ALPS - The Swiss Grand Challenge Programme on the Cray XT3

Dominik Ulmer *and* **Neil Stringfellow**, *Swiss National Supercomputing Centre CSCS*

ABSTRACT: *With the installation of the first Cray XT3 in Europe in 2005, the Swiss National Supercomputing Centre CSCS made a strategic change towards large-scale massively-parallel computing. Consequently, CSCS started a new resource allocation scheme called "Advanced Large projects in Supercomputing (ALPS)" in 2006. Four projects targeting scientific breakthroughs by means of large-scale computing were accepted, covering molecular dynamics, climate modelling, and physics of the earth's magnetic field.*

KEYWORDS: CSCS, Cray XT3, MPP, grand challenges

1 Introduction

1.1 CSCS

The Swiss National Supercomputing Centre CSCS (Centro Svizzero di Calcolo Scientifico) was founded in 1991 to support the academic research institutions in Switzerland with HPC services. It is an autonomous administrative entity of ETH Zurich and operates under a four-year performance contract with global budget. CSCS is located in the Italian-speaking part of Switzerland on the Southern side of the Alps.

CSCS' supercomputer portfolio currently comprises a Cray XT3 MPP system with 1664 dual-core Opterons at 1GB memory per core and a 48 node IBM P5-575 Infiniband cluster, each node with 16 Power-5 CPU and 32 to 80 GB shared memory. Smaller systems include a single-rack Cray XT3 system with 74 dual-core Opterons for hosting the Swiss national weather forecasting suite on a 7km model raster and a 32 node HP visualisation cluster for parallel rendering and remote display.

As of 2007, the centre provides of a staff with a headcount of 35. Besides a small administrative overhead, human resources are split approximately evenly between technical staff for operating and maintaining the technical infrastructure, and scientific staff for user support, application portfolio management, porting, optimisation, and performance analysis as well as data management and visualisation.

1.2 Strategic re-orientation of the centre in 2004

From 1991 to 2003, the centre's technology strategy was to play an important national niche role by providing unique vector computing resources, represented by a succession of NEC SX installations, from generation SX-3 to SX-5. The last system, a NEC SX-5 with 16 CPU and 128 GB of shared memory in a single node, was operated from 2000 to 2007. In 2001, the vector compute facility line was augmented with an IBM SP4 cluster of 256 Power-4 CPU and a Colony interconnect.

The technology strategy described above was the consequence of specific boundary conditions of CSCS: Due to the size of economy of the country, the centre provides only of limited investment capacities. However, the scientific community, which is mainly represented by researchers at ETH Zurich and EPFL Lausanne, carries out world-class research with corresponding computing requirements. As a result, CSCS followed a low-risk strategy, concentrating on a specialised computer architecture, which was introduced late in its life cycle. Thus, a small, but dedicated user community segment could be established for the centre.

In 2004, a new governance model was implemented for CSCS, based on a performance contract, which states objectives of the centre for a period of four years. One of the main objectives of the performance agreement 2004-

2007 is to create a new technical base, which addresses a much wider spectrum of potential users and research fields and which enables them to carry out very large-scale simulations.

As the economic boundary conditions remained unchanged, this objective can only be achieved by shifting the technology strategy from specialised HPC technology in a mature state to leading-edge, general-market supercomputing technology in a very early stage of its life-cycle.

1.3 Introducing the Cray XT3

In 2004/2005, CSCS conducted an open public procurement for a massively parallel computer, codenamed “Horizon”. The project was a joint effort between CSCS and the Paul-Scherrer-Institute, a multi-disciplinary research centre for natural sciences and technology in Villigen, Switzerland, which decided to co-invest into this new HPC facility. The target was to procure a system of at least 1’000 processors, which would let scientific application scale to the full size of the system.

Based on the results of sustained performance - measured mainly by the HPC Challenge benchmark suite and scientific applications of PSI - and total cost of ownership, CSCS chose a 1’100 CPU Cray XT3 system for Horizon. The system was delivered in summer 2005 and was the one of the first Cray XT3 systems worldwide and the first of its kind in Europe. After an extensive acceptance period and several months of deployment and of access by pilot users, the system was opened for general usage in January 2006. Due to its big success and popularity in the user community, it was extended two times, first by adding additional 50% of cabinets in summer 2006 and then by a dual-core upgrade in April 2007, letting the machine grow from initially 5.6 Tflops peak performance to 17.2 Tflops.

In January 2007, the Swiss national weather service MeteoSwiss decided to move its forecast suite from the NEC SX-5 to a separate single-cabinet Cray XT3, thus finalising the technical re-orientation of the computing centre. Due to the success of this single-cabinet system for operational weather forecasting together with the high popularity of the XT3 systems among the remainder of CSCS user community, a five cabinet Cray XT4 system with 896 compute cores has been purchased, and this is due to arrive at the end of May 2007. This system will be used for the next generation high-resolution weather forecasting of MeteoSwiss as well as some applications with high memory bandwidth requirements which might not be able to take full advantage of the dual-core upgrade on the Cray XT3.

2 The concept of the ALPS Programme

One of the main objectives of the Horizon procurement was to provide a computing base for very large-scale simulations. Previously, almost all compute time on CSCS machines was from jobs of no more than 16 to 32 processors, with a significant share of single processor jobs.

Already when opening the new Cray XT3 system in January 2006 for general production, after a few months of pilot usage, the pattern had shifted significantly to larger jobs, mainly in the order of 64 to 127 CPU, four times bigger than the largest jobs on previous CSCS systems (Figure 1).

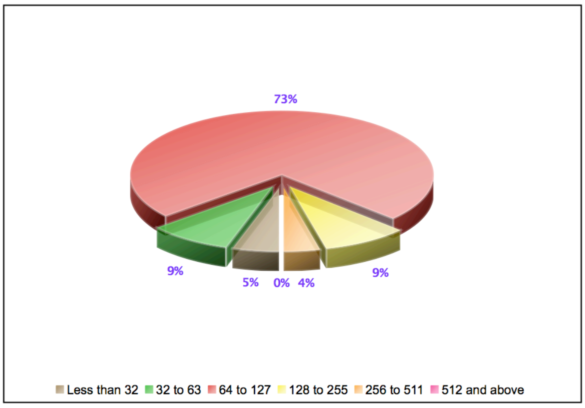

Figure 1: Usage by number of processors on the Cray XT3 in January 2006

Only 9 months later, the pattern had shifted again significantly to higher processor counts per job. In September 2006, 38% of the cycles of the machine were consumed by jobs between 256 and 1024 processors. As the higher number of large jobs opened more slots for small backfill jobs, the share of small jobs up to 63 CPU grew also from 14% to 19% of the machine (Figure 2). Interestingly, it could be observed that even on the old systems of CSCS, user started to run bigger jobs: In January 2006, 83% of the compute time of the IBM SP4 went to jobs of 16 CPU or less. Only 4 months after the arrival of the Cray XT3, in April 2006, this segment shrunk to 49% of the usage, indicating a major shift in the approach how scientists address their research problems on CSCS’ computers.

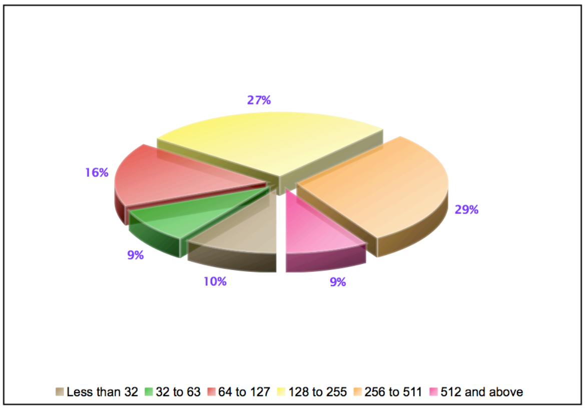

Figure 2: Usage by number of processors on the Cray XT3 in September 2006

Formally, the two main objectives of the performance contract of CSCS for 2004 to 2007 were already fulfilled in mid-2006: a new, leading-edge supercomputer architecture was introduced, which enables the researcher to address large-scale simulations. However, it turned out that CSCS was almost too successful: demand for compute time on the Cray XT3 in the annual call for proposals was very high, sometimes exceeding the available capacity by factor 20 before scientific review, and still by factor 5 after scientific evaluation of the proposal. Allocations for promising research projects, especially those that were targeting large-scale computational problems, had to be cut back significantly.

To enable and sustain this new level of scientific innovation, CSCS decided to introduce a new scheme for computational projects at CSCS, which complements the existing allocation procedure. The new scheme, called ALPS (Advanced Large Projects in Supercomputing) would specifically address projects targeting scientific breakthroughs by means of large-scale simulation. A guaranteed share of the compute capacity of CSCS was set aside for the new scheme. If a project would be accepted for ALPS, it would be fully funded for the whole life span of the project, with a maximum of three years. Therefore, the scheme must be highly selective and only the very best projects will be funded. Furthermore, it was required that the project has a collaborative component between CSCS and the research group, e.g. on the field of parallel software engineering or visualisation, in order to deepen the relationship between the centre and its major users. Finally, CSCS will assign a key account manager to each ALPS project who follows the project closely, can help to acquire the necessary prerequisites, and intervenes in the case that the project is not on track.

3 ALPS projects accepted in 2006

Initially, approximately 8.5 Million CPU hours of compute time on the Cray XT3 were set aside for the ALPS programme. This corresponds to the aggregate useable CPU hours over two years on the 564 processor extension of the Cray XT3, made in August 2006. Due to the huge success of the call, the quality of the proposals, and the arrival of the IBM P5 cluster as additional resource at CSCS, the actual allocation was 16 Million CPU hours over 2 years, or almost twice as much. The allocation was given to four projects, ranging from molecular-dynamics simulations in life science to planetary science. The principal investigators for all four projects are researchers at ETH Zurich.

3.1 Andrew Jackson: Convection and Magnetic Field Generation in Earth and Other Planets

This project, in collaboration with Gary Glatzmaier from the University of California Santa Cruz, received 3,500,000 CPU hours, an equivalent of 145,833 CPU hours per month, for the simulation of the fluid dynamics in the core of earth-type planets and for the first time also of gas giants, which lead to the generation of a magnetic field.

Aspects of the research include

- Modelling the dynamics where the Coriolis forces are large enough compared to the viscous forces that they resemble the situation in the Earth's core.
- Simulatation of thermal convection in gas giants and how this leads to the observed surface cloud features and zonal wind patterns.
- Carrying out the first simulations to account for the oblate geometries of the gas giants as opposed to simulating these planets as purely spherical objects.
- Performing calculations to gain insights into magnetic polarity reversals.

This project, which involves a number of home grown codes that can scale to 1'000 processors or more on the Cray XT3, poses particular challenges in the field of I/O. It is estimated that the project needs about 200 TB of offline or near-online storage. A single simulation suite will require about 9 to 41 TB of online storage. Because the data must be post-processed and analysed on other systems like the centre's visualisation cluster, CSCS has started a project for deploying a centre-wide parallel file system, to which all major systems of CSCS will have access.

Since the programs used in this project have been written by the project members, there is wide scope for making changes to optimise the scientific output of the applications. Code developments in this project have initially focussed on optimising the applications for the Cray XT3 platform and modifying the I/O strategies to allow increased scalability to over 1,000 processors. Further optimisations are still possible for these codes, but memory bandwidth constraints are becoming the critical factor and therefore these applications will be tested for their suitability on the new Cray XT4 system to be delivered to CSCS at the end of May 2007. A further possible development of one of the codes would be the introduction a multigrid approach although this would be a major collaborative effort going beyond the traditional optimisation work carried out at CSCS.

3.2 Michele Parrinello: Modelling Protein-Protein Interactions at the Atomic Level

Michele Parrinello, a former director of CSCS and co-author of the famous Carr-Parrinello-Method for molecular dynamics simulations, studies in this project the interaction of three different kinds of protein-protein interactions which control aspects of cellular life and are implicated in serious diseases.

The specific systems of interest to the medical and biophysical community that are to be studied are

- Amyloid fibrils, which aggregate in the brain of Alzheimer patients.
- Cyclin-dependent kinases, which are related to cell replication and therefore, to uncontrolled multiplication of cancer cells, as well as also being implicated in Alzheimer's disease.
- HIV Protease and the interaction with local elementary structures, in order to gain a better insight into methods of developing protease inhibitors - a key for developing drugs against HIV.

4,000,000 CPU hours or 166,667 CPU hours per month over 2 years have been allocated to this project. The main simulation applications are community codes, NAMD and ORAC. The simulation jobs in this project launch multiple instances of the codes, which communicate every few time steps with each other.

One of the code developments from CSCS personnel in this project will include the introduction of a robust, asynchronous parallelisation scheme for multiple instances of NAMD, increasing the centre's expertise in this code in synergy with the work to be carried out for project of Viola Vogel. For the ORAC code, improvements to the scalability of the code will be carried out such that it might be considered a useful tool for molecular dynamics simulations by a wider research community.

3.3 Christoph Schär: Climate Change and the Hydrological Cycle from Global to European/Alpine Scales

One of the major effects of global change in central and Eastern Europe is the higher annual variability between summer temperatures and the higher probability of strong precipitation events. Switzerland located with its alpine topography in the centre of Europe is a particularly affected by the consequences of these developments.

Christoph Schär and his group study in this ALPS project the role of climate change on the European summer climate and water cycle. They interleave a climate model, using the code ECHAM-HAM ported by Mark Cheeseman from CSCS to the Cray XT3 platform, and a weather model, based on CLM, to simulate the development of the alpine climate in the future.

The project aims to study the likely effects of climate change on the hydrological cycle of the European Alpine region by carrying out simulations for the period 1950-2050. The numerical experiments aim to gain new insights by taking advantage of three major advances

- Detailed inclusion of aerosol processes in the climate simulations
- The use of very-high resolution climate change experiments (down to 2km)
- The combination of global and regional/local modelling frameworks

A total of 3,750,000 CPU hours over two years, or 156,250 CPU hours monthly, has been allocated to their research. Although not as extreme as the storage requirements of the group of Andrew Jackson, but still significant, are the I/O requirements: 185 TB of offline or near-online storage and 5 TB of online storage.

The Cray XT3 system has already proved to be a very good platform for the ECHAM code and the ECHAM-HAM derivative, and further optimisation work may be carried out to tune this application on the XT3 and XT4 machines. As the CLM is a derivative of the LokalModel used by MeteoSwiss for the operational weather forecasts, the optimisation work already carried by Cray and CSCS on the XT3 platform should be transferable to this code as well. Beyond the work already carried out, major efforts are continuing in the area of I/O optimisation and the possible inclusion of the parallel NetCDF library as the main method used for generating data.

3.4 Viola Vogel: Towards Simulating a Cell Adhesion Site at Ångström Resolution

Viola Vogel and her team received the largest allocation of 4,800,000 CPU hours over 2 years (200,000 CPU hours per month) to investigate the mechanisms by which force is transmitted in force-bearing proteins and how force might change the structure and function of proteins at cell surfaces. Previous studies of molecular adhesion have been carried out in collaboration with

Klaus Schulten from the University of Illinois at Urbana-Champaign, with Steered Molecular Dynamics (SMD) being a vital tool in their research.

Previously, the group has carried out research into the activation mechanisms of the bacterial adhesion molecule FimH, which is active at the tip of bacteria such as E. coli. The bacteria use this sticky protein at the tip of their filaments to adhere at surfaces. Not only does E. coli achieve astonishing forces with this mechanism, it also automatically regulates the force according the external forces which are applied: it turns out that the bacterium grips harder to the surface, the stronger water or other media pull at the bacterium. Understanding this mechanism is key for fighting infection paths how bacteria colonise tissues in the body.

The current project will study how force might change the structure and function of various proteins including integrins and how force is transmitted through protein structures. Previous research by the group has shown that the activation of an adhesion site can be triggered by forces applied at some distant part of the molecule. In order to account for these long-range effects, this project requires full molecules to be simulated as well as molecular fragments, and therefore some of these simulations will involve over a million atoms.

For her research, it is particularly important to compare the computer simulation results with laboratory experiments, and the research into FimH consisted of both experiments and computer simulations by members of the group. Investigations of the mechanisms involved in activating these adhesive molecules is greatly facilitated by scientific visualisation, and CSCS has therefore started to become involved in software development in NAMD and the associated visualisation package VMD.

4 Current Status

The demand for time on the Cray XT3 has been extremely high, and after allocating the time for the ALPS program the remaining CPU Hours were 30 times oversubscribed before scientific review. It was therefore decided to only distribute half of the requested monthly allocation to each ALPS project for the first 4 months, and then give a much larger allocation after the dual-core upgrade had taken place. This has not proved to be a limitation for most of the projects since it allowed for an extended period of preparation and planning before the main calculations begin.

As part of the upgrade to dual-core, a series of benchmarks of the major codes used at CSCS were performed by both application support personnel and members of the user community. These benchmarks covered codes from a representative sample of all research fields currently active on the CSCS Cray XT3 systems, and were used to set up a charging mechanism to account for dual-core usage when compared to the allocations that had been given out assuming single-core usage.

The benchmarks of molecular dynamics codes used on CSCS systems performed well in dual-core mode, with NAMD performing extremely well. Whilst all projects will be adjusted to ensure that they gain access to a level of resource which is equivalent to the number of single-core CPU Hours originally committed, the projects which make heavy use of NAMD will benefit from the fact that the dual-core performance of this code is better than the number of single-core equivalent CPU Hours which CSCS charges for a dual-core node. The codes used in the projects from Andrew Jackson and Christoph Schär will be further studied to determine whether they would be best served by running on the new Cray XT4 system which will arrive at CSCS at the end of May.

The ALPS scheme has given extra impetus to the project to introduce a site-wide parallel file system and to the reorganisation of the tape archive, particularly due to the large storage requirements coming from two of these projects, and these improvements to CSCS' infrastructure will be of benefit to the whole user community.

5 Conclusions

The installation of the Cray XT3 enabled CSCS to support large-scale numerical experiments. The establishment of the ALPS programme created a mechanism to select and support projects targeting breakthrough science based on these large-scale simulations. CSCS profits from these projects in several ways:

- The challenges faced with these projects will serve as input for continuing infrastructure improvements and future procurement decisions.
- Problem sizes ran from ALPS projects may be the future problem sizes tackled by many users from the other programme scheme at CSCS.
- Skill requirements originating from the ALPS projects may be of benefit in CSCS recruitment decisions.
- The ALPS projects give CSCS the opportunity to establish itself as a highly skilled HPC competence centre working closely with the scientists.

Code development taking place in the ALPS programme will benefit a wider community because of the direct return to community codes and due to the extra skills gained by CSCS staff.

Acknowledgments

The authors would like to thank colleagues and users as well as Cray staff who together made this success story happen.

About the Authors

Dominik Ulmer is Chief Operating Officer at the Swiss National Supercomputing Centre CSCS. Neil Stringfellow is a Senior Application Analyst and the ALPS Programme Manager at CSCS. They can be reached at: CSCS, Galleria 2, via Cantonale, 6928 Manno, Switzerland, E-mail addresses: dulmer@cscs.ch and nstring@cscs.ch.