



THE UNIVERSITY  
OF ADELAIDE  
AUSTRALIA

# **Investigation into High Performance Computing Technologies for Geophysics**

Tristan M. Wurst,

B.Sc

Geology and Geophysics  
School of Earth and Environmental  
Sciences The University of Adelaide

October 2010

# Contents

Contents .....	ii
List of Figures .....	iii
Abstract .....	v
Chapter 1: Introduction .....	1
1.1 Overview .....	1
Chapter 2: Background .....	4
2.1 A New Paradigm .....	4
2.2 Parallelisation .....	5
2.3 Cloud Computing .....	6
2.4 Workflow for E-science .....	8
2.5 The Trident Workbench .....	12
2.6 MT Concepts .....	13
2.7 MT Processing .....	15
Chapter 3: Approach .....	18
3.1 Approach .....	18
3.2 Applications of Parallel Processing to MT Data .....	18
Chapter 4: Discussion .....	25
Chapter 5: Conclusion .....	31
References .....	33
Tables and Figures .....	39

## List of Figures

Figure 2.1: Sequential Execution of Independent Processes at Different Times .....	39
Figure 2.2: Parallel Execution of Different Processes at the Same Time .....	39
Figure 2.3: Graph Showing the Increase in the Number of Transistors on Processors with Time .....	39
Figure 2.4: Relationship Showing the Power Ceiling Wall with Increasing Number of Transistors on Processors (Mudge, C. personal communication, 27 September 2010)..	40
Figure 2.5: Depiction of a Data Centre in which the Servers are Located in Shipping Containers (Mudge, C. personal communication, 27 September 2010) .....	40
Figure 2.6: A Workflow Utilising the Trident Workflow Workbench Showing the Nodes Called Actors and the Flow of Data between Them .....	41
Figure 2.7: The Proposed Plate Scale Observatory on the Juan De Fuca Plate (Barga et al. 2008).....	41
Figure 2.8: Comparisons between Frequency and the Properties of Different EM Induction Techniques (Heinson, G. Personal communication, July 12, 2010).....	42
Figure 2.9: General Flowchart of the Processing Methodology .....	43
Figure 3.1: Depiction of the Serial Processing of the Time Series Data from the Field to the EDI Target Format .....	44
Figure 3.2: Depiction of a Shared Nothing Parallel Processing Framework of the Sites. ....	44
Figure 3.3: Overview of the Different MT Processing Methodologies of the Three Different Processors .....	45
Figure 3.4: Detailed Overview of Processor One’s MT Processing Methodology.....	46
Figure 3.5: Sequential Execution of the MT Processing Using Existing Processing Codes.....	47
Figure 3.6: Parallel Execution of the MT Processing Using Existing Processing Codes .....	47
Figure 4.1: An Example of Good MT Data for Orthogonal Components of the Electric and Magnetic Fields Sampled at 500Hz.....	48

Figure 4.2: Power Spectra of the Time Series for the By Component Shown in Figure 16.....	48
Figure 4.3: Power Spectra of the Time Series for the Ex Component Shown in Figure 4.1.....	49
Figure 4.4: Power Spectra Map of the Time by Component Showing the Change of the Power Spectrum with Time.....	49
Figure 4.5: Power Spectra Map of the Ex Component Showing the Change of the Power Spectrum with Time.....	50
Figure 4.6: An Example of Time Series Data Affected by 50Hz Electrical Noise.....	50
Figure 4.7: Power Spectra Map of the By Component Showing the Change of the Power Spectrum with Time.....	51
Figure 4.8: Power Spectra Map of the Ex Component Showing the Change of the Power Spectrum with Time.....	51
Figure 4.9: An Example of a Time Series Affected by Intermittent Transmitter Noise.....	52
Figure 4.10: Power Spectra Map of the By Component Affected by the Transmitter Showing Both Stationary and Nonstationary Noise.....	52
Figure 4.11: Power Spectra Map of the Ex Component Affected by the Transmitter Showing Stationary Noise.....	53
Figure 4.12: A Coherence Plot Characteristic of Good MT Data.....	53
Figure 4.13: A Coherence Plot Characteristic of Poor MT Data.....	54
Figure 4.14: An Overview Flowchart of how the WALDIM Dimensionality Analysis Program Works.....	55

## Abstract

The processing of magnetotelluric (MT) data is typically carried out on a desktop computer and as a result suffers from a number of drawbacks. The time taken to process the data on the desktop computer is unacceptably long and can take approximately a month. The limited amount of random-access memory (RAM) in the desktop computer limits the length of the time series that can be used in the bounded influence remote referencing processing (BIRRP) program. Cloud computing is a new high performance computing (HPC) technology that can be accessed over the internet and has the potential to address the drawbacks presented by the desktop computer. Cloud computing reduces the cost of HPC by pooling computing resources on a large scale. Cloud computing offers on-demand resources allowing the user to use only what they need and to change the type of resources they require to suit an evolving need. To utilise the HPC capabilities of the cloud, a problem must exhibit a high degree of parallelisation. MT processing is particularly well suited to cloud computing because of its inherent ability to parallelise by the number of stations. To enable automatic utilisation of the cloud resources, workflow technology can be used in conjunction with the existing MT processing codes. This new approach to MT processing presents the opportunity to address other inefficiencies in the processing. As the cloud is accessible over the internet, this presents the opportunity to perform some processing in the field. The ability to process data in the field is advantageous because it allows for near instant feedback about the quality of the obtained data. This feedback can then be used by the survey team to change the survey to optimise the quality of the obtained data if required. However, to achieve this, a number of new processing techniques need to be introduced into the workflow.