Grid research in Ireland

Brian Coghlan,
Department of Computer Science,
Trinity College Dublin,
Ireland.

Andy Shearer,
Information Technology Centre,
National University of Ireland, Galway,
Ireland.

Michael Manzke,
Department of Computer Science,
Trinity College Dublin,
Ireland.

John Morrison,
Department of Computer Science,
University College Cork,
Ireland.

Abstract ***Grid-Ireland has been established to link compute clusters at three partner sites in Ireland: the Departments of Computer Science in Trinity College Dublin, University College Cork and NUI Galway. Expansion to other academic institutions is in preparation. Here we outline the research themes being explored in the early life of this national grid***

Keywords: Grid, Execution Models, Condensed Graphs, Algorithms

1 Introduction

In the future, as many people have said, computing will become a ubiquitous strategic resource, which users may avail of any time, anywhere. The subject is in its very earliest stage at present, with a very great number of basic research and practical issues to be investigated. Initial seed funding has been provided by Enterprise Ireland to guarantee the establishment of a working national grid between compute clusters at three partner sites in Ireland, that is, the Departments of Computer Science in Trinity College Dublin, University College Cork and NUI Galway.

This project has one primary objective: to establish *Grid-Ireland*. This requires:

- Hardware: Compaq will donate a 4-way SMP gateway machine per site.
- Software: initially the Globus services.
- Management: the Dept.Computer Science at Trinity College will do this.
- Interconnect: the Irish academic network services will be used.

A secondary objective is to begin cooperative research work between the three sites. The themes are clear:

Understanding: the most basic objective is the understanding of system state and estimation of future state of individual nodes, complete clusters, and heterogeneous compositions of these to form a computational grid. This will be explored using control theory via the analagous problem of real-time grid input-output, which is a very difficult concept due to the non-determinism of the grid.

Evolution: traditional models of grid computation (control-driven models) encourage programming practises that are polarised between the shared memory model common to symmetric multiprocessors and the message passing model common to ensembles of independent nodes. There is, however, a middle ground where lateral thinking and systematic exploitation may bear fruit; this hybrid model is essentially evolutionary.

These two themes conserve the existing modus operandus; what if a more radical approach is appropriate? What if grids were better served by a totally different execution models and substantially different algorithms than heretofore?

Revolution: in contrast to shared-memory or message-passing models, or some hybrid of these, data-flow and demand-driven execution models take a revolutionary approach, viewing an execution as a set of computational nodes, connected by paths along which data values move in a manner determined by the topology of the graph and by the rules underlying that model. Their superset, condensed graphs, offers a generalisation of both data-flow and demand-driven configurations.

Algorithms: the grid is different, ergo the algorithms must be different? We will explore suitable applications algorithms for the grid, particularly the emerging techniques for adaptive grid simulations. The programme will involve both the development of parallel Monte-Carlo algorithms and testing of these in a clinical environment both locally and nationally.

The grid is very interesting in terms of behaviour because it is extreme - tightly coupled processors will be interconnected via a high latency low bandwidth network. Moreover it is diverse. A great deal of basic research is needed into the conceptual models for almost every aspect of the grid environment, and these efforts will need to be driven with hard applications in diverse fields, more than is typical at present.

2 Understanding

A major aim of those involved in grid research is to enable applications to utilise resources in an optimal way, by recognizing current utilisation of the grid, by exploiting those resources that are under-utilised, and by avoiding or moving away from those that are over-utilised. This requires monitoring, analysis, and control mechanisms. Here we fail at the first hurdle - even current monitoring mechanisms are extremely minimal.

Moreover, there is little or no understanding of how the grid behaves. The computational grid is recognized to be a very dynamic system. Monitoring and analysis must be conducted on-line - offline approaches are not viable. Furthermore, attempting to control a poorly understood system is a recipe for catastrophe. There have been a number of famous failures of the power grid.

Figure 1: 16-node switched-SCI cluster at Trinity College Dublin.

The most essential thing is to understand the dynamics of the grid, with all its non-determinism. Here we propose to investigate concepts from process control theory. This very fledgling work is being conducted in the context of research activities related to the global state estimation and optimisation of compute-clusters. Measurement is crucial. This is a subject in which Trinity College Dublin has long had an interest, particularly at the instrument and analysis level [3].

3 Evolution

For traditional (control-driven) execution models, maximally-optimized message-passing provides the upper bound for efficiency. The optimizations may be complex and counter-intuitive - maximally optimizing a message-passing program requires significantly greater effort and expertise than for shared memory. The reasons for this are clear: message passing requires explicit data placement and communication, whereas for shared memory this is done

Figure 2: left: tracer/analyzer for the Scalable Coherent Interconnect (SCI), and right: in a heavily instrumented environment (see http://www.cs.tcd.ie/coghlan/scieuro/).

implicitly. The grid will exacerbate this situation. The motivation for shared memory is to eliminate the need to expend so much effort.

From the programmer's point of view, it is easier if at all times all processes have a consistent view of the contents of shared data structures. However, parallel programs require synchronisation for correct (race-free) execution. Many shared memory optimizations exploit this fact and weaken consistency. Generally the weaker the consistency, the less the message traffic, and on the grid this will assume greater importance. The unfortunate reality is that, despite this, message-passing is the most efficient solution, at least in critical places in a program [4]. In an attempt to overcome this, we propose to explore the realm between these extremes. There are a number of very interesting possibilities. This is very speculative research.

4 Revolution

Like both dataflow and demand-driven execution models, *condensed graphs* (CGs) [5] view an execution as a set of computational nodes, connected by paths along which data-values move, in a manner determined by the exact graph, and the rules underlying that model. CGs however, offer a generalisation of both data-flow and demand-driven configurations: each can be straightforwardly transformed into a CG equivalent, and furthermore CGs can be constructed which contain elements exhibiting both behaviours. The same model is inherently parallel, but can equally well express sequential, imperative-style computations.

An important property of CGs is their multilevel, hierarchical, structure. Replacing a cluster by a single node representing the same computation may condense structured graphs; this is analogous to a *fold* operation in a source-level program transformation. The dual operation, *evaporation*, is comparable to a program transformation *unfold*. By this device, it is possible to alter how much detail of a computation it is desired to expose at any given time, thereby making the representation more or less abstract.

Condensed Graphs can act equally well as both an abstract or physically realized machine, and as an intermediate representation: compiling the C intermediate code now requires, not an actual translation into a different representation, but rather is a matter of transforming the code into graphs which are less abstract, and which have more specificity to given hardware. Many of the techniques of source-to-source program transformation can be carried over into the CG world, allowing code optimisations to be expressed as *graph transformations*. Having done this we can then either execute the "intermediate" representation directly, using the Condensed Graph Abstract

Machine (CGAM) model; or it can be used as a true intermediate representation, compiling the CG code down to native code as a final step.

The Centre for Unified Computing in University College Cork is implementing CGs on a rich variety of platforms, and this work is already well under way. The grid offers a valuable opportunity to test the process of obtaining different machine-level CG realizations.

5 Algorithms

It is clear to the emerging grid community that the grid is so different that existing application-level algorithms are often quite unsuited. NUI Galway intends to explore and advance the algorithmic base via real medical applications. NUI Galway is currently developing algorithms for determining the passage of low energy optical photons through the body, and intends to apply them to the higher energy regime where different nuclear cross sections and scattering behaviour is expected.

These algorithms will be in the area of simulations for radiotherapy purposes [1]. Patient treatment planning is a crucial aspect of any course of radiotherapy. In essence this consists of some scans that determine the best machine configuration to give the optimum dose to the patient. The best method is via Monte-Carlo simulations where the track of a large number of particles is followed through the scattering material. Although analytical techniques have been used they are not as accurate. Monte-Carlo simulations, however, require the use of many particles to gauge accurately the degree of scatter and attenuation, see Figure 3. Such calculations can also help determine when a particular round of treatment has not given the tumour the desired dose. Figure 4 illustrates this problem.

NUI Galway intends to develop similar Monte-Carlo code but without the requirement for a large computer to be on-site within each hospital. This can be done by taking the code developed for optical imaging and photodynamic therapy purposes, and modifying the underlying physics for X-rays rather than photons. The results could be used for setting radiotherapy machines, and so are expected to be of interest to the major local hospitals employing radiotherapy at each of the grid sites, that is, in Dublin, Cork and Galway.

Finally we intend to investigate the emerging technique of adaptive mesh simulation [2]. In a typical analytical simulation a fixed grid is used to describe a physical phenomena and the forces at each node are calculated for a series of time steps. A better solution is to vary the grid spacing and structure dynamically through the calculation. The grid with its different architectures is an ideal platform for both calculating the node forces and the subsequent grid topology.

6 Summary

Grid-Ireland has been established to link compute clusters at three partner sites in Ireland. Plans are now in preparation for expansion to nine sites by the end of 2001. Four research themes are being explored in the early life of this national grid; this will rise to fifteen by the end of 2001.

References

[1] Ahnesjo, A. and Aspradakis, M.M., *Phys. Med. Biology*, pp.99–155, Vol.44, 1999.

[2] Lohner, et al, Parallel Unstructured Grid Generation, *Computer Methods in Applied Mechanics and Engineering*, pp.343, Vol.95, 1992.

[3] Manzke, M. and Coghlan, B., Non-Intrusive Deep Tracing of SCI Interconnect Traffic, *Proc.SCIEurope'99*, pp.53–58, Toulouse, September 1999.

[4] MacLaren, J.M. and Bull, J.M., Lessons Learned when Comparing Shared Memory and Message-Passing Codes on Three Modern Parallel Architectures, *Proc.Int.Conf.*

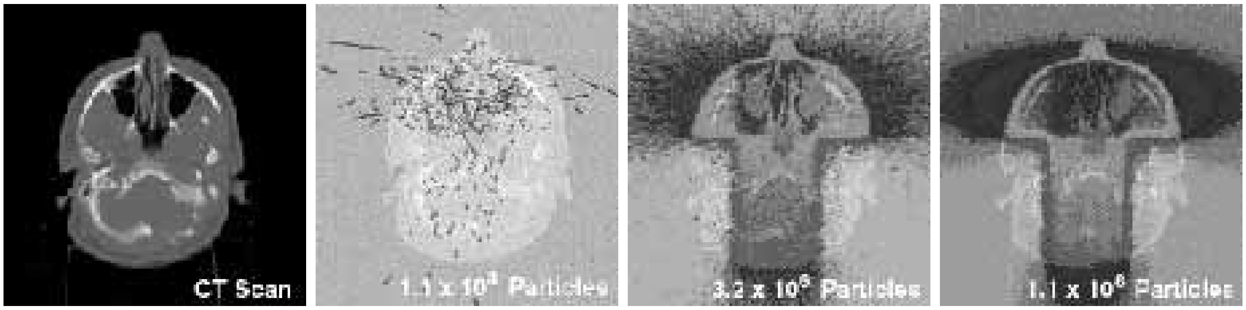

Figure 3: A scan followed by simulations involving 10,000 to 10 million particles [Lawrence Livermore Labs].

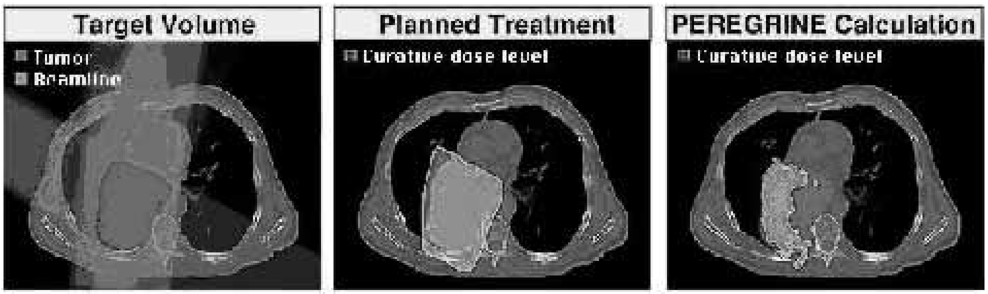

Figure 4: The tumour is shown in the first figure, the planned treatment in the second (analytical approach). In the third a Monte Carlo simulation shows that not all the tumour has been irradiated.

High-Performance Computing and Networking, pp.337–346, Amsterdam, April 1998.

[5] Morrison, J.P. and Dalton, N.J., Condensed Graphs: A Multi-Level, Parallel, Abstract Machine, *HPCS'99*, Queen's University, Kingston, Canada, June 13-16, 1999.