Performance Models for Electronic Structure Methods on Modern Computer Architectures

Joseph Antony

A thesis submitted for the degree of Doctor of Philosophy at The Australian National University

May 2009
Acknowledgements

I am indebted to my supervisor Alistair Rendell for his encouragement and support all throughout my candidature. My thanks also goes to Peter Strazdins and Andrey Bliznyuk, my co-supervisors. A warm thanks goes out to Ben Evans and the NCI National Facility team both as a PhD candidate and as a staff member. I’d like to take this opportunity to thank Mike Frisch and team Gaussian in Wallingford CT for not only sponsoring the ARC linkage grant that led to this thesis, but for all their help, advise and encouragement over the years.

I’ve been enriched by my interactions with past and present students, staff at ANU Computer Science; I’d like to especially thank:
Rui Yang, Pete Perrapong Janes, Jin Wong, Warren Armstrong, Jie Cai, V. Ganesh, Josh Milthorpe, Yang Xi, Arrin Daley. A special mention and thanks goes to Andrew Over and Bill Clarke. Bill had informed me about the state-of-the-art in full machine simulation, C++ run-time intricacies and Roman-era archaeological digs in Syria. Discussions with Andrew often revolved around SG1 and the UltraSPARC IIICu while Armin van Buuren was in ‘A State of Trance’. Joint work with Pete Janes enabled the development of the LPM model in Chapter 3, often sketched out on Saturdays after Pete’s fishing trips from Lake Burley Griffin. Rui Yang deserves an special mention for all his help throughout my thesis years, both as a joint collaborator (Chapter 5) and for his feedback and input often over coffee in his office.

A big vote of thanks goes to:
SGI (Todd Churchwood, Anthony David, Roberto Gomperts);
IBM (Hugh Blemmings, James Kelly);
IBM academic program (for access to compilers and BLAS libraries);
Apple ADC (Tess Collins);
ANU DCS (Pascal Vuylsteker, Steve Blackburne, Eric McCreath, Bob Edwards, James Fellows, Steve Hanley, Hugh Fisher, Robin Garner, Chris Johnson, Henry Gardner);
ANUSF (Jason Ozolins, Rika Kobayashi, Robert Davy, Jonathan McCabe, David Singleton, Judy Jenkinson, Margaret Kahn, Robin Humble, Bob Gingold);
ANU DoI (Allan Williams, Markus Buuchorn, Andrew Wellington, Doug Moncur);
Cray Australia (Lindsay Hood);
Stuart Watson and Richard Walker.
On a personal note: a special thanks to ‘F & M Inc.’

This work was made possible as a result of funding from the Australian Research Council in conjunction with Gaussian Inc. and Sun Microsystems Inc. under ARC Linkage Grant LP0347178.
Abstract

Electronic structure codes are computationally intensive scientific applications used to probe and elucidate chemical processes at an atomic level. Maximizing the performance of these applications on any given hardware platform is vital in order to facilitate larger and more accurate computations. An important part of this endeavor is the development of protocols for measuring performance, and models to describe that performance as a function of system architecture. This thesis makes contributions in both areas, with a focus on shared memory parallel computer architectures and the Gaussian electronic structure code.

Shared memory parallel computer systems are increasingly important as hardware manufacturers are unable to extract performance improvements by increasing clock frequencies. Instead the emphasis is on using multi-core processors to provide higher performance. These processor chips generally have complex cache hierarchies, and may be coupled together in multi-socket systems which exhibit highly non-uniform memory access (NUMA) characteristics. This work seeks to understand how cache characteristics and memory/thread placement affects the performance of electronic structure codes, and to develop performance models that can be used to describe and predict code performance by accounting for these effects.

A protocol for performing memory and thread placement experiments on NUMA systems is presented and its implementation under both the Solaris and Linux operating systems is discussed. A placement distribution model is proposed and subsequently used to guide both memory/thread placement experiments and as an aid in the analysis of results obtained from experiments.

In order to describe single threaded performance as a function of cache blocking a simple linear performance model is investigated for use when computing the electron repulsion integrals that lie at the heart of virtually all electronic structure methods. A parametric cache variation study is performed. This is achieved by combining parameters obtained for the linear performance model on existing hardware, with instruction and cache miss counts obtained by simulation, and predictions are made of performance as a function of cache architecture.

Extension of the linear performance model to describe multi-threaded performance on
complex NUMA architectures is discussed and investigated experimentally. Use of dynamic page migration to improve locality is also considered.

Finally the use of large scale electronic structure calculations is demonstrated in a series of calculations aiming to study the charge distribution for a single positive ion solvated within a shell of water molecules of increasing size.