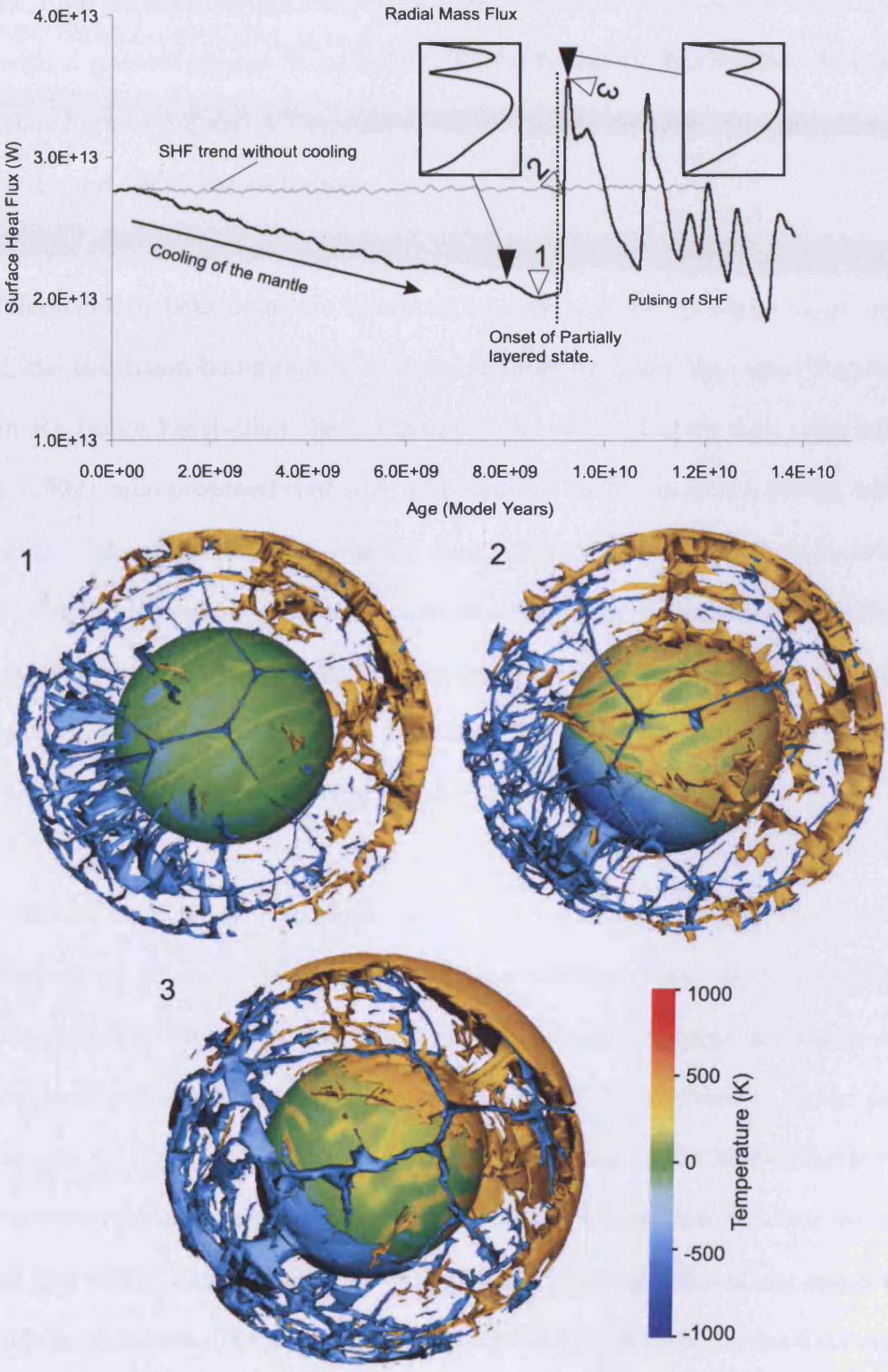


Figure 4.8: High vigour model case with cooling. Zero time refers to the start of cooling, the case had previously 'run in' to allow it to develop away from the initial condition. Case 407, Cl660 = -8 MPa/K, Initial Ra = 7.76×10^6

contrast, high Ra and less negative Cl660 cases display a wide domain of partial layering with a gradual change from layered to whole mantle convection. It must be noted that Figures 4.2 and 4.3 represent 'static' models defining the convective state of a set Ra and Cl660 for each case.

The differences in behaviour are mirrored to some extent by those cases allowed to cool, the transition being relatively unremarkable for lower Ra cases (Figure 4.7). At high Ra (more Earth-like) the transition is dramatic (Figure 4.8), contradicting Allègre (2002), who proposed that such a breakdown could occur in a benign manner. The nature of the transition from layered convection to partial layering is therefore to some extent Ra dependent, with higher Ra resulting in larger fluctuations in the surface heat flux. An important point to note is the relative temperatures of the cases. Low Ra cases are significantly hotter than high Ra cases as they are less efficient at losing heat, convection being more sluggish.

The results of layer breakdown observed in evolving cases represent the same process observed as pulses in partially layered non-evolving cases. It is the transition to partial layering caused by the reduction in convective vigour (cooling), which produces heat pulses. There are some differences in the expression of the pulses. For example, in Figure 4.8 the first 2 pulse peaks occur $\approx$340 Myrs (Earth years) apart but are not separated by an extended return to complete layering as in the non-evolving cases (Figure 4.4). The main pulses are then followed by much lower amplitude fluctuations. The most likely cause of this difference is the constant cooling

of the model. The progression from full layering to partial layering and on to near whole mantle convection does not allow the model to re-equilibrate at any given state.

The mechanism of breakdown has been visualised for three cases and remains consistent, indicating that it is a robust process, capable of operating at a variety of Ra and Cl660 combinations. A sketch representation (Figure 4.9) illustrates the global nature of mantle avalanches and their corresponding heat pulse. A superficially similar process has been modelled in spherical geometry by Zhong et al. (2007), who observed strong degree 1 up and downwelling features in response to continental insulation of the surface. It is important to note that the mechanism behind Zhong, Zhang, Li & Roberts's degree 1 convective pattern is entirely different to the mechanism in this work. The transit time of ≈100 Myrs for an avalanche to descend and plumes to rise is internally consistent with model velocities, which suggest ≈60 Myrs average one-way transit time. This agrees broadly with previous estimates for mantle transit time (Tackley 2000).

Table 4.7 demonstrates that the interval between heat pulses is of order 100-200 Myrs. This broadly agrees with weak cyclicity found by analysing LIP's. Intervals between 105 Myr and 730 Myr were found for the last 3.5 Gyr and ≈170 Myr interval for the last 1.5 Gyrs (Prokoph et al. 2004). This link is somewhat correlative and does not in itself indicate a mechanistic connection. Davies (1995) also finds time scales of order several hundred Myrs for partially layered cyclicity, using a parameterised model. The trend of reducing intervals between pulses as Ra reduces suggests that

the 'pulsing' behaviour of the models is being driven by cold material descending from the surface of the model as suggested by the parameterised models of Davies (1995). This indication is borne out by the visualisation of a cold avalanche triggering the heat pulse. Therefore, an expected trend in the long term cooling of the Earth could be for mantle avalanches and associated effects to become more frequent but less dramatic over time. A scaling of pulse interval with Ra of: Interval (yrs) = $cRa^{1.2}$ (where c is of order 1 for the cases presented) underlines the dependence of peak interval on Ra. The scaling is based on cases with a significantly more negative Cl660 than Earth at -8 MPa/K. Thus, for Earth, c is likely to be less than 1. The scaling also becomes invalid once Ra exceeds the value required for 2 layer convection.

Mantle avalanches have been previously described (Tackley et al. 1993, 1994, Butler & Peltier 2000), similar avalanches are observed in this work. However, due to the greater computing resources available for this study, simulations could be run at high resolutions with much more Earth-like Ra for a longer time. For example, Tackley et al. (1993) show the development of their model over 2 Gyrs of model time. Taking their model surface velocity of 0.6 cm/yr and converting to Earth-time (Equation 4.6) the resulting 240 Myrs is not sufficiently long to observe the cyclic nature of mantle avalanches. Evidence of upwelling and subsequent surface heat flux maxima in response to avalanches, was not found by Tackley et al. (1993). Again, it is likely that the relatively short time evolution of their model case was responsible.

Adopting the assumption that some form of phase change induced layering is likely, at

least for the early Earth, the implications of layered to partial layered transition and avalanche induced heat pulses are significant. Possible models are: a varying degree of partial layering for all of Earth history or full layering for a period, followed by a move to partial layering. Full layering for all of Earth history can be discounted as seismic tomography studies clearly show material passing through 660 km (Van der Hilst et al. 1997, Montelli et al. 2004). The two remaining models are similar in that the major process of interest is the pulsing introduced by partial layering. This either occurs over the entire history of Earth, or begins when layered convection starts to break down. The model of a single event transition from layered to whole mantle convection is effectively ruled out by this study, given the wide range of Ra over which partial layering occurs.

Parameterised modelling of the thermal evolution of Earth often encounters thermal catastrophe, where the temperature of the mantle tends towards unreasonably high values well before 4.5 Ga (reviewed in greater detail by Korenaga 2008). In a partially layered mantle, heat pulses could significantly disrupt the central assumption of parameterised modelling that heat flux decays consistently over time. When the cooling rate (heat flux) of the planet is highly time dependent, typified by a series of maxima followed by accelerated cooling, assigning a consistent cooling rate is likely to produce inaccurate results. It is possible that the heat flux at the present day would plot on the 'tail' of a peak as illustrated by Figure 4.5. Therefore, when measurements are projected back in time they would produce an artificially high 'average' heat flux, leading to thermal catastrophe in a parameterised model. It must

be noted that there are other methods of averting thermal catastrophe. For example, Labrosse & Jaupart (2007) employed the assumption that thermal losses from ocean lithosphere have remained reasonably stable from 3 Ga to present, despite higher mantle temperatures in the past. The assumption prevents thermal catastrophe in the models before 3 Ga. It also suggests that plate tectonics did not operate in its current manner prior to 3 Ga. The results of the work presented here also suggest that mantle convection might have operated in a different manner (more layered) in the distant past.

The mantle avalanche mechanism has been observed to produce converging surface motion towards the downwelling. In the models presented here, the surface velocity boundary condition is free slip. On Earth, plates lie above the mantle and are coupled to its motion. With an avalanche starting at 660 km, one might expect the rapid inward motion of the mantle in response to this dramatic event to act on the plates as shown in Figure 4.9. Whether such a process could explain the periodic accumulation of supercontinents would require further investigation. The interval between avalanche/pulse events in the models is a few 100 Myrs, this is of the same order as the supercontinent cycle, an exciting but rather speculative possibility.

A further impact of pulse events on plate tectonics would be expected from the hot plumes triggered by an avalanche. From both the visualisation and surface heat flux data (Figures 4.5, 4.6 and 4.8) it is evident that a significant amount of heat is delivered to the upper mantle by a pulse. The heat is however, not delivered as

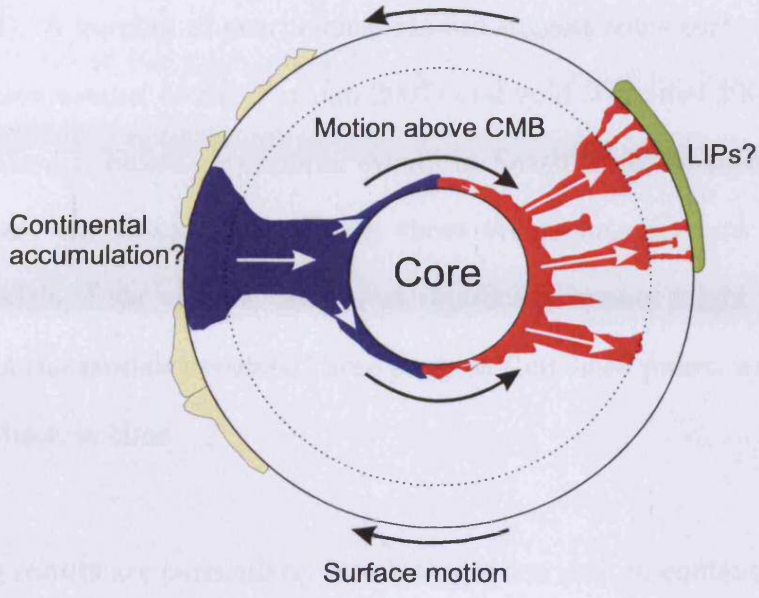


Figure 4.9: Sketch of layer breakdown and subsequent heat output. (not to scale)

a single plume, rather is is more diffuse and takes the form of many smaller plumes spread over much of the hemisphere. On Earth, this could produce clustering in time of Large Igneous Province (LIP) eruption, illustrated in Figure 4.9, which are broadly associated with one hemisphere at the time of emplacement. Based on Earth times of ≈100 Myrs for a pulse to reach maximum, the clustering would be expected on this timescale. LIPs do indeed appear to show some clustering in time (Ernst & Buchan 2002), again raising tantalising possibilities.

Large scale global magmatism would of course have an observable geochemical effect. A probable outcome of the pulsing would be a certain degree of periodicity in the level of melt extraction from the mantle. Expected effects might be continental building episodes and the transfer of trace elements into the upper mantle (Stein &

Hofmann 1994). A number of geochemical studies suggest some sort of periodicity in 'mantle depletion events' of He (Parman 2007) and gold (Frimmel 2008). If this geochemical variation is linked to periodic events in Earth's mantle such as avalanches and heat pulses, one would have to take these events into account when building conceptual models of the mantle. The most significant impact might be expected in the Archean as the models presented here suggest that heat pulses would have been larger further back in time.

The modelling results are particularly interesting when seen in context with the ideas of Condie (1998), who provides a detailed conceptual description of potential mantle avalanche mechanisms. Based in geological and geochemical observations, Condie makes a number of suggestions. He proposes that supercontinents fragment over large upwellings and collect over large downwellings, but that plumes tend to follow sometime after breakup. In his model, plumes are triggered by downwellings reaching the core mantle boundary. This global process is similar to the ideas presented in Figure 4.9. Condie also proposes evolution of layering over time, with the mantle more strongly layered before 1 Ga with only 3 major avalanches in the Archean. The reduced magnitude and increased frequency of avalanche events after 1 Ga is attributed to the reduction of Ra, limiting the effectiveness of Cl660 as a barrier to flow. Again, the models agree well with this time dependent trend. Finally, Condie suggests that the duration of Precambrian events is of order 50-80 Myr and that there is a clustering of plumes in time, statements this modelling work is entirely consistent with. Differences between this work and that of Condie arise in the mechanism for repeated

avalanching. Condie suggests a large role for subducting plates and continental insulation, whereas this work focusses solely on the influence of the 660 km phase change. Therefore, while many of the observations in this work match those of Condie (1998), his approach is more surface driven, an aspect that was not covered in the modelling. Further differences arise in the timescales of avalanche events. The modelling work would suggest 100-200 Myr intervals between avalanches, Condie's 'superevent cycles' vary between 800 Myr and 400 Myr. There are indications that model cases run at early Earth Ra (10^9) would produce longer intervals between pulses, for example the long tail of Case 408 (Figure 4.5). Despite the caveats outlined above, the level of agreement between modelling results and the overall conceptual model of Condie is remarkable.

A number of possible impacts of periodic pulsing of the mantle have been discussed. It is not certain that this mechanism has been active on Earth as there are a number of simplifying assumptions to the models, which could have an unforeseen impact. For example, the model is isoviscous radially, does not simulate surface plates or allow temperature dependent viscosity. However, evidence for periodic behaviours emerging from geochemical studies, problems with simple parameterised thermal evolution models and the need to reconcile observations favouring both layered and whole mantle convection, suggest something is missing from the story of Earth's mantle. The evolution of a partially layered regime could be part of the answer.

4.6 Conclusion

- The transition from layered mantle convection to partial or whole mantle convection at high Ra is characterised by significant periodic fluctuation of surface heat flux, driven by mantle avalanches. This contradicts the suggestion in the literature that such a transition could go almost unnoticed (Allègre 2002).

- Mantle avalanches initiate at the 660 phase boundary as a result of cold material ponding above. The effect of an avalanche is to sweep hot material together, spawning significant plumes.

- The mechanism of periodic layer breakdown has been shown to be consistent across a range of Ra and Cl660, supporting the conclusions of previous parameterised modelling (Davies 1995).

- For partially layered cases, repeated discrete avalanches and subsequent 'heat spikes' are observed, the interval between these events appears longer at higher Ra for a given Cl660. The interval is 100-200 Myrs, of similar order to the time scale of supercontinent formation and the interval between large igneous provinces.

- Partial layering, leading to periodic 'pulsing' of the surface heat flux provides a mechanism for global episodic mantle depletion events, inferred from geochemical observations.

- If Earth's mantle was initially layered and has transitioned to partial layering or even whole mantle convection over its history, models of the thermal history

and geochemical evolution must take this into account.

Chapter 5

Distributed Storage of High-Volume Environmental Simulation Data: Mantle Modelling

5.1 Abstract

Advances in the modelling of Earth's mantle often face limitations due to the volume of output data that can be stored. To address this limitation, a feasibility study of a peer-to-peer distributed storage system for the archiving of large datasets produced by the Earth mantle modelling code TERRA is presented. The manner in which the nature of such data affects the indexing, duplication and performance requirements of such a system is analysed and a data-oriented overlay network to improve efficiency is proposed. Duplication methods are analysed and small scale test bed performance measured.

Glossary

Terms, which may not be familiar to some readers (computer scientists will probably want to start at section 5.2).

- Hashing, using a mathematical algorithm on a file to produce a short numerical or alphanumerical string, which is dependent on the content of the file. If the file changes then using the same algorithm will result in a different hash, alerting the user.
- Distributed hash table, a method of using file hashes to map files to particular resources in a structured way ie. in a network.
- Overlay Network, a virtual network in which connections between nodes do not have to match physical connections between the machines in question. The web is an overlay network.
- API, a library of pre-written code, which can be used by an object oriented programming language like Java to build systems without having to write every single section of the code oneself.
- JRE, Java Runtime Environment, the virtual machine, which allows one to run Java bytecode on a system.
- P2PS, Peer to peer simplified, a API for writing peer to peer programmes in Java.

- Peer, a peer is the basic entity of a peer to peer system, the term is used to refer to individual nodes in a network and usually comprises a programme running on a given machine. A rendezvous peer is a peer with a known address, which can be contacted by other peers for information to further peers.

- Instantiation, the act of creating a new instance of an object such as a Java class, this is a key operation of object oriented programming, which allows existing code to be extended in a structured manner.

5.2 Introduction

Many large-scale scientific instruments such as the LHC at CERN, telescope arrays and supercomputing clusters are rapidly converging on a single abstract definition, a device with the potential to produce bytes in almost limitless quantities. This has been termed the 'data deluge' (Vazhkudai et al. 2005), which is likely to impact on almost every scientific discipline. This chapter addresses that issue through the storage and indexing requirements of data produced by the TERRA Earth mantle simulation code (Baumgardner 1985, Bunge & Baumgardner 1995, Yang & Baumgardner 2000, Oldham & Davies 2004), primarily run on beowulf-style supercomputing clusters in the 256-2048 processor range. Large though the storage attached to such machines may be, the amount available to a single user is usually limited. When running TERRA at 10 km resolution (1024 cores, ≈300 GB RAM), the frequency of visualisation output is limited and long runs may not be possible, even when the runs are within the computational capacity of the cluster. Additionally, as model output data

remains in a single users account, it is difficult to ascertain if a model with particular parameters has already been run. The result is unintended duplication of models, which at high resolutions becomes expensive. The storage system is required to be searchable and to allow selective retrieval of data, precluding the use of tape storage.

The TERRA model is scalable depending on the computing resources available. A typical base resolution may only produce several gigabytes (GB) of output. However, every step-wise improvement in resolution is accompanied by an 8-fold increase in output. At high resolution, a single model run produces over 600 GB of output. When a suite of models are run, the level of data output easily outstrips the storage capacity of smaller clusters, considered here as comprising 200-300 processors. The continued growth of computing power will only increase this flow of data. Although storage technologies are generally improving at a similar rate as computing power, most researchers share their supercomputing clusters with a relatively large number of users. Storing all data output on the cluster is not usually possible. Fortunately, many institutions have large numbers of workstations in computing laboratories where only a fraction of the local disk is in use. Assuming 30 GB of free space per machine, a 60 machine lab has 1.8 TB of storage available. Typically there are 10s of such labs, upgraded on a rolling 3 year cycle, with new workstations having at least twice the amount of local storage as the previous generation. Therefore, a network of 1000 machines where 1/3 are upgraded each year will gain 9-10 TB of storage every year for 3 years and 19-20 TB per year in the following 3 years. The potential for a system able to exploit these resources is significant. The performance and reliability of the

operating environment was evaluated, taking account of the inherently untrustworthy nature of open access computers on an institutional network. The novelty of this work arises from the integration of the storage mechanism with the structure of the data, to allow more efficient operation.

5.3 Related Work

The concept of storage over a number of physical resources itself is not new, search engines, high traffic web pages and grid computing all implement distributed storage for reasons of availability, performance and integrity of data. More 'informal' distributed computing is a more recent concept and consists of systems such as Gnutella for file sharing and Freenet, which allows anonymous storage within an overlay network on the Internet.

5.3.1 Centralised and Larger-Scale Systems

Many distributed storage systems are deployed in a Grid context, Storage Resource Broker (SRB) is a typical example. SRB is aimed at storage and sharing of terabytes-petabytes of data. Developed at the San Diego Supercomputer Centre, SRB is middleware that acts as a link between geographically separated data stores (Rajasekar et al. 2002). SRB uses a federated architecture with metadata catalogues (MCAT) that are local to a particular data store and enable access to the stored data. As with many such systems, the underlying implementation is hidden from the user and files are accessed using a global persistent identifier (a unique 'name'). SRB also utilises data duplication, caching and load balancing (Rajasekar & Wan

2002). OceanStore (Rhea et al. 2001) is a research project aimed at building a global data store with a multi-exabyte storage capacity. Large server pools are distributed globally in peer to peer relationships, data is freely shared but encrypted. Implementations of OceanStore use: erasure coding, secure hashing with a distributed hash table and dedicated routing control. In addition, OceanStore is self-healing through the use of update cycles.

5.3.2 Peer to Peer

Peer to peer (P2P) networking emerged with the rapid uptake of systems such as Napster. First generation networks relied on index servers (in a similar manner to SRB) and one-to-one transfers. Subsequent generation P2P systems such as BitTorrent (Cohen 2003) exploit the parallelism inherent to P2P networks by sharing files as fragments. The result is cost effective distribution of data to a large number of users (Yang & Veciana 2004). Current generations of P2P software such as Gnutella and Freenet are intended to be completely decentralised to avoid single point failure. Other techniques, such as distributed hash tables, enable structured overlays to be built by a groups of peers using strict algorithms relating data and its location within a network, without the need for a central controller.

The P2P approach to distributed systems assumes an environment of 'edge' workstations, where unreliability is expected. This situation is similar to an institutional network, where hardware is accessible to a large number of users and cannot be considered trustworthy. A number of storage oriented systems based on P2P technologies

have been proposed, many intended as a form of fault tolerant backup for user files. Backup systems include SNS (Ye et al. 2003), Farsite (Adya et al. 2002) and pStore (Batten et al. 2001). These three systems implement methods such as; clustered architecture, a versioning system, data replication and routing algorithms to map data onto resources. However, none of these systems are intended to store and index structured scientific data. FreeLoader (Vazhkudai et al. 2005) is intended for scientific data sets of several terabytes (TB). The assumptions made by FreeLoader are that data is written once then accessed multiple times, primary copies of data are backed up elsewhere and that scientific workstations are often of higher specification than standard machines. FreeLoader implements striping by breaking files into 'morsels', which can provide superior performance to that of the local hard disk (Vazhkudai et al. 2005). However, it is primarily a high performance distributed file system, not a data archive. An emerging concept in P2P systems is the use of overlay networks to enhance search efficiency (Andersen et al. 2001, Stoica et al. 2001). In the scenario discussed here, the most useful aspect is the trend toward data-oriented architectures (Aberer et al. 2005), in which metadata are used to virtually group similar data to enable more efficient routing of queries.

The development of P2P systems is assisted by specific P2P APIs, which provide a framework on which to build. Peer to Peer Simplified (P2PS) (Wang 2005), is a set of P2P protocols implemented as an API. The main implementation of P2PS is in Java and contains sufficient classes to develop many different types of P2P application. In this respect, P2PS is similar to JXTA (Gong 2001), the main differences

being technical rather than conceptual. The key feature of P2PS relevant to this project is that customised internal message types (as XML Advertisements) can be easily defined. Communication between peers is based on abstract pipes, thus an apparent direct connection between two peers may in fact pass through a number of intermediate peers. The P2P infrastructure provided by P2PS is based on subnets of standard peers connected by rendezvous peers, which allow communication across subnets. This type of centralised-decentralised architecture allows greater scalability and efficiency of the network as a whole (Ramaswamy et al. 2005).

Key features taken from the above systems, which have been incorporated into the system concept are; duplication for reliability, data striping, indexing, data-oriented overlays and assumed unreliability of peers. P2PS was used as a basis for the test bed implementation.

5.4 Mantle Modelling - TERRA

5.4.1 What is TERRA?

The Earth's mantle is defined as the region of predominantly solid rock located between the generally more buoyant crust and the more dense liquid iron outer core. It is around 2900 km thick and makes up over 80% of the planetary volume. Although the mantle behaves as an elastic solid, transmitting seismic waves on short timescales, on a geological time scale its behaviour is analagous to a highly viscous fluid, allowing convective heat transfer from the interior to the surface. TERRA is a finite element

3D mantle convection simulation code (Baumgardner 1985, Bunge & Baumgardner 1995, Yang & Baumgardner 2000, Oldham & Davies 2004). Earth's mantle is treated as a spherical shell of highly viscous fluid with physical properties based on inversion of observations of the real mantle, lab experiments and *ab-initio* mineral physics calculations of likely mantle minerals. It solves the equations for the conservation of mass, momentum and energy and steps forward the evolution of the mantle system. Appropriate boundary conditions are prescribed at the surface and the core-mantle boundary. The code is parallelised with MPI and is scalable depending on resources available.

The model resolution is referred to by the value of the parameter 'mt'. The mt parameter relates to the manner in which the finite element mesh is divided. The mesh is based on a regular icosahedron, where each of the 20 sides of this regular polyhedron is an equilateral triangle. These equilateral triangles are paired to give 10 diamonds. The diamonds are treated as quasi-squares and are repeatedly subdivided. The number of subdivisions, which must be a power of 2, along each side of a diamond is given by mt. For Earth simulations the number of radial layers (nr) is set to

$$nr = \frac{mt}{2} + 1 \tag{5.1}$$

to maintain a reasonable element aspect ratio. At each time-step, five values are solved for at each grid node; temperature, non-hydrostatic pressure and the three components of velocity. A simulation may be run for many 10,000's of time steps. Note that the number of unknowns increases by a factor of 8 for each factor of 2 rise in mt. Table 5.1 summarises some typical mt values. As TERRA is research code, it

does not have a GUI and is executed from the command line, a recompile is usually required before each run. It is however, extremely computationally stable. Figure 5.1 demonstrates a typical work flow for running a TERRA job.

mt	*Mean Resolution (km)*	*No. of nodes*
64	88	1,351,746
128	44	10,649,730
256	22	84,541,698
512	11	approx. 655,000,000

Table 5.1: Resolution for different values of mt.

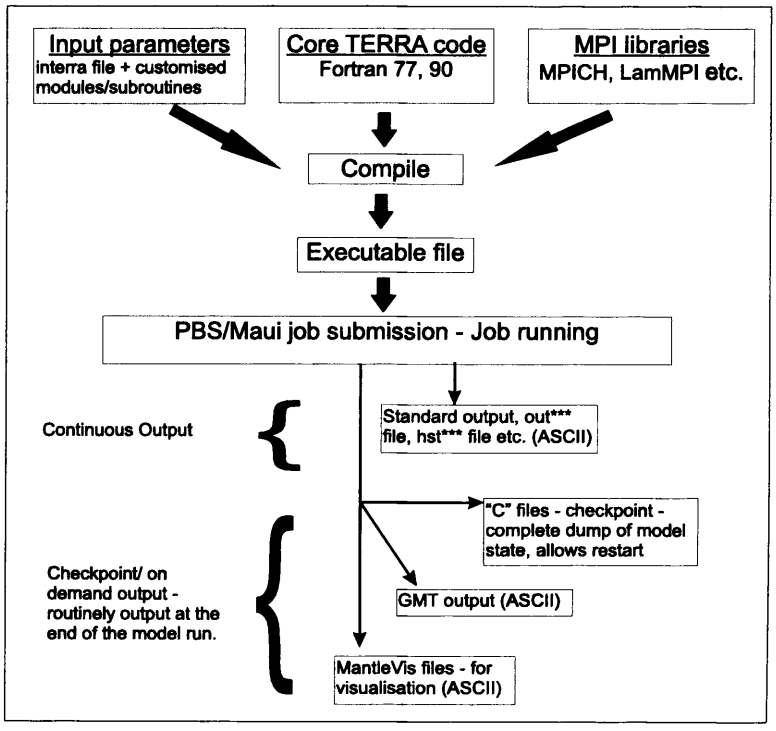


Figure 5.1: TERRA Workflow

5.4.2 Data Volumes

Terra has two fundamentally different output types. First, there are continuous output files which represent the bare minimum of useful summary output. These files are usually less than 10 MB in size since they are only summaries of actual data. The second class of files have values from the actual computational grid. These files are appreciably bigger. They include checkpoint files (to allow restart of the calculation) and output files for visualisation packages such as GMT (Wessel & Smith 1998) and MantleVis (Heath 2002). All these file types have the tendency to be very large at high model resolution (see Table 5.2 for examples). Many of these files are output multiple

Total file size in Megabytes			
mt level	*Checkpoint Files (binary)*	*MantleVis*	*Whole Grid GMT*
mt64	57.6	22	168
mt128	416	179	1248
mt256	3300	1430	9270
mt512	26000	11400	69067

Table 5.2: Data volume for various types of TERRA output

times during a simulation, particularly if an animation of the models' evolution is to be produced. Each visualisation or checkpoint file output by TERRA relates to a specific point in time during the model run. Provided that data was output on several occasions during a model run, the loss of one output does not invalidate the rest of the data. The storage system is therefore intended to be more resistant to the loss of small portions of several files than to the loss of an individual file. Additionally, the resolution of a TERRA run (see Table 5.1) affects the number of timesteps required. For each increase in resolution, the number of timesteps must be doubled in order

to simulate an equivalent time period. Table 5.3 gives an example of a typical run schedule for 2-3 years of work for one researcher and represents approximately 200 days processing on a 550 Gflops cluster. As a minimum, data integrity must be maintained for a similar time span. While model runs can be re-created (re-running with original code and parameters), the cost of doing this is can be exceptionally high. Re-running code should be avoided as much as possible. Additionally, an archive of previous model runs, accessible to a wider research group or community, would allow researchers to avoid unnecessary duplication and again reduce costs.

Total data in Gigabytes		
mt	*No. of runs*	*Totals*
64	60	129
128	40	804
256	10	3240
512	2	6562
		10,734

Table 5.3: Output requiring storage from a typical research project, based on 5,000 timesteps at mt64, 10,000 at mt128, 20,000 at mt256 and 40,000 at mt512. Outputting 2 checkpoints and a set of MantleVis files for every 500 time steps.

5.5 P2PS Based Distributed File System

The design of the system was driven by the requirements of storing TERRA output. However, the reliability and performance requirements are common to many research areas that utilise modelling. The fundamental elements of the initial implementation of the storage system were; standard P2PS peers, a file transfer mechanism and custom advertisements for advertising files held by a peer and for issuing instructions to the network. Each peer in the system maintains a list of the files it holds and

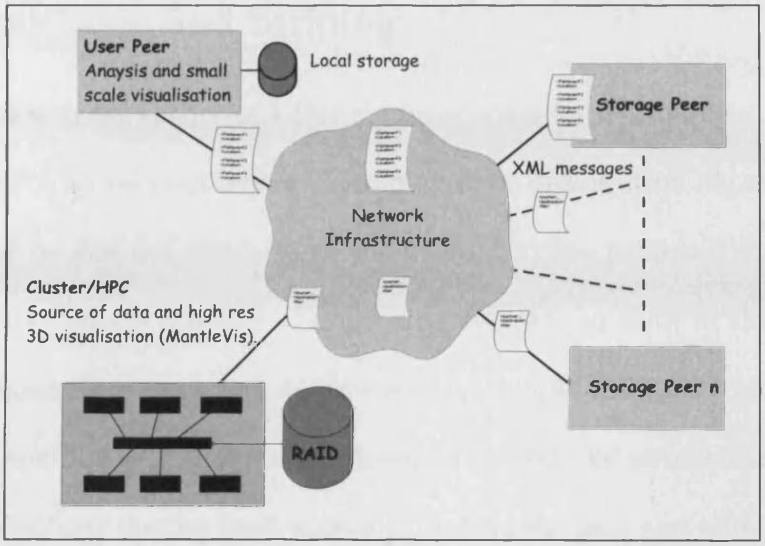


Figure 5.2: Key components the p2p storage system. The network is intended to represent an institutional LAN, however it could equally represent a WAN or the public internet.

advertises them to the network. A user peer connects to the network and constructs a query for available data. A list of available data are then shown to the user. To transfer specific data, a transfer advertisement is constructed containing a file identifier, the source and the destination where it is to be sent. When this advert is published the source peer initiates data transfer. Once the transfer advert has been published, the user peer is no longer required for a file transfer to take place, therefore the user peer could be described as having a 3rd person perspective on the file transfer. This does not prevent the user peer downloading data to itself if required. Figure 5.2 summarises the main hardware components involved.

5.5.1 Duplication and Striping

Due to the low trust nature of a P2P network some form of duplication is essential if data integrity is to be maintained. Complex data distribution algorithms, such as those described by Rabin (1989), were not utilised in the proposed implementation, however they may offer a method of improving storage capacity in the future. Here, simple duplication was employed to provide fault tolerance. To improve transfer performance a method of striping was chosen to exploit the parallelism inherent to a P2P system. Previous studies have shown that data striping can allow transfer rates over a network equal to or exceeding that from the local hard disk (Vazhkudai et al. 2005). Duplication and striping are best performed by the same algorithm, allowing 'chunks' of files and duplicates of chunks to be logically indexed. The basic function of duplication and chunking is:

$$D = \sum(d0_i...dn_i) \tag{5.2}$$

Where D is the original file, d is a chunk of the original file, 0 is the first chunk, n is the last chunk and i is the duplicate number. The manner in which data may be stored and retrieved on demand is demonstrated in Figure 5.3 where chunks and duplicates are numbered sequentially from 0. Mapping of chunks to peers is carried out in a different order for each file. This ensures that, unless the number of duplicates exceeds the number of peers no one peer holds duplicates of the same chunk. When the last peer is reached the mapping loops back to the first peer.

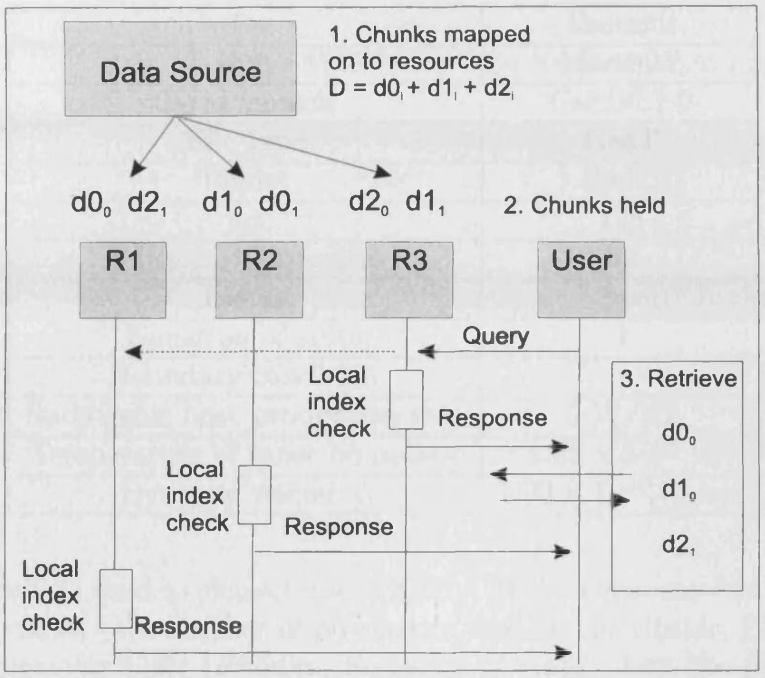


Figure 5.3: Distribution and return of chunks.

5.5.2 Indexing and Metadata

In any information storage system the user must be able to query the archive. For TERRA this would include the user searching for relevant data prior to initiating a new model run, thus saving computing time and money. General metadata for a TERRA model run is a combination of input and summary output parameters. There are some 40-50 values that are of direct relevance to a model run, too many to integrate into the simplified indexing system of the test bed. Instead, a subset of values most relevant to archiving and retrieval are outlined in Table 5.4. The metadata consist of values defining both the physical constraints of a model run and key information relating the computational method (TERRA). This information gives the user an indication of the type and resolution of model and provides the technical details

Tag	*Example*
User	MartinW
Terra version	Cardiff.1.0
File type	GMT
Planet	Earth
mt	128
Number of processes	32
Plates	no
Equation of state	1
Boundary condition	1
Radiogenic heat production rate	0 W/Kg
Temperature of inner boundary	2.85×10^{03} K
Dynamic viscosity	1.00×10^{23} Pa-sec

Table 5.4: Metadata used to characterise TERRA model runs. mt - model resolution, Number of processes - the number of processors used on the cluster, Plates - whether the model implements plate tectonics, Equation of state - how the mantle material behaves in response to changes in pressure and temperature, Boundary condition - how the model handles the upper an lower boundaries, Radiogenic heat production rate - the amount of heat generated within the mantle, Temperature of inner boundary - heat from the Earths' core, Dynamic viscosity - the reference viscosity of the model.

required to restart the run from checkpoint files. Metadata are attached as a header on each file chunk. In addition to model metadata, the header records chunk number, duplicate number and other internal system data. This will allow each peer to derive its own local index from the files it holds. In the simplest scenario, the system could be searched by publishing a query to the whole peer network, each peer searching its own local index and returning any matches. However, this 'query flooding' approach is not particularly scalable and makes poor use of available bandwidth (Ripeanu 2001). A data-oriented overlay network, based on the metadata within the system would streamline querying. There are of course many parameters (over 50) that could be chosen. However, a small subset of parameters can capture the main aspects

of interest to the modeller. A hierarchy of parameters may be constructed and the model outputs grouped accordingly, with the most discriminating parameter first (e.g. resolution). Provided that the hierarchy is carefully constructed, the possibility of a particular model theoretically occupying two groups simultaneously is unlikely. For example, the minimum reference viscosity which may be used in a given run depends on the resolution. A run at mt128 and a run at mt256 are extremely unlikely to share a viscosity value. The only situation where this may happen is where a high resolution case is deliberately run using a viscosity value, which would normally be run at lower resolution. The aim being to test the numerical convergence of the solution. These runs are relatively rare. Figure 5.4 demonstrates how a query might propagate through a series of groups, each group filtering the data such that fewer peers need to be queried. Such a structure requires that lower groups in the hierarchy depend in some way on the group above.

5.6 Evaluation

A test bed of the distributed storage system was developed to test the performance of the operating environment and gain an insight into potential future behaviour. This is important as the system must strike a balance between providing reasonable data throughput and the need to minimise the impact on the workstations donating storage. The duplication method was modelled to establish its resilience.

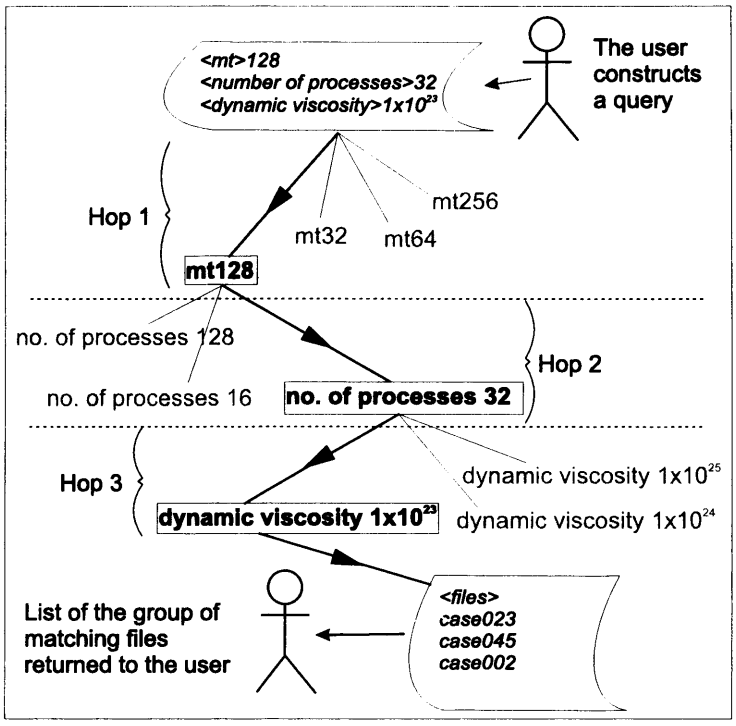


Figure 5.4: Example of querying, groups in bold are those relevant to the query, other possible but unused groups are in grey. Both number of processes and viscosity are strongly dependent on resolution (mt).

5.6.1 Performance Test Bed

A test bed of 5 PCs was set up, consisting of 1 Pentium M notebook and 4 Pentium 4 desktop workstations. All machines ran Windows XP Professional with JRE 1.5.0. Six sample file sizes were produced using TERRA output; 1 MB, 3.7 MB, 7.4 MB, 11.1 MB, 14.9 MB and 18.6 MB. A set of these files was allocated to each storage peer. Communication was by TCP pipes. Under evaluation were; the ability of peers to discover each other, File queries and advertisements, transfer advertisements, latency of the P2P network, rates of data transfer and the ability of the transfer mechanism to maintain data integrity. To derive a baseline of machine reliability an 8 day survey of 27 randomly chosen machines was undertaken; ≈85% were found to respond to a

ping request at any given time.

During test bed evaluation, 42 files, totalling 367 MB were transferred between peers operating in 2, 3 and 5 peer configurations. All workstations on which peers were installed were located within the same subnet, rendezvous peers were not implemented in the test bed. A single peer issued commands to the system. Qualitatively, all peers detected each other and were able to advertise their files. The controlling peer was able to view advertised files and issue transfer commands, which were carried out by the target peers. One peer failed during file transfer in the 5 peer configuration. Out of 42 files transferred 97.6 % arrived intact. Figure 5.5 demonstrates effective transfer rates achieved during file transfer between peers in the test bed. File transfer rates for larger files were marginally lower than those for smaller files, though this may be an artefact of small sample size. Data transfer was a processor intensive task for the duration of the transfer, however, idle peers use very few resources on their host machines. System memory usage fell within the ≈25 MB consumed by Java.

The latency of the P2PS network was defined as the time taken for a 'ping' to be answered using P2PS input pipes, output pipes and message listener. Although this operation will certainly take longer than a simple ping of a machine it is a more objective measure of the true latency of the P2P system as it represents the total time taken for the peers to react to a command. Figure 5.6 compares simple ping and P2PS ping. The Java VM and the peers themselves contribute the majority of the latency to the system, with typical P2PS ping times of 500-700 ms, compared to pure ping times of 5 ms. There was no difference in transfer rates or latency between

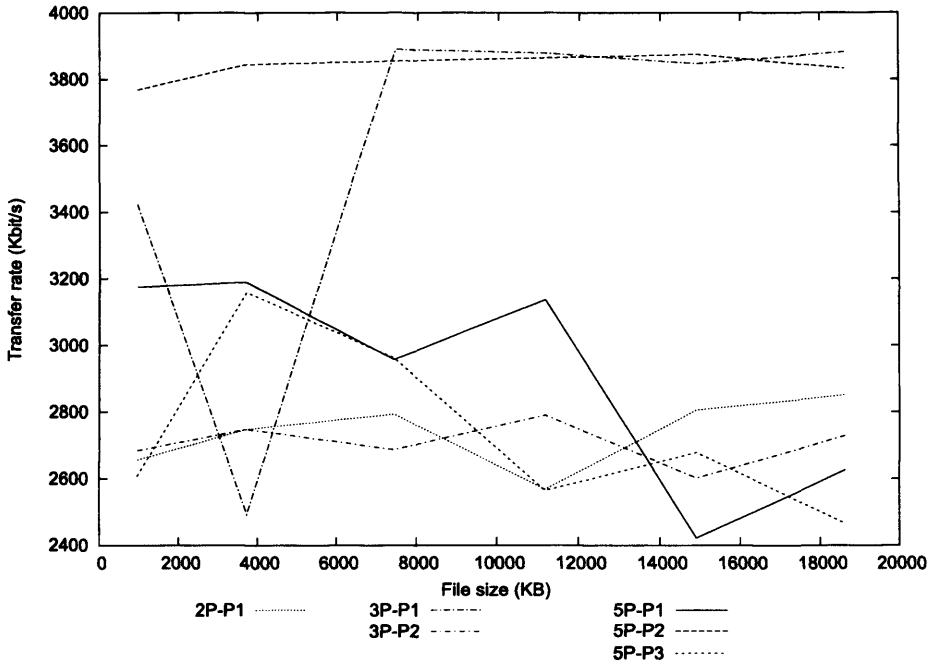


Figure 5.5: Transfer rates between peers in the test bed The speed relates to the data sending rate of the given peer in the given configuration. Key: (no. of peers in configuration)P-P(peer number).

2, 3 or 5 peer configurations.

5.6.2 Levels of Duplication

In the light of the observation that only ≈85% of peers may be available at any particular instant, the level of duplication must ensure that data remains available in such situations. There are two extremes of duplication. Where there is only one copy of each chunk, the loss of one peer results in the loss of access to the whole file. On the other hand, where the number of duplicates equals the number of peers each peer holds a complete copy of the file and only 1 peer needs to survive to retrieve the file. Obviously, the former case is almost certain to result in data loss, while the latter would allow very little data to be stored. The duplication system is most vulnerable

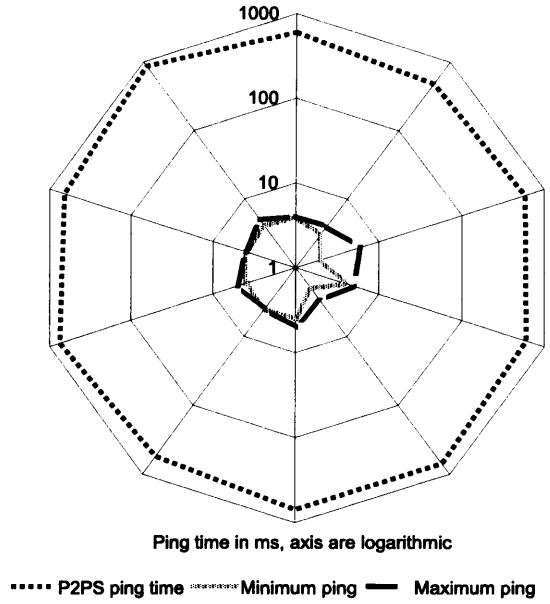


Figure 5.6: Comparison of typical 'pure' ping between workstations on the network and the actual latency of a P2PS peer responding to a request.

when the number of chunks is greater than the number of peers, concentrating the chunks. Table 5.5 demonstrates the reliability of different levels of duplication. The

Duplicates	*MPL*	*MSPL*	*Data Vol. (MB)*
1	1	1	500
2	50	2	1000
3	66	3	1500
4	75	4	2000

Table 5.5: Example employing 100 theoretical peers and a 500 MB file split into 100 chunks. MPL - absolute maximum number of peers that could be lost without chunk loss. MSPL - maximum number of sequential peers that could be lost before chunk loss occurs.

main risk in this duplication structure is the loss of consecutive peers. Excluding large-scale failures such as power outages, the probability of individual machine failure within different rooms across campus can be considered unrelated. Therefore, the failure of one peer is a discrete event and does not make it more likely that the

consecutive peer will also fail. For multiple files the risk is mitigated by the choice of a different sequence of peers each time a file is written so that different files do not share the same sequence of peers. This suits the nature of the data, as the loss of one file does not invalidate the rest of the files.

The probability that sufficient peers will be lost to render a given file irretrievable is:

$$If : M < k, P = 0 \tag{5.3}$$

$$If : M \geq \frac{(Nk - N + k)}{k}, P = 1 \tag{5.4}$$

$$Otherwise : P = \frac{((N-k)!M!)}{(k(M-k)!(N-1)!)} \tag{5.5}$$

where: P is the probability of file loss, N is the total number of chunks, k is the number of duplicates and M is the number of chunks lost.

The probabilistic problem is a variation on the 'socks problem', where one must determine how many randomly selected socks one needs to take out of a drawer before getting a matching pair. In this case there are many different colours of sock and a variable number of feet!

5.6.3 Discussion

The immediately noticeable feature of Figure 5.5 is the raggedness of the lines and the stronger performance of peer 2 in the 5 peer configuration and peer 1 in the 2 peer configuration. Part of the variability is almost certainly due to small file sizes used in this test bed, which allows small anomalies to dominate the data. The machine responsible for the higher transfer rates was a 1.83 GHz Pentium M notebook with 1 GB of RAM, the other machines were of lower performance with 512 MB of RAM. The differences could be due to: the amount of free system memory availably, the cleanliness of the windows install or issues with TCP over the network. The latter cause being the most likely candidate for introducing inconsistency to the results. Data transfer rate therefore appears strongly effected by the performance of the machine sending the data and the reliability of the network. This is significant as it indicates that faster data transfer may be possible when IT labs are upgraded, particularly with more RAM. The relatively high processor usage during data transfer strongly favours file striping across a relatively large number of peers to avoid undue performance impact on the workstation and inconvenience to general users. In this test, rendezvous peers were not implemented. The final system implementation will require a number of rendezvous peers through which new peers can communicate across subnets. Rendezvous peers will need be located on more reliable machines and new peers will be bootstrapped with the address of a rendezvous peer when they join the network.

The test bed confirmed the effectiveness of using customised advertisements to is-

sue commands to the system. This method of controlling the system is very flexible, as P2PS peers respond only to advertisement types they are interested in. Interest being registered by a peer issuing a query for a particular advertisement type. Thus, a peer can have a series of modes, each responding to and communicating via a different set of advertisements. The use of advertisements also allows the user peers in the system to issue commands and then leave the network, confident that the commands will be carried out. Further advertisement types that would be useful are, delete commands, disk space information, and general system health monitoring, which could be collected by the peers and transmitted in response to a query. Theoretical analysis of the duplication method suggests that 3 duplicates of every chunk would be sufficient to ensure that a file could be retrieved when up to 66% of peers on which the chunks are stored are unavailable. Even when only 2 duplicates are made of each chunk, provided the peers holding both chunks no not fail simultaneously, 50% of peers could fail before the file is irretrievable. 4 duplicates of each chunk is not a significant improvement over 3 duplicates, and uses significantly more space.

5.7 Conclusion

In summary, a P2PS based storage system is very feasible in a university setting and imposes few additional costs. The inherent structure of TERRA output is used to group model runs using key metadata to improve P2P search efficiency. The XML advertisements central to P2PS allow a high degree of flexibility and are readily adapted for use in a distributed storage system. Data transfer rates of 3-4 Mbit/s over a 10 Mbit/s connection favour file fragment sizes of 5 MB, as the 15 seconds

taken to transfer this file size is unlikely to cause a noticeable performance impact for the users of the machines. The method of duplication strikes a good balance of resilience and storage usage at a duplication level of 3. Multi-file integrity is further ensured by changing the order of peers for each set of chunks.

Chapter 6

MantleStor: Distributed Indexing and Storage of Earth Mantle Modelling Data

6.1 Abstract

The storage of scientific modelling data is a multi-faceted problem; there are issues of sheer volume, data indexing and cost. TERRA is a scalable 3-dimensional spherical geometry code, capable of modelling Earth's mantle with realistic parameters such as vigour of convection. While many details of the mantle are still being incorporated into the model, the computing issues surrounding TERRA have shifted from a question of how to produce a realistic model to how to visualise and analyse the models in a meaningful manner. The main visualisation tools used with Terra are; MantleVis (a parallel visualisation tool), Generic Mapping Tools (Wessel & Smith 1991) and

Paraview (paraview.org). With these types of visualisation comes data and when over 100 million data points are under consideration, this data volume becomes very large. File systems attached to the supercomputers where the data is produced are able to take up some of this data, however, these resources often need to be shared and can be expensive, especially when maintenance costs are considered. To explore one of the alternatives, this work presents MantleStor, a distributed indexed storage system, based on Peer-to-Peer technology and intended to operate over a significant number of standard institutional workstations. The results of a trial implementation over 8 physical machines are presented and analysed, with a discussion of the performance of the indexing system, system resilience and data transfer. MantleStor is able to handle one index search per second in ideal circumstances and one every 10 seconds in a more challenging environment. Data integrity can be maintained with over 30% loss of storage machines. MantleStor's parallel file download is able to make best use of available network bandwidth.

6.2 Introduction

Modelling Earth's mantle in realistic geometry and at high resolution is an endeavour, which produces large data volumes (100's of gigabytes). A system for storing such data was explored as a feasibility study (Wolstencroft et al. 2006, and Chapter 5), a larger scale implementation of this concept of a peer to peer (P2P) based distributed storage system is presented here. Features include file storage, search, retrieval and deletion. The search function is based on a flexible application-specific metadata indexing scheme.

A distributed storage system is of interest as large volume data storage is expensive but typical institutions have many workstations with significantly underutilised hard disks. Accessing this unused pool of storage was the first motivation for this study. A further motivation lies in development of an indexing system for such model data. Primarily conceived as a means of searching the stored data, the indexing method is not exclusive to a particular data type and can function with any data for which relevant metadata are available. The target user of this system is a small research group with access to computational resources in the small to medium-sized Beowulf cluster range, but with limited storage space for large datasets. Long term (5 years and over) data storage is not a requirement of the system as subsequent generations of clusters make models run on previous clusters almost trivial to re-run.

The novelty of MantleStor arises from the integration of P2P concepts, data storage and metadata indexing of Earth mantle modelling datasets.

6.3 Previous Work

MantleStor combines a number of concepts. Inspiration was drawn from a P2P systems such as Gnutella, while other aspects of the system are motivated by distributed search and data integrity. The system is intended for a small number of users and a large number of autonomous peers, this style of operation has similarities with Grid (or more recently: 'Cloud') computing environments.

The key component of MantleStor is the indexing system. It is inspired by search engines in general but is tailored for users who are asking for something other than a keyword match. An archive search needs to find a particular parameter and return matches within a given range. Thus search is a 2 part operation, first finding matching parameters then finding the values of those parameters that match the range defined. This is different from the type of search that Google does, where results are based on a measure of relevance (Page rank) to the keyword or search term. In this respect, the MantleStor index could be described as a simple distributed database.

The storage aspect of the system is similar in its goals to a number of other systems such as Farsite (Adya et al. 2002), SNS (Ye et al. 2003) and pStore (Batten et al. 2001). The shared goal is to utilise otherwise wasted storage space in a similar manner to how Condor (Thain et al. 2005) makes use of otherwise unused CPU cycles. The intent was not to produce something of the scale and complexity of systems such as SRB (Rajasekar et al. 2002, Rajasekar & Wan 2002) and OceanStore (Rhea et al. 2001) but to implement a lightweight and functional solution to the problem of finding storage for large datasets.

A P2P approach was used as the key assumptions of inherent unreliability, 'edge' environment and low machine resource overhead match with the institutional computer room environment in which MantleStor will be expected to operate. In addition, the availability of P2P Application Programming Interfaces (APIs) such as P2PS (Wang 2005) allowed rapid implementation and testing of MantleStor.

6.4 The TERRA Mantle Model

TERRA is an established mantle modelling code and is extensively described in the literature (Baumgardner 1985, Bunge & Baumgardner 1995, Yang & Baumgardner 2000). Very briefly, TERRA is a finite element, 3-dimensional spherical geometry whole-mantle thermal convection/circulation model. Resolutions up to 15km are possible, allowing very Earth-like simulations. Chapters 2, 3 and 4 provide more detailed descriptions of the different aspects of TERRA and the underlying science. The most relevant aspect of TERRA for MantleStor is the large amount (in terms of information not file size) of summary data it produces along with the large volume visualisation and checkpoint files (which require storing). The summary data consist of at least 50 setup and output parameters, which display varying degrees of uniqueness. In general, model runs that are similar in their behaviour share similar parameters. Only model cases that share exactly the same setup parameters, initial conditions and have been run for the same number of time steps will have identical output parameters. The parameters all relate to the physical behaviour of the system and their use within the indexing scheme is discussed in section 6.5 (below). The parameters can include input parameters for a model and output parameters that result from that simulation.

6.5 Indexing Strategy

The novel aspect of this storage system is its use of TERRA metadata to tag the data. When searching the store, a list of files matching a required parameter is returned, rather than showing a list of all files available. To highlight how this is

valuable, a typical model parameter is described.

The Rayleigh number (Ra) is a non-dimensional parameter, which represents a measure of convective vigour in a system. The larger the Ra, the more vigorous the convection. It can be used as a measure of how Earth-like a particular model case is, at present the Earth is thought to have a Ra of $\approx 10^8$ (Bunge et al. 1997, Weeraratne & Manga 1998). As Ra increases, models demonstrate more complex behaviour and a broader range of length scales (Figure 6.1). Thus, when searching an archive of model cases with the intention of answering a particular question, some Ra ranges will be of more relevance than others. Indeed, different sized planetary bodies will have different Ra e.g. Mars has a lower Ra than Earth. Many metadata will be

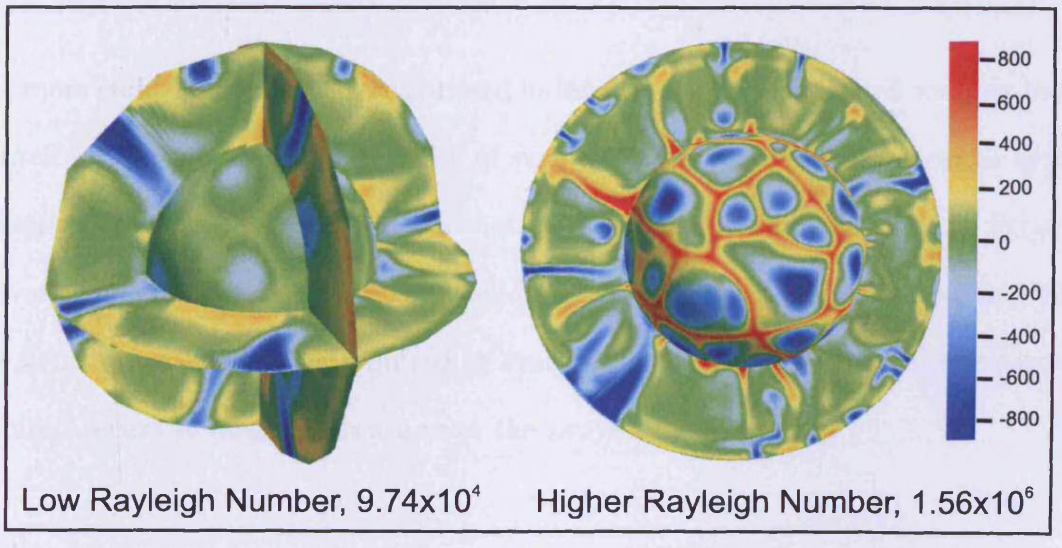


Figure 6.1: An example of low and high Ra model cases, scale represents temperature anomaly in degrees Kelvin from the average.

common to a significant number of model cases that may be run; for example, the radius of the mantle, which would only change if one was simulating another planet.

There are more specific metadata, which are defined by the purpose of the modelling. A typical example would be cases run to assess the effect of increasing Ra on the amount of heat given out at the surface of the modelled mantle (as in Chapter 2). Such models would have identical metadata apart from the Ra.

If only one set of cases were stored, the archive would be simply a store. However, as more cases from different projects are added, searching the store will allow connections between different runs to be more easily made. For example, a straightforward case run as part of a given set may be compared with a more complex run or used as an initial condition for further runs. This allows more effective comparison and discovery of existing data than was the case previously.

The more complex overlay network based index proposed in Chapter 5 was not implemented in MantleStor as the benefits of such an overlay became less clear as system design progressed. The rationale was that P2PS is already an overlay on the Ethernet network and that a further overlay would simply complicate implementation. Additionally, the scale of testing required to evaluate such an overlay based index system required access to more hardware than the project could muster.

6.6 System Structure and Implementation

MantleStor is implemented in Java and is based on P2PS (Wang 2005), a JXTA inspired P2P API (also implemented in Java). P2PS uses Extensible Markup Language (XML) Advertisements for discovering and communicating between peers. The

Advertisement system of P2PS has been extended with the addition of the following Advertisement types: PeerStatAdvertisement (collects statistics such as available disk space), SearchAdvertisement, FileAdvertisement (returns the details of a file in response to a search), TransferAdvertisement (a command to send a given file to a given destination) and FileDelAdvertisement (a deletion command for a given file). Other than these additional advertisement types P2PS is used 'as is' to provide P2P communication between MantleStor peers. This section describes how MantleStor operates on a technical level, some aspects are slightly simplified for clarity. Full source code can be found in digital Appendix A as a PDF and runs to ≈3200 lines of code over 76 pages.

6.6.1 Communication

P2PS advertisements provide the communication structure of MantleStor. Advertisements consist of XML messages that are passed between peers, which are active once they have been published. Advertisements are of a given type and may hold a variety of data, set at instantiation. A peer will only 'hear' a given type of advert if it is actively listening for it, this is achieved by publishing an Advertisement Query for that type of Advertisement. Thus, the search system is asynchronous, an indeterminate amount of time may elapse between Query and corresponding Advertisement publication. For example, when a peer boots, it publishes a FileQuery, this dictates that when the peer detects a FileAdvertisement, an event is triggered by the P2PS peer, allowing MantleStor to process the Advertisement and act on the contents (call

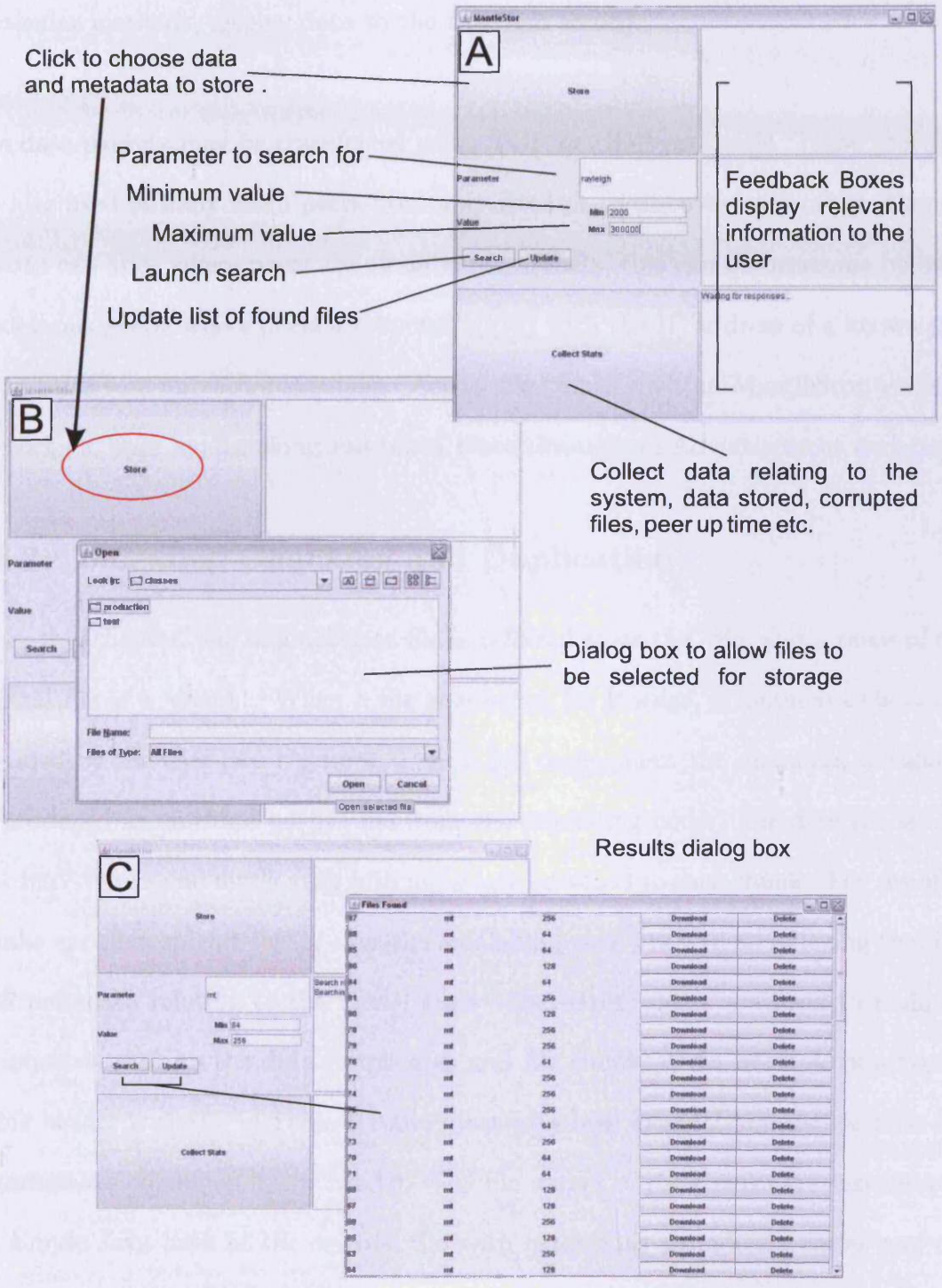


Figure 6.2: MantleStor GUI, (A) Overview of the GUI and its functions, (B) Storing a file, (C) the search response. Note: 'search' clears the returns list, generates a new advertisement and then waits before showing returns, 'update' simply shows all returns since the last time 'search' was clicked.

particular methods, display data to the user and so on).

The data packets may be transferred using TCP or UDP protocols. These protocols are also used initially when peers boot and attempt to discover each other. Complications can arise where peers are on different subnets, this can be overcome by using rendezvous peers, where peers are bootstrapped with the IP address of a known peer on an otherwise unreachable subnet. Actual file transfer within MantleStor uses simple sockets, once handshaking has taken place through an Advertisement exchange.

6.6.2 Indexing, Chunking and Duplication

In this chapter, the original data file is referred to as the 'file' and a piece of this original file is a 'chunk'. When a file is selected for storage, a metadata file is also provided by the user (see Figure 6.2). In a full deployment the metadata file should be produced as an extra output file from the modelling code. The data file is then split into chunks and duplicated with metadata attached to each chunk. The resulting chunks are then spread evenly over the available peers. As well as indexing the data with metadata relating to the model cases themselves, there is a need to maintain an internal track on the data (duplicates and file chunks). An example of a typical chunk header is shown in Table 6.1 and illustrates how internal MantleStor data and metadata are combined in the header. The file names of the chunks are derived from the simple Java hash of the original file with extensions to indicate chunk and duplicate number. The format is: hash.chunkNo.duplicateNo (e.g. 893782476.1.2). A more sophisticated hashing algorithm could easily be substituted, if required.

Entry	*Description*
893782476	The hash code - 1st part of chunk filename
1	Chunk number
3	Total number of chunks the file was broken into
2	Duplicate number
3	Total number of duplicates of each chunk
example.txt	Original file name
15	Total number of lines in the header
mt	First parameter
64	Value
ird	Second parameter
2	Value
.	
.	
.	
line 15	End of the header
18.467 30.379 300.000	Data begins here, may be ASCII or binary
17.858 29.499 300.000	
17.260 28.617 300.000	
16.672 27.732 300.000	

Table 6.1: An example file chunk header, the information is written to every file chunk as it is created. Note: the top 7 lines of the header are always present, the metadata is expandable depending on the length of the metadata file specified with the original data.

The actual chunking and duplication is carried out as follows:

Input: file to be split (data), a metadata (meta) file in plain text and the number of duplicates to make.

```
Check both input files exist and can be read

if (data.size <15 MB)
   chunk size = data.size/3
```

```
else (chunk size = 5 MB)

data.size/chunk size = total number of chunks

hash = positive integer of data

Duplicate loop (iterates for number of duplicates) {

    Chunking loop (iterates for number of chunks) {

        Individual chunk {
           Write to header:
           file hash
           chunk number
           total number of chunks
           duplicate number
           total number of duplicates
           original file name
           total length of header

           Write contents of the meta file to header
           Write (chunk size) bytes from data to chunk

           Close the chunk
        }
     }
}

If (file chunking was successful)
delete the original files.
```

Result: a set of chunks (with a set number of duplicates each) of the original file, each tagged with metadata relating to the chunking system and the data itself.

6.6.3 Storing: Distributing Files

Once the file chunks have been produced, they must be distributed as evenly as possible over the available peers. This is done by making a list of available peers and distributing a file chunk to each one in turn. When the last available peer has been sent a chunk the system loops and begins at the first peer again. When dealing with files of the same size, there will be the same number of chunks of each file. If many such same-sized files are being distributed consecutively, there is the risk that all will be sent to the same set of peers. The consequence of this would be that the loss of sufficient chunks to make one file unavailable would result in all the files that were distributed together being unavailable. To avoid this, the first peer is taken from the list using a randomly generated array index. This should minimise the chance that the loss of one file will result in the loss of all files stored.

Distribution is carried out as follows:

```
Publish ServiceAdvertisement requesting peers send IP addresses
Wait 20s
Discovered IP addresses = IP Array
Starting at random index of IP Array
  while (chunks remain){
    File sending loop{
       connect socket to remote peer
```

```
        send chunk filename

        wait for remote response (file name received)

        send chunk as bytes
        }

   if(chunks remain)
      get next index of IP Array
   if (last index has been reached)
      go back to start of IP Array
}
```

The recipient side of the file transfer is implemented as a file receive thread, which runs continuously listening for incoming file send requests. When a direct connection is detected the thread instantiates a new thread to download the individual file. Thus many files can be downloaded simultaneously. The 'filenames' in the code below are the chunk names, which are unique.

File Recieve:

```
Receive thread detects incoming connection on port 80

   new receive thread instantiated with new socket{

    read incoming filename from socket

    create new empty file with name filename in stor directory

    acknowledge sender

    read bytes from socket and write to new file

    close streams/socket
```

```
}
```

6.6.4 Local Index

The local index is a 2D array containing all metadata of all file chunks held by a given peer. The list is populated, maintained and searched by the LocalIndex class. The metadata is read from the chunk headers (described in Section 6.6.2) and stored in the local index as follows:

Input: the path of the store directory.

```
Instantiate 2D ArrayList = theLocalIndex

Load all files in store into File array.

for each file{
   put the filename into theLocalIndex column 0

   for each parameter in the file header{

      if(Column head contains parameter)
         put the value into the column.
      else
         make a new column with the parameter as head
         put parameter into new column
   }
}
```

The important technical aspect of the local index is that if a particular file does not have a particular parameter associated with it a blank marker is entered into the local index array. This allows many completely different sets of metadata to be stored in

the same array. A small example of what the index might look like is shown in Table 6.2.

Chunk name	*mt*	*bheat*	*ird*	*age*	...
6748399484.1.1	64	2500	2	100000	...
8755646473.3.2	128	2850	1	-	...
6748399484.6.1	64	2500	2	100000	...
2453268074.1.2	256	-	-	300000	...
...	...	...	...	...	...

Table 6.2: An example local index array.

The local index is searched as follows:

Input: Parameter of interest, minimum value and maximum value

```
Search(Param, Minvalue, Maxvalue){

  Refresh the local index

  Find Param in the first row of local index

  Iterate down Param column checking each
  entry against Minvalue and Maxvalue

  Add the filenames of any matching chunks
  to an ArrayList of matching files

Return the ArrayList of matching files.
}
```

6.6.5 Searching MantleStor as a Whole

MantleStor is searched from the GUI of the user client (Figure 6.2) by entering a parameter, a minimum value and a maximum value. For example, if the parameter is Rayleigh number (described in section 6.5) a minimum value might be 2000 and a maximum 300000. If both values are identical then only that value will be searched for. Search returns all files found that satisfy the search criteria.

Searching the distributed archive:

Input: param, min and max value.

```
Client:

Search button clicked

   SearchAdvertisement constructed using input
   SearchAdvertisement published.

   Client waits 20 seconds...*

-----------------------------------------
Peers:

detect SearchAdvertisement

search Local index

Make list of matching files

For each matching file make FileAdvertisement

publish FileAdvertisements
-----------------------------------------

Client:
```

```
Incoming FileAdvertisements are added to an ArrayList

After 20 seconds...*

List of found files displayed to user.
Options: Download or Delete.

Update button - refreshes found file list with contents
                of newly arrived FileAdvertisements.
```

6.6.6 Retrieve/Delete

Once the list of files matching a particular search is displayed (Figure 6.2), the user may download or delete files stored in the system. The mechanics of these tasks are shown below. The Assembler mentioned is a thread, which runs in the background checking whether enough chunks have been downloaded to reconstruct the initial file. When the Assembler finds this is the case it runs the Unchunker. Each file being downloaded has its own Assembler, which is only interested in chunks with a given root filename (hash code).

Download/retrieving files:

```
Client:

Download button clicked

Assembler for file instantiated

TransferAdvertisement constructed with file and destination

TransferAdvertisement published
```

```
-----------------------------------------
Peers:

TransferAdvertisement detected
TransferAdvertisement used to search for matching files.

List of matching files produced.

For each file{
   instantiate sender.
   connect socket to destination
   send filename
   send file as bytes
   close socket/streams
}
-----------------------------------------

Client:

Receive thread detects incoming connection on port

   new receive thread instantiated with new socket{

    read incoming file name from socket

    create new empty file

    acknowledge sender

    read bytes from socket and write to new file

    close streams/socket
    }

Incoming files are placed in /tmp.

Assembler checks every 30 seconds:
if(all chunks present)
   run UnChunker
```

The Unchunker is the same as the chunker but in reverse. The first chunk of a set is passed in as an argument to the UnChunker.

UnChunker:

Input: a chunk of a set located in tmp directory

```
Read the header of chunk

  get total number of chunks

  get original filename

if(all chunks present in /tmp)

   write metadata to a new text file

   Create a target file with the same name as the original data file.

   Assembly loop{
      Open each chunk sequentially.
      Read past the header.
      Read the data from the chunk and write to target file
      Close the chunk
   }

Close the target file

if(file assembled correctly)
delete chunks from \tmp
```

Result: an identical copy of the original file and metadata file.

Files may also be deleted from the whole system by the user.

Deleting a file:

```
Client:

Delete button clicked for file

FileDelAdvertisement published for file

-----------------------------------------
Peers:

Detect FileDelAdvertisement

Search local store

Files matching FileDelAdvertisement are deleted.
```

Note: both filename and chunk hash must match those specified in the FileDelAdvertisement before a file is deleted.

Complete source code (.java and PDF) and MantleVis application can be found in Appendix A.

6.7 Application and Testing

The primary goal of testing was to establish the effectiveness of the indexing system, with tests of request rate and precision carried out for increasing numbers of peers. Other aspects, such as data transfer rate and resistance to failure were also measured and assessed in a user context. For example, the data transfer rate measurements answer the question of, "how long will it take me to store my data", rather than purely abstract measurements. Particular attention was paid to the effects

of the operating environment, as Chapter 5 has demonstrated that the success of any deployment will strongly depend in it.

6.7.1 Request and Precision

Since a distributed archive such as MantleStor must be searchable within a reasonable time, the maximum request rate and the precision of the index system were assessed. Test beds of 1 to 8 peers were used, explicit rendezvous peers were not used and peer discovery used TCP multicast. Inter-peer communication was by P2PS pipes. Each test was carried out by first populating the system with 100 known files (of negligible size $\approx$100 KB) and associated metadata. The initial 100 files were split into 3 chunks and duplicated 3 times, resulting in 900 tagged chunks held within the system. The request test consisted of a series of 20 queries generated at a given interval, the number of returns were counted and compared to the number of returns expected from the known metadata. Results were recorded as a decimal value between 0 and 1, where 1 is perfect return from all 20 queries.

The precision test consisted of a series of 30 randomly generated queries with a given interval. The queries were submitted to the system and to a local store on the machine generating the search containing a complete copy of all chunks stored in the system. The local store return was considered perfect and was compared with the system return. Unlike the request test the accuracy of the response was also taken into account, any returns that did not match the query/local return were considered failed. As with the request test, the results were recorded as a value between 0 and

1, where 1 is a perfect return from all 30 queries. The query intervals used for both request and precision were: 20 s, 10 s, 5 s, 2 s, 1 s, 0.5 s, 0.25 s and 0.125 s.

For analysis, results for both the request and precision tests were scaled with the best return rate (usually from the 20 s interval) used to represent 100%. This allowed comparison between different network speeds and machine specifications, primary interest being in the scaling of the system with increasing query load. The results are presented in Figures 6.3, 6.4 and 6.5.

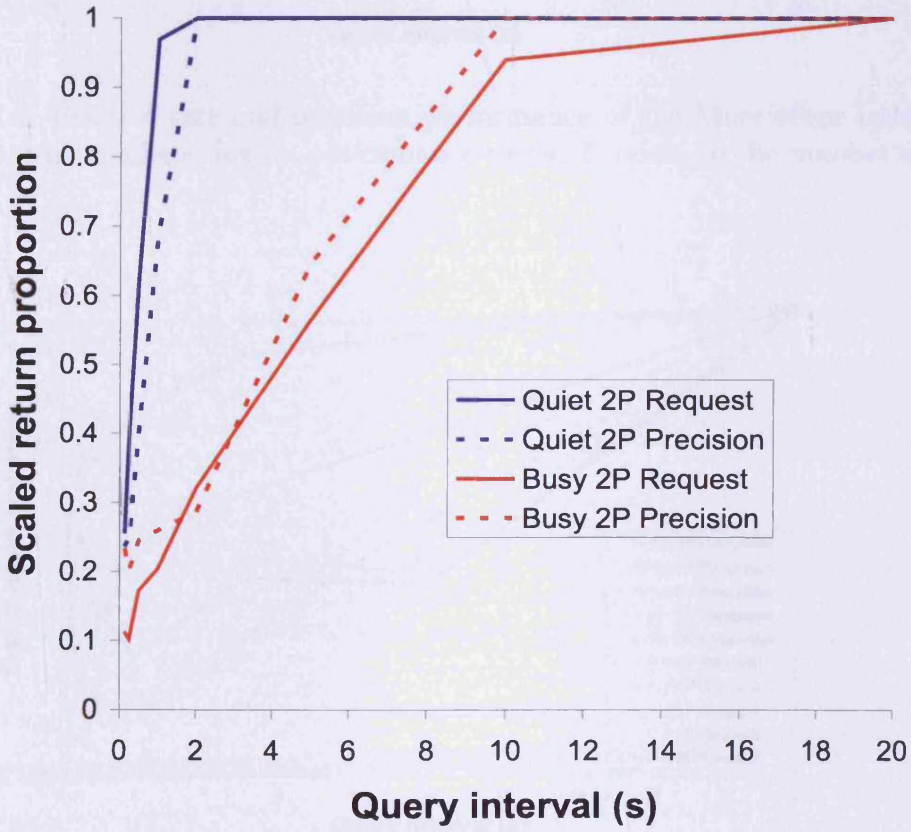


Figure 6.3: Example of the effect of network traffic on request rate and precision performance. P refers to the number of storage peers.

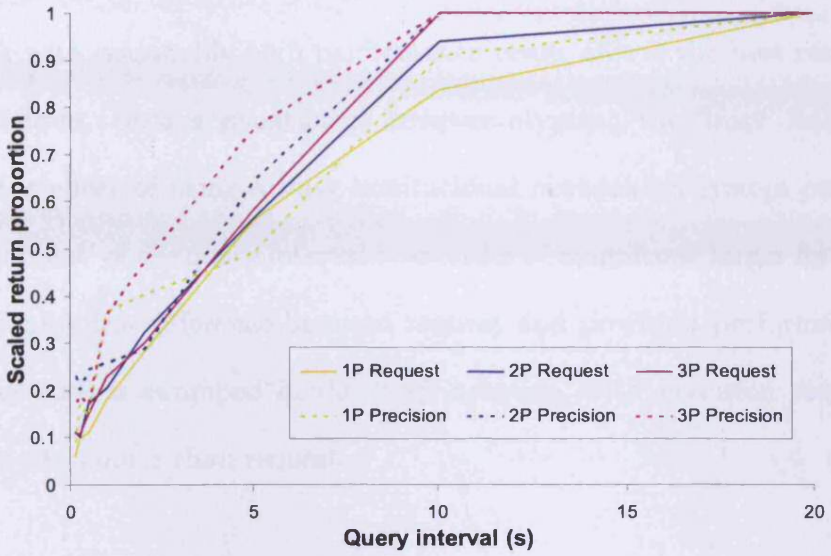


Figure 6.4: Request rate and precision performance of the MantleStor indexing system, low peer numbers, higher performance peers. P refers to the number of storage peers.

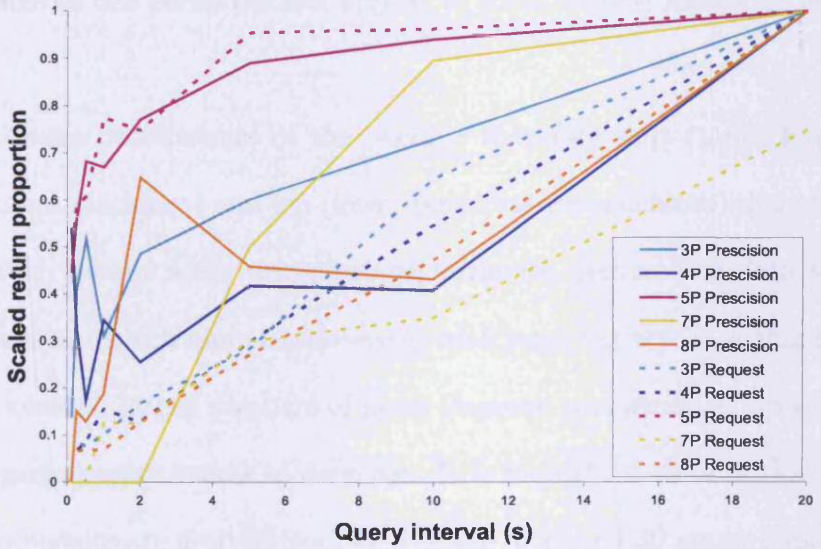


Figure 6.5: Request rate and precision performance of the MantleStor indexing system, higher peer numbers, low performance peers. P refers to the number of storage peers.

A number of observations emerge from the results. Figure 6.3 shows a fast, quiet network with reasonably high performance peers, this is the best result of the whole suite of tests. Such a situation is however atypical, the 'busy' lines on Figure 6.3 shows the effect of using a busy institutional network on system performance. The 'tipping point' of the query interval is an order of magnitude larger for a busy network. Additionally, the difference between request and precision performance seen with a quiet network is swamped in the busy network, with precision seeming to achieve better performance than request.

The most likely sources of failure within the system are: multicast discovery failure, where a peer fails to receive an Advertisement, peer overload, where a peer cannot update rapidly enough to respond to incoming queries and hardware failure. Most of the failures in this series of tests appear to have been of multicast or peer overload.

The hardware performance of the peers is therefore important. Figures 6.4 (higher performance machines) and 6.5 (lower performance machines) demonstrate how scaling is much poorer when using low performance peers. The data for precision and request are also much more inconsistent, with large variations within the plots (Figure 6.5). In general, larger numbers of peers improve performance, up to 5 peers, beyond 5 peers performance seems to decrease. It is important to note that the request and precision metrics are derived from an average of 20 and 30 measurements respectively, thus the inconsistency is unlikely to be due to a lack of smoothing. The relatively poor system performance is caused both by lower performance peers being unable to

update and query their local archive rapidly enough to keep up with incoming requests and by failures of multicast. This local index update process is somewhat inefficient by design as the assumption is that there can be no guarantee that a file held in the store and archived at runtime will be present at an arbitrary point thereafter. Thus the local index must be updated regularly. Low performance peers are Pentium 3 based (256-512 MB RAM), high performance peers are Pentium 4 (or better) based (512-1024 MB RAM). It is likely that machines with 256 MB of RAM are forced to use swap space, dramatically reducing their performance and introducing inconsistency into the results. Operating system and software was Windows XP and JRE 1.6.0-05. Therefore, although MantleStor will run on a P3 machine, a P4 or faster is necessary for good performance in a busy environment. However, simple subjective tests when a file was stored, searched for and then returned showed no great variation between low and higher performance peers in terms of user experience. This is due to the fixed waits built into the system, which allow slower peers to catch up. These waits were not used in the precision and request tests as the system was deliberately overstressed.

In a realistic environment, a minimum query interval of $\approx$10 s is advisable. This is not a major hindrance to the usability of the system considering the time it takes to download very large files (Section 6.7.2 below).

6.7.2 Data Transfer

As MantleStor is intended for the storage of large files, the scaling of the data distribution and retrieval of system was measured. Building on the work of Chapter

5, five sample files were produced with sizes of: 27 MB, 54 MB, 80 MB, 107 MB and 214 MB. These were then stored in a MantleStor system of between 3 and 8 peers, searched for and retrieved. Storage and retrieval were timed and the effective transfer rate calculated. Results are shown in Figures 6.6 and 6.7.

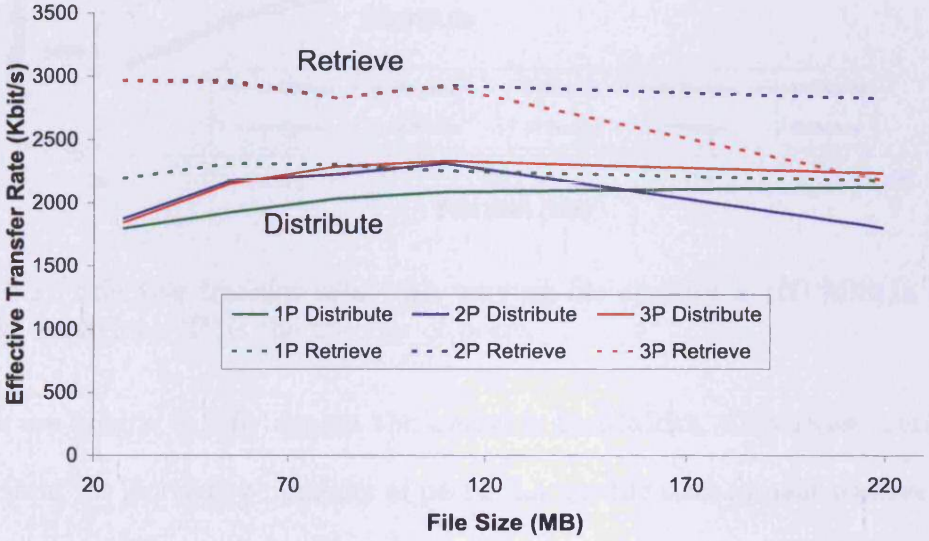


Figure 6.6: Effective transfer rate with varying file size for a 10 Mbit/s Ethernet network. The transfer rate includes time spent handshaking etc. thus it represents performance in a user focussed manner. 'P' is the number of peers.

Both Figures 6.6 and 6.7 show a marked difference in terms of transfer rate between distributing and retrieving, this is due to the nature of MantleStor. Distributing a file is a serial task, the client maps one file at a time onto one peer. By contrast, retrieving a file is a parallel task, many peers are able to send their chunks simultaneously, utilising bandwidth more effectively. In both graphs this observation is confirmed by the performance of the lowest number of peers (1P in Figure 6.6, 3P in Figure 6.7) the retrieval rate is lower in both cases, indicating that smaller numbers

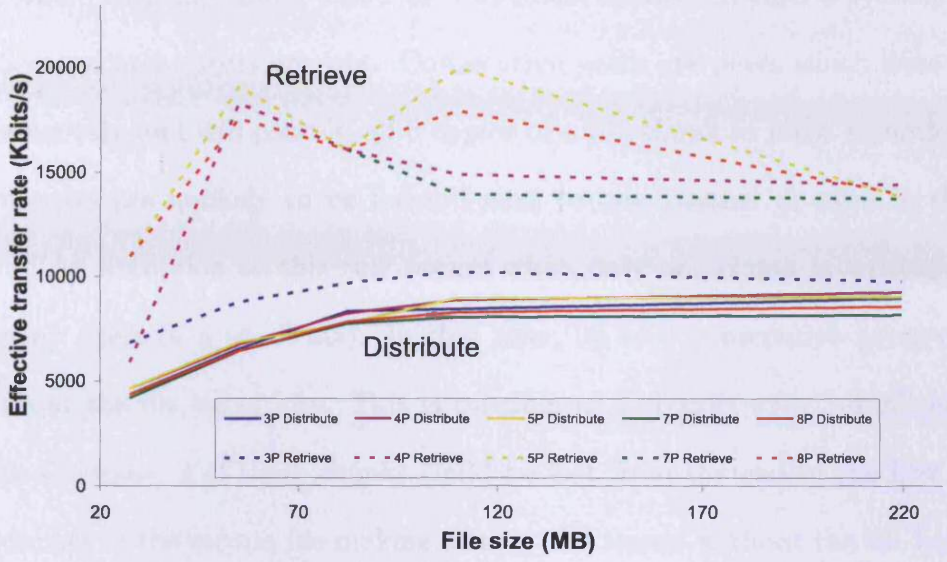


Figure 6.7: Effective transfer rate with varying file size for a 100 Mbit/s Ethernet network. As before, 'P' is the number of peers.

of peers are unable to fully exploit the available bandwidth. Otherwise, performance is consistent for increasing numbers of peers. Larger file sizes appear to have a higher transfer rate, this is the result of the increasing time spent actually sending data vs the system waits, which influence all transfers regardless of size. The jaggedness of the retrieve plots in Figure 6.7 are the result of the Assembler running every 30 s, sometimes downloads fit nicely within this 30 s cycle, sometimes not. The effect is moderated as file size increases (increased time to download).

6.7.3 System Resilience

For any distributed storage system, some level of resilience is essential if the integrity of the archive is to be maintained. MantleStor implements a duplication

system where each file chunk exists as 3 identical copies. In such a system a file is lost if 3 consecutive peers are lost. Consecutive peers are peers which were written to consecutively and will contain all 3 copies of a file chunk in most scenarios. Consecutive peers are unlikely to be located next to one another or even in the same building. An exception to this rule occurs when only one chunk is written to each peer (many peers or a small file). In that case, up to 4 consecutive peers could be lost without the file being lost. This is possible as 2 chunks with 3 duplicates each comprise 6 chunks, 2 of these chunks could be lost from the end of the first file and the beginning of the second file making a total of 4 losses, without the file being lost. The more usual situation, where each peer holds many chunks precludes this possibility. Thus the worst case scenario is that the loss of 3 peers results in the loss of the file.

As the critical number of consecutive losses remains equal to the number of duplicates (3), the percentage of peers that must be consecutively lost becomes smaller as the number of peers increases (Fig. 6.8). In a qualitative sense, the likelihood of consecutive loss reduces as the number of peers increases. A more likely scenario involves the loss of random peers from various positions in the original mapping. In the best case scenario where all losses occur in this manner, a file may be retrieved with many of the original peers missing. Fig. 6.8 illustrates both the best and worst case scenarios together with data from a test scaling to 8 peers. The real data plots in the envelope between the worst and best case scenarios with a preference for the best case. The shaded area indicates where larger numbers of peers would be expected to plot. The system appears to be resilient against losses of around 30%. In previous work we

showed that a network of this type may have up to 15% of machines unavailable at any given time (Chapter 5). Therefore, the duplication system is applicable to the operating environment.

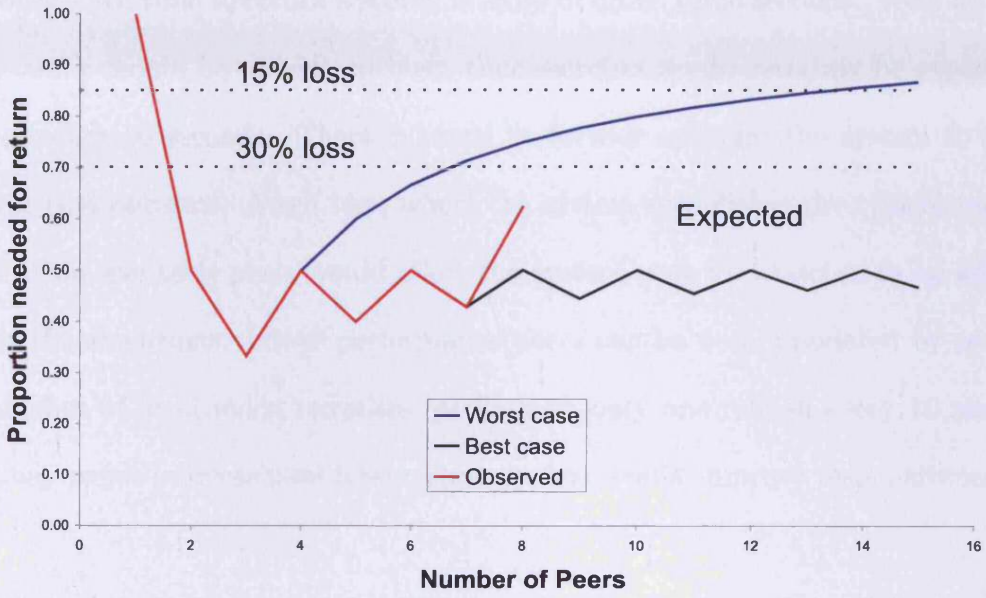


Figure 6.8: Results of the resilience test, worst and best case curves are synthetic data based on the manner in which chunks are mapped to peers.

6.8 Concluding Discussion

MantleStor, when deployed, would be expected to operate over ≈60 machines (peers). The analysis of the system performance presented above extends to 8 peers, therefore we use the trends observed in this relatively small number of peers to predict the behaviour that would be expected in a 60 peer configuration. The host machines would ideally be Pentium 4 or above with more than 1 GB of RAM, connected by a fast Ethernet network (100 Mbit/s or higher). In such an environment, the indexing

system will return accurate searches within a second or two (well within the wait built into the search method). However, in a more realistic environment where hardware performance may be less than ideal or the network has lower bandwidth (10 Mbit/s) or is busy, then time to return a search is more of order 10-20 seconds. With 60 peers, the network should be considered busy, thus searches would certainly be expected to take of order 10 seconds. There is scope to further optimise the system to run in difficult environments. A self test, where the system establishes the typical response times of the available peers would allow the system wait for searches to be adjusted to suit the conditions. Lower performance peers can be accommodated by reducing the number of local index refreshes, perhaps to only one refresh every 10 searches. This may result in occasional false responses, but would improve responsiveness.

Considering the resilience of MantleStor to be consistent with modelling and observations (Figure 6.8), ≈18 peers could be lost from a set of 60 before file integrity is at risk. The system could be made more robust by increasing the number of duplicates, should the infrastructure be particularly unreliable. The cost of this would be to reduce the overall amount of data that could be stored.

Transfer performance is obviously dependent on the network, assuming a 100 Mbit/s network, ≈9 Mbit/s distribution speed and ≈15 Mbit/s retrieval speed can be expected (Figure 6.7). A typical set of TERRA checkpoint files for a mt=256 run might constitute 3 GB of data, which would take 45 minutes to distribute and 27 minutes to retrieve. With a 10 Mbit/s network (Fig. 6.6) 3 GB of data would take 3 hours

20 minutes to distribute (2 Mbit/s) and 2 hours 23 minutes to retrieve (2.8 Mbit/s). Again, the performance of MantleStor is very much constrained by the hardware, though a 100 Mbit/s network is only 4-5 times faster than a 10 Mbit/s network, not 10 times as theoretical data rates might suggest.

The reason for such low performance could be that files are split prior to distribution, leading to a longer cumulative transfer time. The encoding/decoding of data being carried out by the Java machine as it sends the chunks may also have a varying negative performance impact depending on whether the data is ASCII or binary. Within the system itself, Advertisements must be converted into XML, another processing intensive task. Higher transfer rates might be achieved if, once discovered, messages were sent one-to-one between the peers, this could also simplify handshaking between peers before chunk transfer takes place, streamlining the process.

It is worth noting that using gzip to compress a set of binary checkpoint files totalling 3 GB takes 20 minutes on an Intel Harpertown based cluster node, unzipping takes 12 minutes and $\approx$7% compression is achieved. This result would of course be different for ASCII files and depends on the speed of the processor and file system, hence it can only really be considered an example. The equivalent 'speed' of this process is $\approx$23 Mbit/s compression and $\approx$40 Mbit/s decompression. Though this is clearly faster than distributing data with MantleStor, the data rates are of the same order of magnitude, and the space savings are very dependent on the type of file being compressed.

To conclude, MantleStor is a fairly simple implementation of a distributed storage application. MantleStor and systems like it are quite clearly limited by the speed and bandwidth of network in which they operate and to some extent the performance of the machines they operate on. The latter factor is more of an issue when using distributed indexes, which require a certain amount of processing. In good conditions, the system can handle a query every second, retrieve 3 GB of data in under 30 minutes and ensure file integrity with up to 30% data loss. Performance deteriorates somewhat when using older hardware and slower/very busy networks, an aspect that could be improved by optimising some operations, such as the local index update and system waits.

Given the above conclusions, the usefulness of a distributed storage system such as MantleStor is dependent on the precise context of its use and the needs of a particular researcher. If data must be transferred over a typical 10/100 Mbit/s Ethernet network, the distribution time penalty is in effect already being paid, therefore MantleStor may be applicable. An example of this situation would be if the data has been generated at a remote computing site such as a national supercomputer. However, if data is produced on a particular supercomputer and will be visualised on the same machine, then it is only worth incurring the time cost of distributing the data if there is nowhere else it can be stored. This balance may change as institutional networks and Internet infrastructure in general, become faster.

Chapter 7

The e-Science Project, Reflections at Half Time.

7.1 Abstract

The use of computing in science has a long history. Modern computer science and science are, however, more disparate then one might imagine. The two disciplines have somewhat different philosophies and priorities. The e-Science project is an attempt to bring together computer scientists and scientists in order to take advantage of the potential of high performance computing and the Internet. This chapter considers both the e-Science project within Earth sciences in the UK and a case study of a small-scale e-Science implementation. Although the overall e-Science project is far from complete, a number of observations can be made. Computer scientists and scientists need to better understand each others' approach and methodology. Strength and depth must continue to be built up in interdisciplinary areas. This observation

is particularly applicable to Earth science where the challenges of understanding the planet require both fundamental and applied research across many different fields. Finally, support for projects must be maintained long term. The potential is great but a lack of good interdisciplinary contact and research could lead to valuable opportunities being squandered.

7.2 Introduction

The task of realising the potential of modern computing is one of the most important challenges in science today (e.g. Nature, Editorial 2008). The ultimate aim is to integrate computing to such an extent that a scientist will be able to use extremely powerful computing when it is required without any complication and from any device anywhere in the world. Needless to say this is a very ambitious goal. The term e-Science (sometimes written *e*Science) is used to cover this truly vast field of interdisciplinary research broadly centred on work at the interface of science, computing and internet technologies. A further motivation of e-Science is to extend computational methods into non-traditional research sectors. The term is dominantly used in the UK, the US equivalent being 'Cyberinfrastructure'.

The universe is increasingly seen as a mass of potential data awaiting collection and processing through the lens of computing in order to make sense of it. Modern interconnected sensor networks and analytical machines are capable of producing volumes of data unheard of in human history. A key example, predicting the impact of anthropogenic climate change, is an integrated data collection, analysis and computational

modelling challenge. The conceptual background for this type of integration can be seen in the suggestion by the mathematician/astronomer Pierre-Simon Laplace (1749-1827) of an 'intellect' with knowledge of the position and state of all components of the universe, which would be able to calculate the future using the laws of physics. In practice this is of course impossible. However, high performance computing and high resolution global data collection may well be able to approximate the future of Earth's climate on a time scale useful to humanity.

Understandably, politicians and other high level research funders have been very keen to promote e-Science and significant funds have been injected into e-Science activities (e.g. Natural Environment Research Council (NERC) e-Science programme[1]). e-Science centres were established throughout the UK. e-Science is seen as important by Universities and many have prioritised a range of projects aimed at bringing more e-Science into their researchers' professional lives (e.g. ARCCA[2]). The examples outlined above lead to the central question posed by this chapter: what is e-Science, how has it been implemented so far and what are the implications for science in the future?

This chapter explores this question from a combined Earth Science and computer science perspective. A number of larger projects are discussed in order to understand how e-Science is being implemented. A small scale e-Science project is presented as

[1]The NERC e-Science programme ran between 2001 and 2007 and spent £14.2 million supporting 8 projects, 2 research centres and 17 studentships. *Source: NERC publication, e-Science: Harnessing the power of the internet for environmental research.*

[2]ARCCA - Advanced Research Computing @ Cardiff, is a combination of high performance computing hardware and programming support for any researcher who needs such a service in Cardiff University. The investment is of order several million pounds.

a case study.

Glossary

- Scientist: throughout this chapter, scientist is used to indicate Earth scientists, biologists, chemists, physicists etc. sometimes referred to as 'hard' sciences.
- Computer Scientist: Computing researcher who is not also a scientist (as defined above).
- Grid Computing: The concept of using computing power with the same ease as electricity from the power grid.
- The Grid: The high speed network of supercomputers built primarily to process data from CERN Large Hadron Collider experiments. Other uses include video conferencing and computational modelling.
- Cloud Computing: commercial variation on grid computing. It includes 'software as a service' where maintained online applications such as word processing are available either free (advertisement supported) or for a fee. Further examples include Amazon's S3 and EC2, which allow users to cheaply rent network storage and use remote multi-core workstations on a pay as you go basis.
- Web 2.0: A collective term for websites offering online applications such as: photo and video sharing, blogs, social networking and mapping. First use of the term is usually credited to Tim O'Reilly, founder of O'Reilly Media. Such sites are often highly dependent on user generated content.

- Semantic web: The idea that language/terminology can be broken down and presented such that a computer can understand it as a human would. E.g. understanding the difference between: 'is the cat black?' as opposed to 'the cat is black'.

- Middleware: Software that connects other pieces of software together, often linking disparate systems of different provenance or as a way of supporting multiple user interfaces (UI) for a single underlying service.

- Wiki: a user created and editable online content management system. Wikipedia is a typical example.

7.3 What is e-Science?

Scientists have been using computing of some form for a long time. The first 'computers' were humans who worked on paper, carrying out calculations and manually processing data such as astronomical images (Nelson 2008). These were followed by mechanical computers such as the (unfortunately overlooked) K1 built by Herman Hollerith. Finally, electronic computers such as the WW2 British Colossus, the US ENIAC and the Manchester University 'Baby' brought in most basic concepts of modern computing. The successful demonstration of the transistor by 1948 lead to the development of the integrated circuit and the type of processors we know today. Developments to date culminate in the construction of the internet, loosely linking most of the world's computers. At every stage, science has been intimately involved in both development and exploitation of the new technology. Generally, mathemati-

cians and physicists were in the vanguard as their work often depends on being able to process a large number of calculations. More recently, the demands of large datasets and the modelling of complex interactions has pulled a significant portion of hard sciences into the computational sphere.

The first difficulty is in separating e-Science from everyday computing. The line is fuzzy; one could consider collaborating with researchers across the globe via email to be a form of e-Science. However, e-Science generally focusses on the most demanding applications from a data and distributed processing aspect. It has been variously defined by different bodies; in fact some avoid defining the term at all. The Research Councils' UK website[3] states that e-Science is: 'the systematic development of research methods that exploit advanced computational thinking'. NERC states that e-Science is: 'the new way of working for scientists in computational and data intensive fields' (slightly re-phrased). For the sake of this work e-Science is defined as: 'using the power of modern computing and networks to enable scientific research that would not otherwise be possible'. The key word in this definition is 'enable'; there is little to gain from developing innovative methods if they do not translate into research.

An example of classical scientific computing is 3D spherical geometry Earth mantle modelling. TERRA (Baumgardner 1985, Bunge & Baumgardner 1995, Yang & Baumgardner 2000), one of the the older codes, has been in development since the early 1980's. It is highly scalable such that, for some questions, one is only limited by the size of the supercomputing cluster at hand. Other limitations originate from

[3]www.rcuk.ac.uk

a lack of full understanding of some more subtle physical phenomena such as the self consistent generation of rigid tectonic plates on Earth's surface. The mode of operation for such a code is simple. Users log in remotely to a large supercomputer, upload source code, compile and run from a Linux/UNIX command line. This process has been used for a long time and can cause problems for new users. A typical Web 2.0 software engineering approach might be to build some sort of UI, possibly web-based, with grid access such that the user could specify input parameters and submit model runs without concerning themselves with details such as compilation or even where the code is running. The problem with this approach is that codes such as TERRA are by their nature non standard. Every user requires different features in their version of the code. What TERRA users actually benefit from is a subversion system to maintain the source code and a Wiki to maintain basic 'how to' documentation for new users and more advanced comments regarding new features of the code. This example highlights the need for an informed approach when tackling e-Science problems. Scientists are not the average computer user and applications are not a science version of Facebook[4]. Scientists have particular requirements and non standard skill sets; e-Science solutions must work with them.

There is a certain degree of debate as to whether e-Science tends to be 'push' or 'pull' in nature. Developments in computing are very rapid and technology is often available before the scientist knows how to use it effectively. The nature of procurement programmes also leads to situations where resources are available before researchers are able to make best use of them; this is technology push. Conversely,

[4]www.facebook.com, one of the current crop of 'buzz' social networking sites.

there are many problems in science, which are in desperate need of a computational solution. For example, many legacy datasets need processing and making available online; this is science pull. From the scientist's viewpoint, science would pull and computer technology would provide the solution. However, when considering complex problems there may be more innovative solutions possible than the scientist can conceive. Thus, computer science must also independently develop technologies, which are not initially aimed at a specific science problem.

7.4 Who is Involved?

e-Science can be viewed as the process of bringing the relatively young discipline of computer science into a closer relationship with older sciences such as Earth science (or vice-versa). This is a somewhat complicated process as computing has a different philosophy from traditional sciences. The difference is largely due to a strong engineering legacy, which is evident in the approach and language of computer science. For example, a hypothesis in computing refers to a statement about a problem e.g. 'I need a way of distributing 100's of GB to 1000's of users' and a proposed solution, which will be attempted. If the solution is found to be effective it may be more widely applied. In traditional science, a hypothesis is a reasoned assumption about the way a given system operates e.g. 'mantle plumes are the cause of hotspot volcanism', which is then vigorously tested in an attempt to disprove it. If it survives repeated testing it will enter established theory. This example highlights a potential difficulty, which can arise in interdisciplinary collaborations. The scientist is attempting to test a concept, while the computer scientist is attempting to build something and learn more

about the way similar problems may be solved. This can lead to a situation where the scientist expects to merely obtain a tool from the computer scientist, which they can use to help test their hypothesis. However, the computer scientist is much more interested in the process of finding the solution; building the tool is often considered more of a proof of concept. Production-ready software tools are generally built by software engineers, not by scientists or computer scientists. This introduces a third actor in successful e-Science, however the contribution of the software engineer may be relatively invisible, for example if an Application Programming Interface (API) has already been written.

7.5 Examples of Larger e-Science Projects

Many projects to date have been as much about developing capabilities as actually doing science. The majority of large projects have focussed on middleware, with the intention that it will be recyclable into other projects. The intent has been to develop a loosely associated tool set for e-Science within the various target disciplines. A selection of larger projects have been grouped by type and briefly described below.

7.5.1 Data Organisation, Assimilation and Availability

A large number of projects take the data deluge provided by modern analytical and monitoring technology as their motivation. Organising and storing such data in an accessible and durable manner is essential if expensively collected data is to be fully exploited. The NERC DataGrid[5] provides a data discovery service, which

[5]http://ndg.badc.rl.ac.uk/

consists of a catalogue of Earth and environmental science data indexed by location (latitude/longitude), time and keyword. The important component is the middleware, which connects the data providers to the search system. No data resides on the NERC data grid so it can be considered as digital plumbing. Data providers such as the British Geological Survey and the British Oceanographic Data Centre retain control over their data; other types of data service maintain a database and provide a web front end for searching. Bioscience services such as Flybase[6] and Fishbase[7] are typical examples. They act as central repositories for DNA sequence and species data respectively. Many such services are implemented as a variation of the Wiki format, which allows users to add and view information in a consistently structured manner.

7.5.2 Distributed Processing

When modelling a physical system, it is often necessary to run many instances of a model in order to better understand the relative sensitivities or to map out a particular parameter space. The problem with such an approach is the high throughput of computing required. If the model is only suitable for high performance supercomputers then such a task is best postponed until the next generation of hardware. However, if the model can be sufficiently simplified to run on a desktop machine without compromising key features, then a distributed computing approach can be used. The are two main methodologies. The structured approach, utilising a system such as Condor[8], can be used in an institutional network setting. A pool of workstations

[6]http://flybase.org/
[7]http://www.fishbase.org/
[8]http://www.cs.wisc.edu/condor/

running client software are organised by a central server, a list of jobs and the codes required are submitted and the system runs them as efficiently as possible over the available machines. If such a system is not available or indeed if the few thousand workstations in a typical Condor pool are not sufficient, an even larger scale solution may be used. ClimatePrediction.net[9] is a worldwide climate modelling project where the computing power is provided by ordinary people voluntarily installing the software and allowing it to use their computers when they are idle. This highly unstructured distributed computing method employs the BOINC[10] software API and is reliant on the goodwill of the volunteers who provide the computing resources. The volunteer approach is only applicable for some types of work; the high public profile of ClimatePrediction.net has been key to its success. The processing of any type of confidential information could not be done with such a public system. The distributed computing approach is also useful for data processing, the SETI[11] project being a typical example. As desktop machines continue to become more powerful (multi-core, large RAM size, GPGPU[12] etc.), this type of project is likely to become much more effective.

7.5.3 Semantic Web

The term semantic web covers a variety of projects, which are concerned with moving common human classification systems such as language into a consistent framework for use by computers. These classification systems arise from the manner in

[9]http://www.climateprediction.net/

[10]Berkeley Open Infrastructure for Network Computing, http://boinc.berkeley.edu/

[11]Search for extraterrestrial intelligence, http://setiathome.berkeley.edu/

[12]General purpose graphics processing unit, using the power of a graphics card to accelerate processing.

which the human brain functions, therefore a certain amount of work is required to help computers 'understand' them. A typical example is the species classification system defined by Carl Linnaeus (1707-1778). The CATE[13] project aims to use the taxonomic system within a database and web portal, which will allow both consensus and alternative taxonomic hypotheses to appear together. This degree of flexibility is necessary as there is significant debate over relationships between related species, particularly with the advent of genetic sequencing. Thus it is desirable to be able to view and manipulate all proposed species phylogenies (trees) together. From a technical standpoint, the database must be able to link to other initiatives such as DNA catalogues or smaller specimen datasets. Programmers must also try and anticipate future developments as data must be easily adaptable and re-usable.

A related issue is that of ontology, a term borrowed from philosophy, which is the study of the nature of existence and its classification. In e-Science it concerns classification and description of data and how such classification is defined. Linnaeian taxonomy is an example of an efficient ontology as organisms are classified into a fairly rigid framework. Other classification systems in science are less well defined and can lack structure. It is necessary for such classification systems to be formalised in some manner that is machine readable. For example, a machine requires some way of understanding that the rock description: 'coarse grained rock composed of quartz', applies to a sandstone and should be associated with other similar rocks.

[13]Creating a Taxonomic e-Science, http://www.cate-project.org/

7.5.4 Visualisation

The ability to visualise results is one of the most important aspects of science. Anything larger than a table of ≈10 numbers is impossible for most people to comprehend, whereas a graph or map plot is instantly accessible. The reach of the internet results in extreme amounts of data being potentially accessible. The challenge lies in how to process and present such data in the most efficient manner. Significant time may be spent preparing a single figure for publication by hand, doing that thousands of times is impossible. Some datasets must be analysed in their entirety, for example the TERRA mantle modelling code (described in section 7.3) produces many gigabytes of spatial and temporal data. Terra data requires a HPC cluster to visualise and render information into a more compact format for exploration, figure production and animation. TERRA data is quite unusual in this respect; many datasets are either of smaller (e.g. regional) extent or consist of discrete data. Figure 7.1 demonstrates the difference between the two classes of data. For many applications the user only requires a subset of the original data; the problem then becomes one of data discovery, exploration and extraction. Data discovery is covered by middleware systems as described in section 7.5.1. If visualisation by data extraction and presentation is added to a data discovery, a valuable tool can be constructed. Access to this type of system is often provided via a web portal such as the NERC funded GODIVA2[14], which integrates ocean models and observational data on a range of scales and makes it available in a 'point and click' interface. More advanced visualisation may be carried out using links to Google Earth or Google Maps. Data available include scalar fields

[14]http://behemoth.nerc-essc.ac.uk/ncWMS/godiva2.html

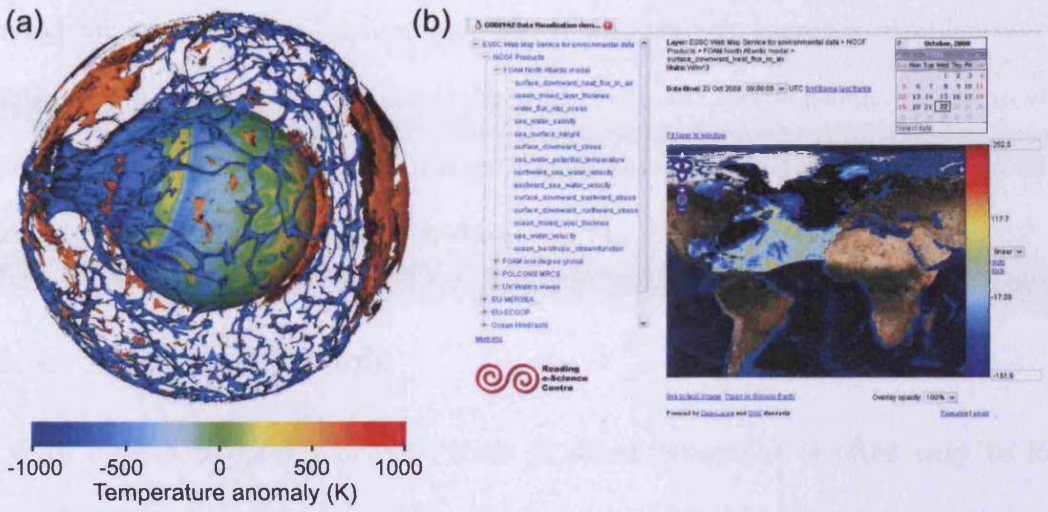


Figure 7.1: (a) An example of TERRA visualisation. Temperature anomaly data from a single time step: 1.5 GB in size, consisting of over 300 million data points, processed using 128 cores of a HPC cluster. A typical model run might contain 30,000-40,000 time steps, all potentially available for visualisation. (b) An example of GODIVA2 portal, showing slective visualisation of 2D ocean model data. URL: http://behemoth.nerc-essc.ac.uk/ncWMS/godiva2.html

such as sea surface temperature and ice volume from a wide range of sources. There are significant advantages to the portal interface: data transfer costs are reduced as only summary data are downloaded; processing occurs remotely so any device capable of running a browser can access the service. New data can be automatically added as they become available. Technically, there is some similarity with GIS (Geographical Information Systems), but such portals are conceptually different in that they are designed for exploring very large datasets, whereas GIS are concerned with the processing and analysis of spatial data. A disadvantage of the portal approach is the amount of effort involved in maintaining and updating the datasets and building the tools to process the many different types of raw data into a usable format. A further limitation lies in the trade off between simplicity (essential for ease of access) and the

demand for as much data as possible, which will tend to increase complexity. This problem is mitigated by linking to programmes such as Google Earth, which provide a more flexible environment. In the future, larger touch screens/surfaces should enable more intuitive manipulation of such data.

7.5.5 High Level Tools

With all the complex e-Science tools in development, it is often easy to forget that researchers requirements will sometimes go beyond anything a generic service can provide. Traditionally this would require custom building their own model or processing tool. If they were fortunate an environment such as MATLAB would be sufficient but more likely the use of a programming language like FORTRAN or C++ would be required. There are however alternatives, which are best described as high level tools. High level tools attempt to simplify the process of building semi-bespoke models and data processing tools. As these tools should be accessible to as many researchers as possible, there is a strong dependence on appropriate UI. Some examples follow.

The Cactus Code[15] is a computational modelling tool kit that includes pre-written modules to handle many functions the scientist would otherwise have to write. As a result, only specific custom modules require hand coding, reducing the time taken to produce a code. Examples of uses include: modelling black hole collisions, smooth particle hydrodynamic methods and a variety of relativistic physics applications.

[15]http://www.cactuscode.org/

GENIE[16] is a more applied, NERC funded attempt to integrate a range of Earth system (climate, ocean, carbon cycle etc.) models into a framework that can run across the grid. The aim is also to permit high level open access to this type of modelling.

The Triana project[17] is a workflow programme, which allows data processing with pre-written modules that are linked together using a graphical UI of blocks representing different tools. Advanced support allows the use of network resources via web services directly from the work flow environment. It has been used for signal and image processing.

7.6 Reflections on e-Science

Having introduced e-Science and covered a selection of projects, it is worth restating the initial question: what is e-Science, how has it been implemented so far and what are the implications for science in the future? The what and the how have been covered, the rest of this chapter discusses the implications of e-Science. The view advanced here is that although some of the tools developed by the e-Science programme are excellent, tools are not necessarily the key benefit from an e-Science project. Tools tend to change with the problem at hand and even with what happens to be in fashion. What remains are: contacts between very different researchers, the

[16]http://www.genie.ac.uk/
[17]http://www.trianacode.org/

skills acquired by teams and individuals and most importantly, a change in the way scientists think about their use of technology.

7.6.1 Case Study: MantleStor

Storing and managing large datasets is and will remain a major preoccupation for the scientific modeller. Data storage capacity increases rapidly as hard disk capacity climbs. This is matched by improvements in performance of processors and RAM size, which lead to ever higher model resolutions. The problem is particularly notable in 3D modelling. If one halves the resolution (e.g. 10 km to 5km) the number of data points increases 8 times (Figure 7.2). More importantly, the size of output and checkpoint files also increases 8-fold. For many research groups, funding to buy large amounts

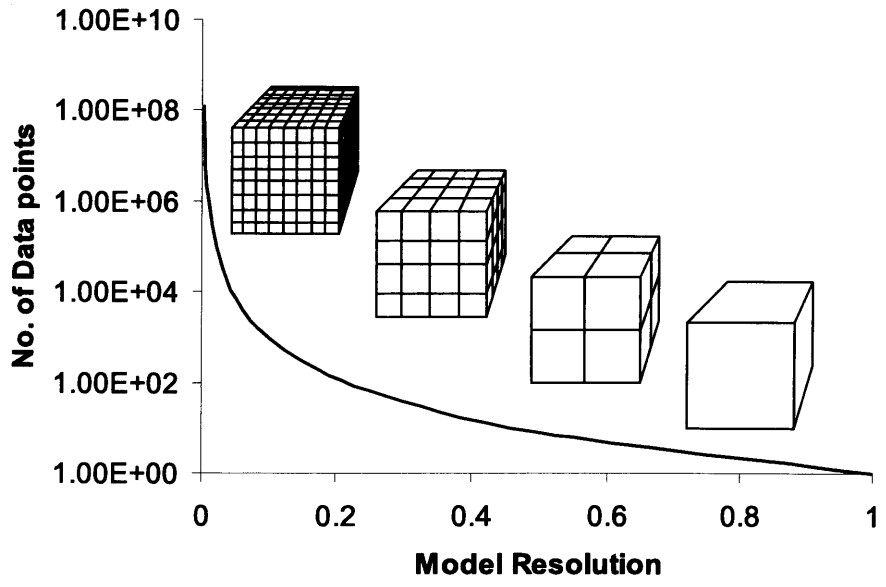


Figure 7.2: Illustrating the rapid growth of data points as 3D model resolution is increased. Note log axis on vertical.

of storage is often limited. There is therefore a need to search out new methods to access more storage. The MantleStor project explored one possibility, to access underutilised storage space on network workstations and consolidate it into a usable storage space. The technical solution exploited peer to peer (P2P) technology to allow each workstation to make its unused disk space available to a central controller. The initial concepts of the MantleStor system have been published (Wolstencroft et al. 2006).

Methodology

The system needed to be able to operate in an environment of mostly standard MS Windows XP workstations, located in rooms of ≈60. The maximum envisioned deployment environment would comprise several hundred machines. There are obviously a wide range of methods, which could have been used to spread data files over such a number of machines. Some rely on OS kernel level modification and are often Linux based and intended for high performance distributed file systems like Lustre (Simms et al. 2007). Other systems such as Storage Resource Broker (Rajasekar et al. 2002) are generally intended for a Grid environment. These approaches were considered too complex for the MantleStor system. Instead, a lightweight Java based P2P system was used. The underlying P2P infrastructure was provided by P2PS (Wang 2005) a P2P API. Each participating machine runs a daemon process, which maintains a local storage space, an index of files and maintains a connection to the network. A 'client' version of the application is run on the users' workstation, allowing data to be stored, searched for and retrieved. A basic but fully functional

implementation of MantleStor was tested using a small pool of machines to establish search speed, data transfer performance and system resilience.

Outcomes

MantleStor was successfully tested and found to be resilient in a test bed scenario (Chapter 6). However, P2P communication difficulties on the wider institutional network prevented a full scale deployment. The difficulty arose during the project as a result of changing network infrastructure; early versions of MantleStor did not experience such difficulties. This highlights a common problem in e-Science, that key underlying hardware can easily change beyond the control of the project.

As a further example of the risks inherent in e-Science, the wider project of which the P2PS API was a part unexpectedly came to the end of its funding during the course of the study. The lead programmers found employment elsewhere and the project was effectively frozen. This complicated the construction and testing of the final version of MantleStor and seriously limits future development of MantleStor as it would require re-writing with a different and better supported API. This provides a cautionary tale, highlighting the importance of choosing a long-term base upon which to build. This is a difficult problem to deal with due to the rapid overturn of new ideas and software in the computer science world.

The available resources also changed during the project. At the start of the project a single small-medium sized Linux cluster with 200-300 GB of storage was available

to the project. At the time of writing this had expanded to 3 clusters, two of which entered the Top500 (06/2008 edition) at number 29 and 88 in the world respectively. The storage attached to these systems effectively removed the problem of data storage but highlighted a further issue, that of indexing. MantleStor features an indexing system, which tags data to be stored with metadata to enable data discovery. This was not only essential for managing data within MantleStor but also has implications for indexing model data in general.

TERRA is a very traditional code and that extends to output data handling. Summary output takes the form of ASCII files and detailed visualisation output is dumped as a set of files in a nominated directory. There is no archiving structure for the output of TERRA runs; collective analysis of cases is carried out using a variety of packages such as GMT, gnuplot and MantleVis. As it stands there is no method for easily discovering if a given combination of input values has been used in a previous case either by a single user, research group or the wider community. The method of indexing used by MantleStor is a step towards a model archive. The simple search tool in MantleStor allows users to search stored model runs based on a series of input parameters, something not previously possible. An extension of this sort of functionality, possibly without the large data storage aspect, could lead to a very valuable tool.

Using MantleStor as an example, a number of general observations can be made regarding small scale e-Science projects. Project outcomes are not particularly predictable, as the range of possible solutions to a given problem changes rapidly and

may overtake the original aims. The most valuable outcomes of a project may well be unintended. e-Science projects are by nature interdisciplinary, even the smallest projects (like MantleStor) have a sizeable number of personnel in the 'background' e.g. writing software libraries or APIs. The importance of these contributions is often overlooked until they are no longer available. Also, the distinction needs to be made between tools that need to be extensively developed to production level and those that are more ad-hoc in nature, the former requiring software engineering skills, the latter representing business as usual for most scientific hackers!

7.6.2 Building on e-Science

In the e-Science projects discussed above, science did provide the driving force (science pull), whether or not this involved using an established method or developing a new one. The question is, whether many e-Science projects are too complicated. An excellent quote from Albert Einstein illustrates a typical physicists approach: "Everything should be made as simple as possible, but no simpler"[18]. It is a test that should probably be applied to e-Science.

Infrastructure projects such as the National Grid Service, improvements in JANET and the HECToR supercomputer have provided top quality services at a national level, leaving the UK in good shape for the future. These projects represent generic capability, which provide a basis on which to build. A number of UK research council funded projects described in section 7.5 have been extremely successful, for example,

[18] Alles sollte so einfach wie möglich gemacht sein, aber nicht einfacher

ClimatePrediction.net and GODIVA2 (both NERC). The longer term legacy of the e-Science programmes is less clear. The Open Middleware Infrastructure Institute (OMII[19]) is an organisation supported by the UK research councils to re-engineer some of the middleware developed by various projects to make it more widely available (EPSRC 2007). The interesting point is that middleware is intended to be fairly generic and therefore one might not expect that it would need re-engineering. Were the original projects too complicated? What the example highlights, is the difficulty of doing something new. Often, once a novel piece of software has been written and tested, it requires extensive revision before it can be considered mature. Further complexity arises as it is tempting to state at the beginning of a project, that it will produce reusable components, but much more difficult to actually do this. This does not indicate that a particular project has failed, just that it should perhaps have had better defined goals. Tools are forever changing so skills and adaptability contribute much more to the capability of a research team than any particular piece of code. The NERC e-Science programme is complete, but the e-Science project as a whole has some way to go. With regard to the future, one would hope that the funding bodies will maintain support for e-Science work. NERC now have 7 funding themes, only one of which (Technology) explicitly mentions computing, although all themes will need to make use of it. After all the excitement of the e-Science programmes it would be unfortunate if in the minds of Earth scientists, computer science is relegated to playing a supporting role.

From a participant point of view, successful e-Science is a two way process, computing

[19]http://www.omii.ac.uk

becoming embedded in science and vice versa. This will of course only work if each part seeks to understand the others' discipline. A generalist natural scientist with computing skills and an awareness of software design might be the perfect eScience researcher, but such people are rare. Science in particular has become extremely specialised, mainly in response to the weight of knowledge needed to even understand where the forefront of understanding lies. If e-Science is to really succeed, researchers who understand how predictive, observational and engineering sciences interconnect are required to mobilise techniques and skills across traditional scientific boundaries. Such people have existed in the past; Johann Wolfgang von Göthe (1749-1832) is remembered primarily as a writer, poet and philosopher but he also made considerable contributions to science with his works on the nature of light and plant morphology. Göthe was of course an exceptional individual, but he serves as an example of a time when academia was less isolationist. The difficulty lies in breaking down structural divisions between disciplines. The 'interdisciplinary trap' is mainly a funding and career progression problem. Outside of specific initiatives (e.g. NERC e-Science) funders who will support method development are difficult to find, due to the current near-obsession with applied science and its economic benefit. As an example, the former UK chief scientist Sir David King used his address at the BA science festival 2008 to call for a focus on applied fields, at the expense of blue skies research (Amos 2008). Once funding has been found and work done, publication can also be difficult as top journals in a particular field may not see the value of interdisciplinary research. This may lead to difficulties when applying for academic posts, which demand a certain level of publication in an expected range of journals (Whitfield 2008).

Many of the problems outlined above could be solved by education. Science requires a stronger emphasis on educating students in 'proper' computing such as programming and scripting, rather than simply focussing on particular commercial software packages. This should probably start in schools, as the digital future requires citizens with full computer literacy. As an example, the world's rapidly growing body of digital information requires an army of digital curators to archive and preserve it. Whether the work of the present survives as long as that of the Greek philosophers will be largely up to these digital curators.

7.7 Summary

This Chapter has provided a general overview of e-Science and some realisations of the concept, mainly centred on Earth science applications. The greater e-Science project is still in its infancy. Its development in the future will be heavily influenced by technological innovation, which by its nature, is unpredictable. Philosophically, it is clear that science and computer science must come even closer together. Individuals who are equally at home in the laboratory, in the field and writing code may not be in great supply but effective interdisciplinary teams should be able to operate in a similar manner. A broader issue is whether sufficient funding will be committed in the long term. Large amounts of money have been spent fostering e-Science projects over the last decade. If a decent level of support is not maintained, skills will be lost. If the e-Science project is maintained, science and computer science will be able to work together to produce truly useful methodological innovations. If the project falters,

then science will have to make do with whatever techniques computer science happens to develop and computing will struggle to access a valuable application space. With sufficient motivation, there may come a time when there will be no need to make a distinction between computer science and science; at that point the e-Science project will be complete.

Chapter 8

Summary

The main aims of this work were to advance understanding of Earth's thermal evolution and to investigate technical aspects of high resolution mantle modelling with particular reference to data indexing and storage. The project as a whole was dominated by computational techniques, with a very strong reliance on: computational methods, high performance computing, visualisation, data handling and analysis. It was access to technology that enabled novel investigations into the geodynamics and history of the deep Earth.

It has been demonstrated that investigating the deep Earth with the 3D spherical geometry code TERRA is a very data intensive exercise. The project has produced in excess of 1.5 terabytes of data, mostly in the shape of visualisation and checkpoint files. In 2005 this data volume would have required almost the entire RAID array of the then available computing cluster. Therefore, a significant aspect of the work was focussed on the handling and storage of data. The scavenging of spare

hard disk space on communal network workstations was investigated as a possible solution. The resulting system, MantleStor, combined a number of techniques from computer science not previously used in Earth sciences. MantleStor is a peer to peer (P2P) based storage system, in which data are indexed in a distributed manner using model metadata. To improve performance and data integrity, data are split into 5 MB 'chunks' and duplicated 3 times across a number of storage clients. The index and search system allows an archive of model runs to be queried in a manner not possible with conventional storage. Peer to Peer Simplified, the P2P API used, is based on XML messages, which allow a high degree of flexibility within the design of the system.

In good conditions, MantleStor can handle one search per second and retrieve 3 GB of data in under 30 minutes. The duplication system ensures file integrity with up to 30% loss. Performance deteriorates somewhat when using older hardware or slow networks due to delayed XML messages and slow updating of the distributed index. The usefulness of a distributed storage system such as MantleStor is dependent on the precise context of its use and the needs of a particular researcher. If data must be transferred from a remote machine in any case, then the cost of using a distributed storage system may not be high. However, if data are able to reside on storage directly attached to a cluster, the cost of distributing them may be too high. Therefore, the applicability of MantleStor is somewhat dependent on the task and while distributed storage is an attractive idea, it may not suit every scenario.

MantleStor was, however, not deployed beyond a test bed in part due to the availabil-

ity of new resources, which were not anticipated or indeed expected at the beginning of the project. Access was obtained to HECToR (UK national supercomputer) and Merlin (Cardiff University HPC cluster), where several terabytes of local data storage could be obtained. The indexing and archive aspect of MantleStor became of greater interest and it is clear that an extensive archive of TERRA model cases should probably be constructed at some point in the near future. The experience of developing MantleStor highlights the ephemeral nature of such e-Science projects; often the original goal is overtaken by technology.

Chapter 7 provides a reflection on e-Science in general, with MantleStor as a primary case study. Other, mostly Earth science, examples were also discussed. General conclusions centre on the relationship between computer science and the more traditional sciences. The wider 'e-Science project' is still in its infancy; science and computer science need to move still closer together. One of the most effective ways of doing this is via well supported interdisciplinary teams. As long as the wider e-Science project is maintained, science will be able to influence the path of a significant strand of computing and computing will have applications for which to develop solutions. Computer science techniques must continue to be effectively harnessed to accelerate our understanding of the Earth and its processes. The cost to science of not doing this could be significant.

The Earth is an example of a small rocky planet orbiting a medium sized star. For obvious reasons it is by far the best studied planet, although the study of Venus and

Mars is catching up. At the time of writing, astronomy has inferred the existence of at least 300 planets orbiting other stars and may have directly imaged 4 (Kalas et al. 2008, Marois et al. 2008). In the near future, exoplanets of similar size and composition to Earth are likely to be detected and it will be the task of geophysical modelling to study these bodies. In order to understand the emerging 'planetary zoo' one must first understand Earth. The evolution of Earth is in itself a complex problem. In this work, a range of possible mechanisms influencing the thermal and dynamical evolution of Earth were considered using high resolution computational modelling.

One of the most fundamental constraints on the thermal history of Earth is the heat transfer efficiency of the mantle. Silicates are poorly conductive so the bulk of heat transfer within the mantle occurs by convection. The efficiency of the convective system, represented by the Nusselt-Rayleigh power law relationship: $Nu = aRa^{\beta}$, can provide a constraint on how much heat Earth is able to lose. Estimating this efficiency analytically is difficult owing to the geometry of the mantle and the many conditions which must be satisfied. Therefore high resolution computational modelling in 3D spherical geometry is one of the few methods, which can provide an answer. Both basally and internally heated systems were modelled to an Earth-like Ra and the basally heated β was found to be 0.29, a slightly lower value than the the value suggested by boundary layer theory of convection ($\frac{1}{3}$) but broadly consistent with recent literature. With $\beta = 0.29$ as opposed to $\frac{1}{3}$, the Nu and surface heat flux are reduced by 32% and 55 TW respectively at Ra $= 10^9$. Diffuse heat input in internally heated

models and the smaller aspect ratio of the resulting convection cells, appears to be responsible for β being closer to $\frac{1}{3}$ ($\beta = 0.337$) for such cases. The differing aspect ratio of basally heated and internally heated models results in consistently lower surface velocities for internally heated models. It had been suggested that β reduces at high Ra, however, this study found no evidence to support that hypothesis. Other mechanisms are therefore required to resolve the problem of back-projected mantle temperature tending to infinity before 2-3 Ga in parameterised thermal evoloution models.

As the intrinsic efficiency of convection in a spherical shell cannot provide a moderating influence on mantle heat flux in the past, other mechanisms are required. A source of much debate regarding the mantle has been the interpretation of geochemical evidence suggesting the presence of a 'primitive' reservoir of material. It had been proposed that the entire lower mantle is isolated from the upper mantle, providing this reservoir. Unfortunately, the elegance of this solution is disrupted by convincing seismic tomography images showing whole mantle convection in action. The conceptual model of Allègre (2002) integrates these two end member views, hypothesising that the mantle has been layered to some degree in the past and has evolved to whole mantle convection today. The most common mechanism, which is invoked to allow evolving layering is the interaction of the convective vigour of the mantle (Ra) with the 660 km mineral phase change. Decreasing convective vigour (Ra) over time theoretically allows a transition from layered whole mantle convection. The relationship between 660 km phase change Clapeyron slope and Ra has been mapped out over a

large range of both parameters. It was found that partial layering of some degree is possible at present day Ra. Some seismological observations indeed appear to indicate some slab deflection at 660 and an effect on plume transit across 660. With regards to the Allègre model, past layering is supported by modelling results. However, strong layering is unlikely to have survived much after 3 Ga. It is unlikely that any large scale heterogeneity in the mantle would have survived to the present. Therefore, other explanations are required for some of the observed geochemistry. For example, a better understanding of the behaviour of key elements such as Helium during melting is emerging, suggesting that assumptions may need to be revised (Parman et al. 2005).

Past layering of the mantle appears possible and the present is likely to be weakly partially layered at best. It is interesting to consider what the impact of a change from layered to mostly whole mantle convection would have. To investigate this mechanism, a series of models were run, which crossed the previously identified threshold values of Ra. The transition from layered mantle convection to partial or whole mantle convection at high Ra range is characterised by significant periodic fluctuation of surface heat flux (SHF), driven by mantle avalanching. This contradicts the suggestion that such a transition could go almost unnoticed (Allègre 2002). The fluctuations represent transition of the mantle into a partially layered regime, rather than a dramatic switch from layered to whole mantle convection. The range of Ra over which partial layering is possible is large at Earth-like Ra, indicating that the mantle could have spent a significant period partially layered. Mantle avalanches, initiated at the 660 phase boundary in response to cold material ponding above, drive periodic

'pulses' in the SHF. The effect of an avalanche is to sweep hot material together at the core mantle boundary, spawning significant plumes in the opposite hemisphere. The mechanism operates in a 'pulse-recharge' manner producing a series of heat inputs into the upper mantle region. The mechanism of periodic layer breakdown has been shown to be consistent across a range of Ra and 660 Clapeyron slop values, supporting the conclusions of previous parameterised modelling (Davies 1995). The interval between 'heat spikes' appears to increase with increasing Ra (going back in time) and is of order 100-200 Myrs, similar to the time scale of supercontinent formation and the interval between large igneous provinces. Models where the mantle was permitted to cool produced a succession of SHF pulses where the initial 2-3 pulses were of similar high magnitude followed by smaller oscillations. This trend corresponds to a reducing degree of partial layering. Should Earth have undergone a similar process, one would expect the Precambrian to show evidence of periodic global melting events which reduce in magnitude over geological time. Global episodic mantle depletion events have been inferred from some geochemical observations (Parman 2007).

The main implication of the work presented here is that non-linearity could be very important in the thermal history of Earth. If Earth's mantle was initially layered and has passed through a partially layered regime, it is likely that there have been global scale mantle avalanche and melting events. This suggests that the smooth average mantle temperature plots produced by many thermal evolution models may not adequately capture Earth's thermal history. Non-linear heat flux from the mantle would also mandate extreme care when using the present state of the mantle to infer its

state in the geological past, as modern heat flux measurements could be misleading if taken out of context.

Bibliography

Abbott, D. H. & Isley, A. E. (2002), 'The intensity, occurrence, and duration of superplume events and eras over geological time', *J. Geodyn.* **34**, 265–307.

Aberer, K., Datta, A., Hauswirth, M. & Schmidt, R. (2005), Indexing data-oriented overlay networks, *in* '31st International Conference on Very Large Databases (VLDB)', Trondheim.

Adya, A., Bolosky, W. J., Castro, M., Cermak, G., Chaiken, R., Douceur, J. R., Howell, J., Lorch, J. R., Theimer, M. & Wattenhofer, R. P. (2002), FARSITE: Federated, Available, and Reliable Storage for an Incompletely Trusted Environment, *in* '5th Symposium on Operating Systems Design and Implementation', The USENIX Association, Boston, Massachusetts.

Allègre, C. J. (1997), 'Limitation on the mass exchange between the upper and lower mantle: the evolving convection regime of the Earth', *Earth Planet. Sci. Lett.* **150**, 1–6.

Allègre, C. J. (2002), 'The evolution of mantle mixing', *Philos. T. Roy. Soc. A* **360**, 1411–1431.

Allègre, C. J., Hofmann, A. & O'Nions, K. (1996), 'The Argon constraints on mantle structure', *Geophys. Res. Let.* **23**, 3555–3557.

Allègre, C. J., Standacher, T., Sarda, P. & Kurz, M. (1983), 'Constraints on evolution of Earth's mantle from rare gas systemics', *Nature* **303**, 762–766.

Allègre, C. J. & Turcotte, D. L. (1986), 'Implications of a two-component marble-cake mantle', *Nature* **323**, 123–127.

Amos, J. (2008), "Climate crisis' needs brain gain'. 8/09/2008.
URL: *http://news.bbc.co.uk/1/hi/sci/tech/7603257.stm*

Andersen, D. G., Balakrishnan, H., Kaashoek, M. F. & Morris, R. (2001), Resilient overlay networks, *in* 'Symposium on Operating Systems Principles', pp. 131–145.
URL: *citeseer.ist.psu.edu/andersen01resilient.html*

Anderson, D. L. (2005), *Scoring hotspots: The plume and plate paradigms*, Vol. 388 of *Plates, plumes and paradigms: Geological Society of America Special Paper*, Geol. Soc. Am.

Ballentine, C. J., Marty, B., Lollar, B. S. & Cassidy, M. (2005), 'Neon isotopes constrain convection and volatile origin in the Earths mantle', *Nature* **433**, 33–38.

Batten, C., Barr, K., Saraf, A. & Treptin, S. (2001), 'pStore: A secure peer-to-peer backup system'.
URL: *citeseer.ist.psu.edu/batten01pstore.html*

Baumgardner, J. R. (1985), 'Three dimensional treatment of convective flow in the Earths mantle.', *J. Stat. Phys.* **39**, 501–511.

Bennett, V. C., Brandon, A. D. & Nutman, A. P. (2007), 'Coupled ^{142}Nd-^{143}Nd isotopic evidence for Hadean mantle dynamics', *Science* **318**, 1907–1910.

Bercovici, D., Schubert, G. & Glatzmaier, G. A. (1992), 'Three-dimensional convection of an infinite-Prandtl-number compressible fliud in a basally heated spherical shell', *J. Fluid Mech.* **239**, 683–719.

Bercovici, D., Schubert, G., Glatzmaier, G. A. & Zebib, A. (1989), 'Three-dimensional thermal convection in a spherical shell', *J. Fluid Mech.* **206**, 75–104.

Berry, A. J., Danyushevsky, L. V., O'Neill, H. S. C., Newville, M. & Sutton, S. R. (2008), 'Oxidation state of iron in komatiitic melt inclusions indicates hot Archean mantle', *Nature* **455**, 960–963.

Boyet, M. & Carlson, R. W. (2005), '^{142}Nd Evidence for early (>4.53 Ga) global differentiation of the silicate Earth', *Science* **309**, 576–581.

Boyet, M. & Carlson, R. W. (2006), 'A new geochemical model for the Earth's mantle inferred from ^{146}Sm - ^{142}Nd systematics', *Earth Planet. Sci. Lett.* **250**, 254–268.

Bunge, H. P. & Baumgardner, J. R. (1995), 'Mantle convection modeling on parallel virtual machines', *Comput. Phys.* **9**, 207–215.

Bunge, H. P., Richards, M. A. & Baumgardner, J. R. (1997), 'A sensitivity study of three-dimensional spherical mantle convection at 10^8 Rayleigh number: Effects of depth-dependent viscosity, heating mode and an endothermic phase change', *J. Geophys. Res.* **102**(B6), 11991–12007.

Busse, F. H. (1989), *Fundamentals of Thermal Convection*, Mantle Convection, Plate Tectonics and Global Dynamics, Gordon and Breach Publ., New York.

Butler, S. L. & Peltier, W. R. (2000), 'On scaling relations in time-dependent mantle convection and the heat transfer constraint on layering', *J. Geophys. Res.* **105**, 3175–3208.

Butler, S. L. & Peltier, W. R. (2002), 'Thermal evolution of Earth: Models with time-dependent layering of mantle convection which satisfy the Urey ratio constraint', *J. Geophys. Res.* **107**, doi: 10.1029/20000JB000018.

Carvazzoni, C., Chiarotti, G., Scandolo, S., Tosatti, E., Bernasconi, M. & Parrinello, M. (1999), 'Superionic and metallic states of water and ammonia at giant planet conditions', *Science* **283**, 44–46.

Castaing, B., Gunaratne, G., Heslot, F., Kadanoff, L., Libchaber, A., Thomae, S., Wu, X.-Z., Zaleski, S. & Zanetti, G. (1989), 'Scaling of hard thermal turbulance in Rayleigh-Benard convection', *J. Fluid Mech.* **204**, 1–30.

Choblet, G. & Sotin, C. (2000), '3d thermal convection with variable viscosity: can transient cooling be described by a quasi-static scaling law?', *Phy. Earth Planet. In.* **119**, 321–336.

Chopelas, A., Boehler, R. & Ko, T. (1994), 'Thermodynamics and behaviour of γ-Mg_2SiO_4 at high-pressure - implications for Mg_2SiO_4 phase-equilibrium', *Phy. Chem. Min.* **21**, 351–359.

Christensen, U. R. (1984), 'Heat transport by variable viscosity convection and implications for the Earth's thermal evolution', *Physics of the Earth and Planetary Interiors* **35**, 264–282.

Christensen, U. R. (1985), 'Thermal evolution models for the Earth', *J. Geophys. Res.* **90**, 2995–3007.

Christensen, U. R. (1998), 'Dynamic phase boundary topography by latent heat effects', *Earth Planet. Sci. Lett.* **154**, 295–306.

Christensen, U. R. & Yuen, D. A. (1985), 'Layered convection induced by phase transitions', *J. Geophys. Res.* **90**, 10291–10300.

Cohen, B. (2003), 'Incentives build robustness in BitTorrent'.

Coltice, N. & Schmalzl, J. (2006), 'Mixing times in the mantle of the early Earth derived from 2-D and 3-D numerical simulations of convection', *Geophys. Res. Let.* **33**.

Condie, K. C. (1998), 'Episodic continental growth and supercontinents: a mantle avalanche connection?', *Earth Planet. Sci. Lett.* **163**, 97–108.

Condie, K. C. (2004), 'Supercontinents and superplume events: distinguishing signals in the geologic record', *Phy. Earth Planet. In.* **146**, 319–332.

Condie, K. C., Marais, D. J. & Abbott, D. (2001), 'Precambrian superplumes and supercontinents: a record in black shales, carbon isotopes, and paleoclimates?', *Precambrian Res.* **106**, 239–260.

Courtillot, V. & Olson, P. (2007), 'Mantle plumes link magnetic superchrons to phanerozoic mass depletion events', *Earth Planet. Sci. Lett.* **260**, 495–504.

Davies, D. R. (2008), Applying multi-resolution numerical methods to geodynamics, PhD thesis, Cardiff University.

Davies, G. F. (1980), 'Thermal histories of convective Earth models and constraints on radiogenic heat production in the Earth', *J. Geophys. Res.* **85**, 2517–2530.

Davies, G. F. (1995), 'Punctuated tectonic evolution of the Earth', *Earth Planet. Sci. Lett.* **136**, 363–379.

Davies, G. F. (1999), *Dynamic earth : plates, plumes, and mantle convection*, Cambridge University Press, Cambridge, New York.

Davies, J. H. (2005), 'Steady plumes produced by downwellings in earth-like vigor spherical whole mantle convection models', *Geochem. Geophys. Geosys.* **6**, doi:10.1029/2005GC001042.

Davies, J. H. & Bunge, H. P. (2006), 'Are splash plumes the origin of minor hotspots?', *Geology* **34**, 349–352.

Davies, J. H. & Stevenson, D. J. (1992), 'Physical model of source region of subduction zone volcanics', *J. Geophys. Res.* **97**, 2037–2070.

Elliott, T., Thomas, A., Jeffcoate, A. & Nui, Y. (2006), 'Lithium isotope evidence for subduction-enriched mantle in the source of mid-ocean-ridge basalts', *Nature* **443**, 565–568.

EPSRC (2007), 'e-Science becomes business as usual', *HIGHLIGHTS from the UK e-Science Programme* (2), 5.

Ernst, R. E. & Buchan, K. L. (2002), 'Maximum size and distribution in time and space of mantle plumes: evidence from large igneous provinces', *J. Geodyn.* **34**, 309–342.

Fei, Y., Van Orman, J., Li, J., van Westrenen, W., Sanloup, C., Minarik, W., Hirose, K., Komabayashi, T., Walter, M. & Funakoshi, K. (2004), 'Experimentally determined postspinel transformation boundary in Mg_2SiO_4 using MgO as an internal pressure standard and its geophysical implications', *J. Geophys. Res.* **109**.

Flanagan, M. P. & Shearer, P. M. (1999), 'A map of topography on the 410-km discontinuity from PP precursors', *Geophys. Res. Lett.* **26**, 549–552.

Forte, A. M., Mitrovica, J. X. & Espesset, A. (2002), 'Geodynamic and seismic constraints on the thermochemical structure and dynamics of convection in the deep mantle', *Philos. T. Roy. Soc. A* **360**, 2521–2543.

Frimmel, H. E. (2008), 'Earth's continental crustal gold endowment', *Earth Planet. Sci. Lett.* **267**, 45–55.

Fukao, Y., Obayashi, M., Inoue, H. & Nenbai, M. (1992), 'Subducting slabs stagnant in the mantle transition zone', *J. Geophys. Res.* **97**, 4809–4822.

Giannandrea, E. & Christensen, U. R. (1993), 'Variable viscosity convection experiments with a stress-free upper boundary and implications for the heat transport in the Earth's mantle', *Phy. Earth Planet. In.* **78**, 139–152.

Goncharov, A. F., Struzhkin, V. V. & Jacobsen, S. D. (2006), 'Reduced radiative conductivity of low-spin (Mg,Fe)O in the lower mantle', *Science* **312**, 1205–1208.

Gong, L. (2001), 'JXTA: a network programming environment', *IEEE Internet Comput.* **5**, 88–95.

Graham, D. W., Blichert-Toft, J., Russo, C. J., Rubin, K. H. & Alberède, F. (2006), 'Cryptic striations in the upper mantle revealed by hafnium isotopes in southeast indian ridge basalts', *Nature* **440**, 199–202.

Gurnis, M. (1989), 'A reassessment of the heat transport by variable viscosity convection with plates and lids', *Geophys. Res. Lett.* **16**, 179–182.

Hansen, U. & Ebel, A. (1984), 'Experiments with a numerical model related to mantle convection: boundary layer behaviour of small- and large scale flows', *Phy. Earth Planet. In.* **36**, 374–390.

Heath, A. (2002), 'MantleVis'.
URL: *http://pcwww.liv.ac.uk/~aeh%20/Software/MantleVis.htm*

Hofmeister, A. M. (1999), 'Mantle values of thermal conductivity and the geotherm from phonon lifetimes', *Science* **283**, 1699–1706.

Honda, S. (1995), 'A simple parameterized model of Earths thermal history with the transition from layered to whole mantle convection', *Earth Planet. Sci. Lett.* **131**, 357–369.

Ito, E., Akaogi, M., Topor, L. & Navrotsky, A. (1990), 'Negative pressure-temperature slopes for reactions forming mgsio3 perovskite from calorimetry', *Science* **249**(4974), 1275–1278.

Ito, E. & Takahashi, E. (1989), 'Postspinel transformations in the system mg2sio4-fe2sio4 and some geophysical implications', *Journal of Geophysical Research-Solid Earth and Planets* **94**(B8), 10637–10646.

Iwase, Y. & Honda, S. (1997), 'An interpretation of the Nusselt-Rayleigh number relationship for convection in a spherical shell', *Geophys. J. Int.* **130**, 801–804.

Johari, G. P. & O, A. (2007), 'Vibrational and relaxational properties of crystalline and amorphous ices', *Thermochim. Acta* **461**, 14–43.

Kalas, P., Graham, J. R., Chiang, E., Fitzgerald, M. P., Clampin, M., Kite, E. S., Stapelfeldt, K., Marois, C. & Kirst, J. (2008), 'Optical Images of an exosolar planet 25 light-years from Earth', *Science Express* . doi: 10.1126/science.1166609.

Katsura, T., Yamada, H., Shinmei, T., Kubo, A., Ono, S., Kanzaki, M., Yoneda, A., Walter, M. J., Ito, E., Urakawa, S., Funakoshi, K. & Utsumi, W. (2003), 'Post-spinel transition in Mg_2SiO_4 determined by high P-T in situ X-ray diffractometry', *Phy. Earth Planet. In.* **136**, 11–24.

Kellogg, L. H., Hager, B. H. & Van der Hilst, R. (1999), 'Compositional stratifcation in the deep mantle', *Science* **283**, 1881–1884.

King, S. D., Raefsky, A. & Hager, B. H. (1990), 'ConMan: vectorizing a finite element code for incompressible two-dimensional convection in the Earths mantle', *Phys. Earth Planet. Int.* **59**, 195–207.

Korenaga, J. (2003), 'Energetics of mantle convection and the fate of fossil heat', *Geophys. Res. Lett.* **30**, 1437.

Korenaga, J. (2005), 'Archean geodynamics and the thermal evolution of Earth', *AGU Mono., Archean Geodynamic Processes* .

Korenaga, J. (2008), 'Urey ratio and the structure and evolution of Earth's mantle', *Rev. Geophys.* **46**. doi:10.1029/2007RG000241.

Korenaga, J. & Jordan, T. H. (2002), 'Onset of convection with temperature and depth-dependent viscosity', *Geophys. Res. Lett.* **29**, 1923.

Kumar, A., Heaman, L. M. & Manikyamba, C. (2007), 'Mesoproterozoic kimberlites in south India: A possible link to ~1.1 Ga global magmatism', *Precambrian Res.* **154**, 192–204.

Kurz, M. D., Jenkins, W. J. & Hart, S. R. (1982), 'Helium isotope systematics of oceanic islands and mantle heterogeneities', *Nature* **297**, 43–47.

Labrosse, S. (2002), 'Hotspots, mantle plumes and core heat loss', *Earth Planet. Sci. Lett.* **199**, 147–156.

Labrosse, S., Hernlund, J. W. & Coltice, N. (2007), 'A crystallizing dense magma ocean at the base of the Earths mantle', *Nature* **450**, 866–869.

Labrosse, S. & Jaupart, C. (2007), 'Thermal evolution of the Earth: Secular changes and fluctuations of plate characteristics', *Earth Planet. Sci. Lett.* **260**, 465–481.

Lowman, J., King, S. D. & Gable, C. W. (2004), 'Steady plumes in viscously stratified, vigorously convecting, three-dimensional numerical mantle convection models with mobile plates', *Geochem. Geophys. Geosys.* **5**(1).

Machetel, P. & Humler, E. (2003), 'High mantle temperature during cretaceous avalanche', *Earth Planet. Sci. Lett.* **208**, 125–133.

Machetel, P., Thoraval, C. & Brunet, D. (1995), 'Spectral and geophysical consequences of 3-D spherical mantle convection with an endothermic phase change at the 670 km discontinuity', *Phy. Earth Planet. In.* **88**, 43–51.

Machetel, P. & Weber, P. (1991), 'Intermittent layered convection in a model mantle with an endothermic phase-change at 670 km', *Nature* **350**(6313), 55–57.

Marois, C., Macintosh, B., Barman, T., Zuckerman, B., Song, I., Patience, J., Lafrenière, D. & Doyon, R. (2008), 'Direct imaging of multiple planets orbiting the star HR 8799', *Science Express* . doi: 10.1126/science.1166585.

Mckenzie, D. P., Roberts, J. M. & Weiss, N. O. (1974), 'Convection in earths mantle - towards a numerical-simulation', *Journal of Fluid Mechanics* **62**(Feb11), 465–538.

Mckenzie, D. & Weiss, N. (1975), 'Speculations on thermal and tectonic history of earth', *Geophys.J. Roy. Astron. Soc.* **42**(1), 131–174.

Montelli, R., Nolet, G., Dahlen, F. A. & Masters, G. (2006), 'A catalogue of deep mantle plumes: New results from finite-frequency tomography', *Geochem. Geophys. Geosys* **7**(11).

Montelli, R., Nolet, G., Dahlen, F. A., Masters, G., Engdahl, E. R. & Hung, S.-H. (2004), 'Finite frequency tomography reveals a variety of plumes in the mantle', *Science* **303**, 338–343.

Moresi, L. N. & Solomatov, V. S. (1995), 'Numerical investigation of 2d convection with extremely large viscosity variations', *Physics of Fluids* **7**, 2154–2162.

Nakagawa, T. & Tackley, P. J. (2004), 'Effects of thermo-chemical mantle convection on the thermal evolution of the Earths core', *Earth Planet. Sci. Lett.* **220**, 107–119.

Nataf, H. C. (2000), 'Seismic imaging of mantle plumes', *Annu. Rev. Earth Planet. Sci.* **28**, 391–417.

Nature, Editorial (2008), 'Community cleverness required', *Nature* **455**, 1.

Nelson, S. (2008), 'The Harvard computers', *Nature* **455**, 36–37.

Nimmo, F., Price, G. D., Brodholt, J. & Fubbins, D. (2004), 'The influence of potassium on core and geodynamo evolution', *Geophys. J. Int.* **156**, 363–376.

Oldham, D. & Davies, J. H. (2004), 'Numerical investigation of layered convection in a three-dimensional shell with application to planetary mantles', *Geochem. Geophys. Geosys.* (12), doi:10.1029/2003GC000603.

Oldham, D. N. (2004), On the possibility of layered mantle convection; numerical simulations in a spherical geometry, PhD thesis, Cardiff University.

Olsen, P. (1987), 'A comparison of heat transfer laws for mantle convection at very high Rayleigh numbers', *Phy. Earth Planet. In.* **48**, 153–160.

O'Nions, R. K. & Oxburgh, E. R. (1983), 'Heat and helium in the Earth', *Nature* **306**, 429–431.

Parman, S. W. (2007), 'Helium isotopic evidence for episodic mantle melting and crustal growth', *Nature* **446**, 900–903.

Parman, S. W., Kurz, M. D., Hart, S. R. & Grove, T. L. (2005), 'Helium solubility in olivine and implications for high $^3He/^4He$ in ocean island basalts', *Nature* **437**, 1140–1143.

Pearson, D. G., Parman, S. W. & Nowell, G. M. (2007), 'A link between large mantle melting events and continent growth seen in osmium isotopes', *Nature* **449**, 202–205.

Peltier, W. R. (1996), 'Phase-transition modulated mixing in the mantle of the Earth', *Philos. T. Roy. Soc. A.* **354**, 1425–1447.

Pollack, H. N., Hurter, S. J. & Johnson, J. R. (1993), 'Heat flow from the earth's interior: Analysis of the global data set', *Reviews of Geophysics* **31**, 267–280.

Prokoph, A., Ernst, R. E. & Buchan, K. L. (2004), 'Time-series analysis of large igneous provinces: 3500 Ma to present', *J. Geology* **112**, 1–22.

Rabin, M. O. (1989), 'Efficient dispersal of information for security, load balancing, and fault tolerance', *J. ACM* **36**(2), 335–348.

Rajasekar, A. K. & Wan, M. (2002), SRB & SrbRack : Components of a Virtual Data Grid Architecture, *in* 'Advanced Simulation Technologies Conference (ASTC02)', San Diego.

Rajasekar, A., Wan, M. & Moore, R. (2002), MySRB & SRB - Components of a Data Grid, *in* 'The 11th International Symposium on High Performance Distributed Computing (HPDC-11)', Edinburgh.

Ramaswamy, L., Gedik, B. & Liu, L. (2005), 'A distributed approach to node clustering in decentralized peer-to-peer networks', *IEEE Transactions on Parallel and Distributed Systems (TPDS)* **16**(10), 814–829.

Ratcliff, J. T., Schubert, G. & Zebib, A. (1996), 'Steady tertrahedral and cubic patterns of spherical shell convection with temperature dependent viscosity', *J. Geophys. Res.* **101**, 25473–25484.

Reese, C. C., Solomatov, V. S. & Baumgardener, J. R. (2005), 'Scaling laws for time-dependent stagnant lid convection in a spherical shell', *Phys Earth and Planet In* **149**, 361–370.

Reese, C. C., Solomatov, V. S., Baumgardner, J. R. & Yang, W. S. (1999), 'Stagnant lid convection in a spherical shell', *Phy. Earth Planet. In.* **116**, 1–7.

Rhea, S., Wells, C., Eaton, P., Geels, D., Zhao, B., Weatherspoon, H. & Kubiatowicz, J. (2001), 'Maintenance-free global data storage', *Internet Computing, IEEE* pp. 40–49. 1089-7801.

Richter, F. M., Nataf, H. C. & Daly, S. F. (1983), 'Heat-transfer and horizontally averaged temperature of convection with large viscosity variations', *J. Fluid Mech.* **129**(Apr), 173–192.

Ringwood, A. E. (1994), 'Role of the transition zone and 660 km discontinuity in mantle dynamics', *Phy. Earth Planet. In.* **86**, 5–24.

Rino, S., Komiya, T., Windley, B. F., Katayama, I., Motoki, A. & Hirata, T. (2004), 'Major episodic increases of continental crustal growth determined from zircon ages of river sands; implications for mantle overturns in the early Precambrian', *Phys. Earth Planet. In.* **146**, 369–394.

Ripeanu, M. (2001), 'Peer-to-peer architecture case study: Gnutella network'. **URL:** *citeseer.ist.psu.edu/ripeanu01peertopeer.html*

Sanz, E., Vega, C., Abascal, J. L. F. & MacDowell, L. G. (2004), 'Phase diagram of water from computer simulation', *Phys. Rev. Lett.* **92**(25).

Schubert, G., Cassen, P. & Young, R. E. (1979), 'Subsolidus convective cooling histories of terrestrial planets', *Icarus* **38**, 192–211.

Sharpe, H. N. & Peltier, W. R. (1978), 'Parameterized mantle convection and the Earth's thermal history', *Geophys. Res. Lett.* **5**, 737–740.

Shearer, P. M. & Masters, G. M. (1992), 'Global mapping of tomography on the 660-km discontinuity', *Nature* **355**, 791–796.

Simmons, N. A., Forte, A. M. & Grand, S. P. (2006), 'Constraining mantle flow with seismic and geodynamic data: A joint approach', *Earth Planet. Sci. Lett.* **246**, 109–124.

Simms, S. G., Pike, G. G. & Balog, D. (2007), Wide area filesystem performance using Lustre on the TeraGrid, *in* 'TeraGrid 2007 Conference,', Madison, WI.

Solomatov, V. S. (1995), 'Scaling of temperature- and stress-dependent viscosity convection', *Phys. Fluids* **7**, 266–274.

Stegman, D. R., Jellinek, A. M., Zatman, S. A., Baumgardener, J. R. & Richards, M. A. (2003), 'An early lunar core dynamo driven by thermochemical mantle convection', *Nature* **421**, 143–146.

Stein, M. & Hofmann, A. W. (1994), 'Mantle plumes and episodic crustal growth', *Nature* **372**(6501), 63–68.

Steinbach, V., Yuen, D. A. & Zhao, W. (1993), 'Instabilities from phase transitions and the timescales of mantle thermal convection', *Geophys. Res. Let.* **20**(12), 1119–1122.

Steinberger, B. (2007), 'Effects of latent heat release at phase boundaries on flow in the earths mantle, phase boundary topography and dynamic topography at the earths surface', *Phys. Earth Planet. In.* **164**, 2–20.

Stevenson, D. J. (2003), 'Styles of mantle convection and their influence on planetary evolution', *Geodynamics* **335**, 99–111.

Stevenson, D. J., Spohn, T. & Schubert, G. (1983), 'Magnetism and thermal evolution of the terrestrial planets', *Icarus* **54**, 466–489.

Stoica, I., Morris, R., Karger, D., Kaashoek, M. F. & Balakrishnan, H. (2001), Chord: A scalable peer-to-peer lookup service for internet applications, *in* 'Technical Report TR-819', MIT.

Tackley, P. J. (2000), 'Mantle convection and plate tectonics: Toward an integrated physical and chemical theory', *Science* **288**, 2002–207.

Tackley, P. J. (2008), 'Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid', *Phy. Earth Planet. In.* **171**, 7–18.

Tackley, P. J., Stevenson, D. J., Glatzmaier, G. A. & Schubert, G. (1993), 'Effects of an endothermic phase change at 670 km depth in a spherical model of convection in the Earth's mantle', *Nature* **361**, 699–704.

Tackley, P. J., Stevenson, D. J., Glatzmaier, G. A. & Schubert, G. (1994), 'Effects of multiple phase transitions in a three dimensional spherical model of convection in Earth's mantle', *J. Geophys. Res.* **99**, 15877–15901.

Tajima, F. & Nakagawa, T. (2006), 'Implications of seismic waveforms: Complex physical properties associated with stagnant slab', *Geophys. Res. Let.* **33**. doi:10.1029/2005GL024314.

Tauzin, B., Debayle, E. & Wittlinger, G. (2008), 'The mantle transition zone as seen by global Pds phases: No clear evidence for a thin transition zone beneath hotspots', *J. Geophys. Res.* **113**, doi:10.1029/2007JB005364.

Taylor, B. (2006), 'The single largest oceanic plateau: Ontong Java-Manihiki-Hikurangi', *Earth Planet. Sci. Lett.* **241**, 372–380.

Thain, D., Tannenbaum, T. & Livny, M. (2005), 'Distributed computing in practice: the Condor experience.', *Concurrency-pract. Ex.* **17**, 323–356.

Tozer, D. C. (1965), 'Physics of convection currents in earths mantle .21. heat transfer and convection currents', *Philosophical Transactions of the Royal Society of London Series a-Mathematical and Physical Sciences* **258**(1088), 252–.

Tseng, T.-L. & Chen, W.-P. (2004), 'Contrasts in seismic wave speeds and density across the 660-km discontinuity beneath the philippine and the japan seas', *J. Geophys. Res.* **109**. doi:10.1029/2003JB002613.

Turcotte, D. L. (1980), 'On the thermal evolution of the Earth', *Earth Planet. Sci. Lett.* **48**, 53–58.

Turcotte, D. L. & Oxburgh, E. R. (1967), 'Finite amplitude convective cells and continental drift', *J. Fluid Mech.* **28**, 29–42.

Van der Hilst, R., Engdahl, E. R., Spakman, W. & Nolet, G. (1991), 'Tomographic imaging of subducted lithosphere below northwest Pacific island arcs', *Nature* **353**, 37–43.

Van der Hilst, R., Widiyantoro, S. & Engdahl, E. R. (1997), 'Evidence for deep mantle circulation from global tomography', *Nature* **386**, 578–584.

Van Keken, P. E., Ballentine, C. J. & Porcelli, D. (2001), 'A dynamical investigation of the heat and helium imbalance', *Earth Planet. Sci. Lett.* **188**, 421–434.

Vazhkudai, S., Ma, X., Freeh, V., Strickland, J., Tammineedi, N. & Scott, S. (2005), FreeLoader: Scavenging desktop storage resources for scientific data, *in* 'Proceedings of Supercomputing 2005 (SC'05): Int'l Conference on High Performance Computing, Networking and Storage', Seattle, Washington.

Wang, I. (2005), P2PS (Peer-to-Peer Simplified), *in* 'Proceedings of 13th Annual Mardi Gras Conference - Frontiers of Grid Applications and Technologies', Louisiana State University, pp. 54–59.

Watson, E. B., Thomas, J. B. & Cherniak, J. (2007), '^{40}Ar retention in the terrestrial planets', *Nature* **449**, 299–304.

Weeraratne, D. & Manga, M. (1998), 'Transitions in the style of mantle convection at high Rayleigh numbers', *Earth Planet. Sci. Lett.* **160**, 563–568.

Wessel, P. & Smith, W. H. F. (1991), 'Free software helps map and display data', *EOS Trans. AGU* **72**, 441.

Wessel, P. & Smith, W. H. F. (1998), 'New, improved version of generic mapping tools released', *EOS Trans. AGU* **79**(47), 579.

Whitfield, J. (2008), 'An indifference to boundaries', *Nature* **451**, 872–873.

Wolstencroft, M., Rana, O. F. & Davies, J. H. (2006), Distributed storage of high-volume environmental simulation data: Mantle modelling, *in* 'IEEE/WIC/ACM International Conference on Web Intelligence', Vol. 0, IEEE Computer Society, Hong Kong, pp. 991–996.

Yamamoto, J. & Burnard, P. G. (2005), 'Solubility controlled noble gas fractionation during magmatic degassing: Implications for noble gas compositions of primary melts of OIB and MORB', *Geochim. Cosmochim. Ac.* **69**, 727–734.

Yang, W. S. & Baumgardner, J. R. (2000), 'Matrix-dependent transfer multigrid method for strongly variable viscosity infinite Prandtl number thermal convection', *Geophys. Astrophys. Fluid Dyn* **92**, 151–195.

Yang, X. & Veciana, G. d. (2004), Service capacity of peer to peer networks, *in* 'INFOCOM 2004', Hong Kong.

Ye, W., Khan, A. I. & Kendall, E. A. (2003), Distributed network file storage for a serverless (P2P) network, *in* 'The 11th IEEE International Conference on Networks, ICON2003', pp. 343–347.

Zhong, S. (2005), 'Dynamics of thermal plumes in three-dimensional isoviscous thermal convection', *Geophys. J. Int.* **162**, 289–300.

Zhong, S. J., Zuber, M. T., Moresi, L. & Gurnis, M. (2000), 'Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection', *J. Geophys. Res.* **105**, 11063–11082.

Zhong, S., Zhang, N., Li, X. L. & Roberts, J. (2007), 'Supercontinent cycles, true polar wander, and very long-wavelength mantle convection', *Earth Planet. Sci. Lett.* **261**, 551–564.

Appendix A

MantleStor

This Appendix can be found on the accompanying CDROM and contains source code for MantleStor along with the application itself. To access the CDROM menu, open the 'Appendices' file in a web browser. Source code is presented as a PDF and in .java files. MantleStor is provided as a MantleStor.jar file, in a directory with packages upon which it depends. Windows batch files are provided to launch the application.

On systems other than Windows, cd to the MantleStor directory and type:

```
java -jar MantleStor.jar client
```

or

```
java -jar MantleStor.jar peer
```

depending on if you wish to run a storage peer or the user client.

Appendix B

Mantle Avalanche Visualisation

This Appendix can also be found on the accompanying CDROM. To access the CDROM menu, open the 'Appendices' file in a web browser. The appendix consists of an animation of a mantle avalanche and subsequent plumes. The animation was produced using a MantleVis visualisation of Case 404 (Chapter 4), which had the following characteristics: Rayleigh Number 6.77×10^6, 660 km Clapeyron slope -8 MPa/K, mt256 resolution $\approx$22 km.