NORDIC COMPUTATIONAL GRAND CHALLENGE SURVEY

April 16, 2007

EXECUTIVE SUMMARY

This report contains the results of a survey that was carried out in the autumn of 2006 among the Nordic research community. The primary aim of the survey was to identify 'Grand Challenge' scientific problems that are considered important to the research community and whose solutions require access to large scale computational or storage resources, possibly larger than any of the existing supercomputer facilities or larger than the aggregate capacity provided by any of the national infrastructures in the Nordic countries.

A total of 34 descriptions of Grand Challenge problems were received. The Grand Challenges come from a variety of disciplines, including earth sciences, biosciences, physics, fluid dynamics and chemistry. The descriptions include various needs for compute and storage resources, ranging from high-end compute resources with tightly coupled processing to more loosely-coupled resources, e.g., connected by grid environments. The key aspects of the proposed Grand Challenges are summarized in this report.

The results provide a basis for increased collaboration between the projects and organizations that are responsible for providing the national infrastructures for computational science. Collaborations that should be investigated include the exchange of resources (and users) between the countries such that researchers can get access to infrastructure resources abroad that are not (or insufficiently) available nationally. The national infrastructures should also collaborate such that large aggregate resources are provided to research communities that address large-scale computational problems. Also coordinated operations and support, a coordinated infrastructure for compute and storage resources, and the joint development of new end-user services and unified interfaces on the Nordic level should be considered. This way, collaborating groups in the Nordic countries can make use of multiple resources across the Nordic countries and need not have to deal with different interfaces and policies that are maintained within the individual countries.

The results of this survey are communicated to the national funding agencies in the four countries and all other stakeholders and provide a basis for further discussions between the funding agencies, research communities and the management of the national infrastructures on how to organize a coordinated Nordic effort. The objective should be that eventually several of the Grand Challenges that are included in this report will be given access to large-scale infrastructure resources, possibly spread over several countries in a way that is transparent to the end-user.

The results of the survey could be used as a basis for a catalogue of computational Grand Challenge problems that exist within the Nordic research community. This catalogue could be updated regularly by carrying out similar surveys in the coming years and by establishing mechanisms to update the catalogue.

April 16, 2007

Gothenburg, Siniša Krajnović, SNIC, Sweden
Member of the SNIC Board

Espoo, Jura Tarus, Finnish IT center for Science (CSC)
Specialist in nanoscience

Trondheim, Jacko Koster, UNINETT Sigma AS
Project Manager of the Notur II project

Contents

1 Introduction **1**

2 Present infrastructure in the Nordic countries **2**

2.1 National organizations 3

2.1.1 Finland 3

2.1.2 Norway 3

2.1.3 Sweden 4

2.1.4 Denmark 5

2.2 Facilities 5

2.2.1 Finland 6

2.2.2 Norway 7

2.2.3 Sweden 8

2.2.4 Denmark 10

2.3 Resource allocation 11

2.3.1 Finland 11

2.3.2 Norway 11

2.3.3 Sweden 11

2.3.4 Denmark 12

2.4 Large computational projects 12

2.4.1 Finland 12

2.4.2 Norway 13

2.4.3 Sweden 13

2.4.4 Denmark 14

2.4.5 WLCG collaboration 14

2.5 Contacts and addresses 15

3 International trends in technology **15**

4 Computational Grand Challenges **16**

5 Observations and Recommendations **17**

6 Workshop **19**

7 Summaries of the Grand Challenges **19**

7.1 Categorization 29

8 Grand Challenge descriptions **31**

8.1 **ABICAT** - *Ab initio* molecular dynamics and statistical studies of catalytic reactions 31

8.2 **ATMOTURB** - From Kolmogorov theories to fluid mechanics of climate change 32

8.3 **BGC** - Biomedical Grid Computation 33

8.4 **BHOO** - A Brief History Of the Oceans 34

8.5 **BIOGRID** - Bioinformatics for the grid 35

8.6 **BRAINSIM** - Full scale, biologically detailed neocortex simulation 36

8.7 **CarboClint** - Carbon Cycle Climate Interactions 37

8.8 **CATOX** - Nanocatalysis at the oxide-metal interface 38

8.9 **CLIQAIMD** - AIMD simulations of complex liquids 39

8.10 **CMPQ** - Computation of novel materials with predictive quality 40

8.11 **COMMIS** - Charge and Optical induced Manipulation of individual Molecules on Insulating Substrates 41

8.12 **FINALICE** - Exploring the nuclear matter with the ALICE experiment 42

8.13 **FINNCMS** - Finnish CMS Project 43

8.14 **GIG** - Gene Interactions in General Pedigrees 44

8.15 **GRAMS** - GRAnd and not so grand challenges in computational Material Science 45

8.16 **HetCat** - Unsolved Problems in Heterogeneous Catalysis 46

8.17 **HiMuMoMo** - New generation of force fields for predictive hierarchical multiscale molecular modeling 47

8.18 **LIPOPRO** - Large-scale atomistic and coarse-grained modeling of lipoproteins 48

8.19 **LOWSOUND** - Generation and propagation of sound in low Mach number internal flows . . 49

8.20 **MHDTurbu** - High resolution magnetohydrodynamic turbulence simulation 50

8.21 **MOIS** - An integrative software system for management, analysis and design in molecular-level chemical and biological calculations 51

8.22 **MomentA** - Nuclear Electromagnetic Moments via Atomic Theory 52

8.23 **MURPF** - Microphysics in Ultra Relativistic Plasma Flows 53

8.24 **NANOGOLD** - First principle design of ligand-protected gold nanoparticles 54

8.25 **NANOQUANT** - Understanding Nanomaterials from the Quantum Perspective 55

8.26 **N-CC-GC** - Nordic Climate Computing Grand Challenge 56

8.27 **NPQCD** - Non-Perturbative Quantum Chromodynamics 57

8.28 **Pss1Å** - Protein structure from sequence to resolution 1Å 58

8.29 **QUADCASC** - Quantum dynamics of collision cascades 59

8.30 **SAM** - Stellar atmosphere modeling 60

8.31 **StaFraLa** - Statistical Fracture Laws 61

8.32 **STP** - Simulation of Tsunami Propagation 62

8.33 **TURBSIM** - Direct numerical simulation of turbulence, transition and separation for flows around bodies 63

8.34 **Vioprocg** - Violent processes in geophysics 64

A Proposal form **65**

B Workshop **66**

1 Introduction

Increasingly more scientific disciplines discover the possibilities and benefits of using computer simulation to advance the knowledge of phenomena and problems that are considered important to understand and challenging to solve. Computational research can be a lot more economical, environmentally friendlyand faster to produce results than laboratory experiments and field observations. Computational research can also help steer the development of theory and experiments, verify theory and produce information that cannot be obtained by traditional observation (e.g., long-term climate modelling).

The use of electronic resources also makes it possible to establish new links between scientific disciplines and combine data and information from a variety of sources to address new scientific territory and problems that cannot be tackled by one scientific discipline alone. Nowadays, important scientific discoveries are being made in areas in between the more traditional sciences and new challenging problems require a coordinated effort from a group of researchers that have skills from several of these traditional disciplines. Intelligent use of compute resources and digital information plays an important role in this.

As researchers are simulating physical phenomena at increasingly larger scales and aim for increasingly better accuracy in the results, the size and complexity of the computational problems and the data sets are growing rapidly. This leads to a constant demand for larger electronic resources to carry out the simulations and process and store data. A modern infrastructure for computational science which includes large scale resources is nowadays essential for doing high quality research in a large number of scientific fields. The infrastructure must provide the necessary resources and services to enable scientists to do research of high international standard in close collaboration with their colleagues in other countries.

Many of today's large scale computational facilities for academic research are shared resources. The facilities are expensive and are therefore usually being financed by large organizations (e.g., universities) and/or national funding agencies (e.g., Research Councils or Ministries). At any time, the facilities are used by multiple research projects. Several applications from different research groups are executed simultaneously on each facility. Even if the facilities are large in terms of the number of processors, memory, and storage, there are limitations to their capacities.

To distribute the available computational resources over the research groups that wish to use the facilities, national committees have been established that decide on the allocation of the resources (CPU-hours). Usually, the total amount of resources requested, by all research groups together, exceeds by far the total available capacity of the facilities. This means in practice that the research groups only get part of the resources that they applied for, even though they have a real need for a larger amount of resources. With limited compute resources available nationally, scientists need to adapt their scientific goals and software applications to the resources that they can get access to and they need to use these carefully, thereby prioritizing and postponing scientific activity. Ultimately, this may lead to a degradation of the research carried out, and the researchers may not be able to address challenging scientific questions that require computationally or data intensive simulation.

The Nordic countries invited the research community in the four countries to describe large scale scientific problems that are recognized as being important in their field, that are challenging to solve, and that require the use of large scale compute and/or storage resources for an extended period of time. This offered scientists the possibility to describe research objectives that are not restricted by current limitations in nationally available resources. The goal was not to obtain a listing of already on-going projects, but to identify well-defined scientific problems that will be of significant interest to the Nordic research community in the coming years and that have special/challenging requirements for the national infrastructures, in particular with respect to aggregate compute and storage capacity.

The survey for Grand Challenge applications was conducted with several other goals in mind. On the national level, these include:

- Encourage researchers to define scientific problems independent of the size of the resources that they have access to (or believe they could get access to)
- Increase the visibility of computational science and infrastructure as an important tool for solving large scale complex problems
- Identify trends in large scale computationally-oriented research

And on the Nordic level:

- Encourage research groups in the Nordic countries to define joint large-scale projects that are not limited by the resources that are available on the national level
- Identify the need for resource sharing between the Nordic countries motivated by science (besides economical considerations). Strengthen Nordic collaboration between the projects and organizations that are responsible for providing the national infrastructure for computational science
- Find a mechanism for identifying computationally intensive grand challenges within the Nordic research community
- Raise awareness on the Nordic level for e-Infrastructure as one of the primary instruments for computational science and identify the need for Nordic level funding for e-Infrastructure
- Assist organizations in the Nordic countries to attract Nordic (or European) funding for e-Infrastructure, by providing scientific cases that justify and require Nordic collaboration

Whenever the term infrastructure (or e-Infrastructure) is used in this report, we mean the combination of hardware resources (computers, storage, networking), software (tools, grids, etc.), end-user services (interfaces, quality assurance, etc.), and operations and support. Resources needed to maintain (and not only store) scientific data and digital repositories are also an integrated part of infrastructure. Usually, the compute resources get much attention in discussions about infrastructures as these often represent the more expensive (and tangible) part of the infrastructure.

The report consists of two parts. Section 2 summarizes the present situation of the national infrastructures for computational science, together with the main user groups that exist within the countries. Sections 4-8 contain an overview and discussion of the Grand Challenges that were submitted by the research community.

2 Present infrastructure in the Nordic countries

This section gives a brief overview of the organization and usage of the infrastructure for computational science in the four Nordic countries. The section describes the present organization, the present compute facilities, resource allocation procedures, and the main scientific disciplines that make use of the infrastructure in each of the four countries.

2.1 National organizations

2.1.1 Finland

The Finnish infrastructure for computational science is a centralized infrastructure. The Finnish Information Technology Centre for Science (CSC) is one of the largest supercomputing centres in the Northern Europe and was established in 1971. Today, it is a powerful information technology centre for high-performance computing and networking recognised at national and pan-European levels. CSC-Scientific Computing Ltd. is owned entirely by the Finnish state, administered by the Ministry of Education, and operated on a non-profit principle. CSC states its mission as developing and providing computing, modelling and information services for universities, research institutions and industry. CSC develops and provides a range of publications and training options and is active in disseminating information on CSC's objectives, services and results to decision-makers, national and international stakeholders and general public. CSC's essential objectives include supporting computational science, transfer of skills and building up cooperation. All services are based on a versatile supercomputing environment, fast data communications and expertise in various scientific disciplines and information technology. Currently, CSC has ca. 150 employees. Turn-over for 2006 was 15.5 MEuro.

2.1.2 Norway

The responsibility for the Norwegian e-Infrastructure for academic use rests with the e-Science programme (eVITA) that was established late 2005 by the Research Council of Norway. The eVITA programme board has an advisory committee (ReInfra) that advises on the strategic development of the infrastructure.

The operational responsibility for the national infrastructure is delegated to UNINETT Sigma AS through the Notur II project and a consortium of partners. Notur II is a 10-year project under the eVITA programme. The Norwegian infrastructure for computational science is a distributed infrastructure that involves a number of HPC groups at several universities. The HPC groups operate supercomputer facilities and provide user support services. The activity of the Notur II project was continued from the Notur I project that ended in 2004. The consortium partners are:

- UNINETT Sigma AS, Trondheim
- Norwegian University for Science and Technology (NTNU), Trondheim
- University of Bergen (UiB)
- University of Oslo (UiO)
- University of Tromsø (UiT)
- Meteorological Institute (met.no)

UNINETT Sigma has a board in which five project partners are represented. UNINETT Sigma reports progress, plans and budgets for the project to eVITA. UNINETT Sigma has an advisory board that consists of the IT-management of the Notur partners. The board advises on operational and long-term aspects of the Notur project.

The Research Council of Norway has appointed a Resource Allocation Committee that has the responsibility of assigning the national compute resources to the research projects. Major objectives of the committee are to review the quality of the science that is carried out on the resources and to maximize the efficient use of

the resources. The committee evaluates applications for resources based on scientific and engineering merit, needs, and impact of the proposed research. The committee prioritizes innovative usage of the resources. The committee reports to the eVITA programme board. See also Section 2.3.

The Notur II project is financed in part by the Research Council of Norway. The contribution from the Research Council of Norway is guaranteed for a 10-year period (2005-2014) and is at least 21.7 MNOK per year. There is additional funding for special purposes and projects. The funding from the Research Council is complemented by cash and in-kind from the consortium partners.

In February 2007, the Research Council of Norway granted UNINETT Sigma 4.9 MNOK to establish a national grid infrastructure (NorGrid) that will include the resources and partners of the Notur II project.

2.1.3 Sweden

The Swedish National Infrastructure for Computing (SNIC) is a metacentre for high performance computing (HPC) and was initiated in 2002. SNIC has six member centers that provide national HPC services. The SNIC member centers are:

- High Performance Computing Center North (HPC2N), Umeå
- Centre for Scientific and Technical Computing at Lund University (LUNARC)
- National Supercomputer Centre (NSC), Linköping
- Center for Parallel Computers (PDC), Stockholm
- Center for Scientific and Technical Computing at Chalmers (C3SE), formerly UNICC, Gothenburg
- Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX)

The responsibility for SNIC's activities under the Swedish Research Council (Vetenskapsrådet) is held by a board that reports to the Director General of the Council. The board consists of a chairman and eight other members, appointed by the board of the Swedish Research Council.

The Strategic Technical Advisory Committee (STAC) serves as an interface to the member centers and as an advisory forum for decision items for the SNIC board through the director. The committee consists of the directors of the SNIC-member centers. Specific technical experts can also be co-opted to this committee. The joint center projects such as SweGrid, and national level collaborations are managed by STAC.

The Swedish National Allocations Committee (SNAC) serves the academic community by seeking to maximize the efficient use of national computational, visualization and data warehousing facilities through the evaluation of applications for resources based on scientific and engineering merit, needs, and impact of the proposed research. The mission of the committee also includes stimulation of new innovative uses of the SNIC resources through interaction with the user community. See also Section 2.3.

In 2005, the SNIC User Group (SNUG) was formed. SNUG's mission is to provide a communication interface between SNIC and its users to provide information on SNIC resources and services as well as to gather comments and user requirements. SNUG also has the mission to organize user meetings and will be responsible for the coordination of the SNIC-interaction meetings in the future. SNUG has the mandate to nominate two of the SNIC board members.

The SNIC budget for 2007 from the Swedish Research Council is 49.2 MSEK. Besides the annual grant from the Research Council, there is additional funding for special purposes and funding from private foundations such as the Knut and Alice Wallenberg foundation.

2.1.4 Denmark

The Danish Center for Scientific Computing (DCSC) was formed in 2001 under the Ministry of Science, Technology and Innovation with the aim of ensuring researchers in Denmark access to processing power in connection with their research projects.

DCSC is a decentralised umbrella organisation binding together the four regional operating centres at the Technical University of Denmark (DCSC/DTU), the University of Copenhagen (DCSC/KU), the University of Southern Denmark (DCSC/SDU), and Århus University (DCSC/AU).

The board of DCSC consists of 9 members. Four members represent the regional operating centres, two members come from industry, two members represent the research councils and there is one international member.

In DCSC's six years of existence, 15 MDKK per year has been invested in computer equipment. The main focus of DCSC is to support hardware, but DCSC can also provide support for software development.

The Danish Center for Grid Computing (DCGC) consists of a series of research and development projects and was established in August 2003 with a three year grant from the Ministry of Science, Technology and Innovation. The project ended ultimo 2006 and is no longer in existence. However, plans are being made to make DCGC a sustainable part of DCSC.

2.2 Facilities

Today, a diverse mixture of hardware resources is put in operation to satisfy the needs of the scientific user community. Some scientific disciplines require access to primarily tightly-coupled parallel computing resources, while other disciplines can make efficient use of more loosely-coupled capacity (cluster) resources, or truly distributed environments as provided by today's grid computing technologies. Also, the programming paradigms used varies in application areas and ranges from throughput computing (including embarrassingly parallel applications) to message-passing (MPI) and shared-memory parallelism (OpenMP). These diverse demands cannot be covered by a single hardware resource. Each of the national infrastructures therefore operates a variety of hardware resources.

A supercomputer facility typically consists of one or more compute nodes, each of which contains a number of processors that share a pool of memory and limited disk capacity (internal to the node). All nodes and processors can usually access a larger shared storage area through a file system, and possibly also a secondary storage (tape) resource. Supercomputer facilities typically use variants of the UNIX operating system (Linux, AIX, etc.). The cost of hardware is usually driven by the number of compute nodes, the speed of the processors, the amount of memory per node, the required I/O performance (memory-disk), and the bandwidths and latencies of the node-interconnect.

In the following section, we list per country the resources that are currently in operation. We emphasize that the configurations of the resources are subject to change and we recommend the reader to contact the relevant organizations for up-to-date information.

2.2.1 Finland

The current computing facilities at CSC consist of three different systems.

Supercomputer. A new Cray XT4 system (louhi.csc.fi) has recently been installed at CSC. The first configuration of the XT4 consists of 2048 computing cores (1024 dual-core AMD Opteron 2.6 GHz processors). This system will be expanded during 2008 to more than 6500 computing cores leading to 70 Tflops/s aggregate performance. The XT4 will be profiled for codes that benefit from the system's fast interconnection network. The XT4 replaces the IBM eServer Cluster 1600 system (ibmsc.csc.fi, 16 IBM p690 nodes, 512 processors, 2.2 Tflops/s) that is currently being phased out.

Application server. The application server corona.csc.fi consists of two Sun Fire 25K servers. The total system is composed of 96 UltraSPARC IV dual-core processors, contains 384 GB memory and has a theoretical peak performance of 460 Gflops/s. A large variety of scientific software programs and databases are also available on the server. Typical size of (parallel) jobs on this server is 1-8 processors.

Supercluster. The HP ProLiant Cluster (sepeli.csc.fi) consists of 256 HP ProLiant DL145 and DL145G2 compute nodes. The cluster has altogether 256 2.2 GHz AMD Opteron dual-core processors, about 1 TB of memory and 40 TB of local disk space. The system also includes 11 TB shared disk space. The system is profiled mainly to small and medium size jobs. However, the recent upgrade with an InfiniBand node interconnect has changed this situation somewhat. Usage statistics show that 12% of the jobs executed on this resource use 32 processors or more.

CSC has also acquired a new HP CP4000BL ProLiant cluster, integrated with a 100 TB high-performance storage system. This cluster uses the recently introduced cClass blades from HP, incorporating a 4X DDR InfiniBand interconnect. The cluster contains 2048 2.6 GHz AMD Opteron compute cores, 4 TB of memory, and has a peak performance of 10.6 Tflops/s. The system will be operation in May 2007.

Name	System	Type	Number of cores	CPU type	Peak perf.	Total memory	Total disk
Louhi	Cray XT4	cluster	2024	AMD Opteron	10.6 Tfbps/s	2024 GB	74 TB
Murska	HP CP4000BL Proliant	cluster	2048	AMD Opteron	10.6 Tfbps/s	4000 GB	98 TB
IBMSC	IBM p690	distr. SMP	256	Power4	1100 Gfbps/s	416 GB	2 TB
Corona	Sun Fire 25K	SMP	192	UltraSPARC IV	461 Gfbps/s	384 GB	2 TB
Sepeli	HP Proliant DL145G2	cluster	512	AMD Opteron	2200 Gfbps/s	1000 GB	31 TB

Other Finnish academic compute resources. The Material Sciences National Grid Infrastructure (M-grid) is a joint project between CSC, seven universities and the Helsinki Institute of Physics (HIP). The project builds on a network of computers that are situated at different locations but are in shared use through grid technology. The systems are connected to the Nordic NorduGrid network, but access is currently limited to M-grid partners and CSC customers. The systems are particularly suitable for high-throughput computing (sequential and easy-to-parallel applications). The aggregate compute capacity is approximately 2.5 Tflops/s. M-grid is based on HP ProLiant DL145, DL385 and DL585 servers equipped with AMD Opteron processors (642 altogether). The University of Oulu has its own AMD Athlon MP based cluster (94 processors) that is also connected to M-grid. (For more information on M-grid, see http://www.csc.fi/proj/mgrid/)

The University of Jyväskylä has acquired an SGI Altix 4700 server with 96 processors and 192 GB of shared memory. The machine is dedicated to nanoscience research.

The Linux cluster at the Biomedicum at the University of Helsinki consists of 120 cores (dual-core AMD Opteron processors). The cluster is used for sequence analysis, gene expression analysis and genetic linkage/association analysis.

Storage. The storage capacity at CSC consists of work directories on different computing systems (about 200 TB, no back-up) and centralized home and project directories (total about 160 TB, back-up).

For long term storage, CSC provides an archive service that is accessible by normal unix commands. The maximum capacity of the tape library with the current tape drive technology (LTO3) is 3 PB. The tape drives will be replaced by LTO4 technology in 2008 which duplicates the storage capacity of the robot to 6 PB.

Staffing and support. CSC employs approximately 150 specialists. Most employees are highly qualified in information technology or one of various scientific disciplines. Around 80% of the employees has an academic degree. CSC's specialists maintain the widest selection of scientific software in Finland and take responsibility for software installations and database version updates. Approximately 200 programs and 70 databases are installed in the CSC server environment. CSC is the national license holder and distribution center for several programs and databases.

During 2006, CSC organized 74 different events (courses, seminars and workshops) with a total of 1956 participants.

National academic network. Funet (Finnish University and Research Network) is a high-speed data communications network serving the Finnish research community. It connects about 80 research organizations and 350 000 users. The redundant Funet backbone utilises 2.5 Gbit/s and 10 Gbit/s links. Funet has 10 Gbit/s connectivity to GÉANT2 via the NORDUnet hubs in Sweden and Denmark. The Funet services are maintained by CSC.

2.2.2 Norway

The status of the HPC resources in Norway as of January 2007 can be summarized as follows.

Hardware

The university partners of the Notur consortium operate a number of HPC facilities. Some characteristics of these facilities are listed below. More detailed information can be found on www.notur.no.

Name	System	Type	Number of CPUs	CPU type	Peak perf.	Total memory	Total disk
Magnum	HP SuperDome	SMP	64	Itanium2	84 Gfbps/s	128 GB	23 TB
Njord	IBM p575+	distr. SMP	992	Power5+	7500 Gfbps/s	1984 GB	70 TB
Snowstorm	HP rx4640	cluster	400	Itanium2	2208 Gfbps/s	400 GB	16 TB
Tre	IBM p690	distr. SMP	96	Power4	499 Gfbps/s	320 GB	>7 TB

The resource njord at NTNU is the resource that is used for the operational weatherforecasting by the Norwegian Meteorological Institute. The nodes of njord are connected by IBM's HPS switch technology. The nodes of the cluster snowstorm at UiT have InfiniBand interconnect. Magnum at UiO is currently the largest single-node SMP in Norway in terms of processors (64 CPUs), while one of the three 32-way shared-memory nodes on tre has largest memory (192 GB).

Storage. The current storage capacity in the Notur project consists primarily of disk systems that are tightly

coupled to the supercomputer facilities and (campus) tape systems for backup and archiving. The compute resources have today access to around 100 TB disk and 300 TB tape. The need for large storage is concentrated to a subset of the research projects that make use of the compute resources, primarily in climate modeling.

Staffing and support. The Notur partners operate a national helpdesk and use a common request tracking system that registers email correspondence. Most questions from users refer to specific resources and resource-specific queues are implemented. The advantage of the national helpdesk is that the support personnel at the various sites use the same system and the users see only this single system.

The Notur project enables scientists to apply for advanced user support projects. Advanced user support aims at helping scientists to improve or extend the computational performance and capabilities of their applications. This can be in a number of ways. Advanced support includes code parallelization, optimization, benchmarking, improving user-interfaces, complex application enabling and software development. A main aim is also to achieve a better utilization of the (expensive) supercomputer facilities. The activity gives priority to scientists with compute- and data-intensive applications as well as scientists that start using the resources in the national infrastructure. Advanced user support is by application and project based.

The centers that operate supercomputer facilities employ staff for the operations and user support. Special 24/7 surveillance is implemented for njord (for met.no's operational weatherforecast).

Other resources. There are a number of resources operated by the Notur partners that are not included in the national user allocation system. These include:

Name	System	Type	Number of CPUs	CPU type	Peak perf.	Total memory	Total disk
Titan	Dell	cluster	-	Xeon	-	-	-
Fimm	IBM e1350	cluster	172	AMD Opteron	825 Gfbps/s	258 GB	15 TB

The cluster titan at UiO is being upgraded at the time of writing this document. The final configuration of titan will have approximately 1000 compute cores. A subset of the nodes of titan is connected with InfiniBand interconnect. Twenty-five of the nodes of the cluster fimm at UiB are connected with SCI/Dolphin interconnect.

National academic network. UNINETT AS operates the national network for research and education in Norway. The primary customers are the universities, regional high schools and research institutions. Its backbone is 2.5 Gbit/s (2006). This will be upgraded to double 10 Gbit/s links during 2007. The HPC resources at the universities are all connected to the UNINETT network via the local campus network. The UNINETT network is connected to NORDUnet with a 10 Gbit/s link.

New HPC resources. During 2007 and 2008, a significant increase in resource capacity (computing, storage and networking) will be implemented. New compute systems will be taken in operation at several of the university partners. The intention is to expand and include the Titan cluster into the Notur project in the first half of 2007.

2.2.3 Sweden

Most systems at the SNIC centers are today based on commodity hardware which gives high performance at relatively low cost. Large existing systems are also equipped with high-end interconnects. SNAC grants access to 12 clusters with a total of about 2700 AMD or Intel processors, counting the six SweGrid clusters individually (although they are considered as one resource in SNAC allocations). There are also clusters exclusively for local

use as well as clusters owned by research groups, but managed by SNIC centres. Funding for these systems has been obtained from other sources such as the Wallenberg foundation.

Name	System	Type	Number of CPUs	CPU type	Peak perf.	Total memory
Sarek		cluster	384	AMD Opteron	1.69 Tfbps/s	1536 GB
Seth		cluster	240	Athlon MP2000+	800 Gfbps/s	120 GB
Lenngren	Dell 1850	cluster	884	Xeon EM64T	6 Tfbps/s	3400 GB
Lucidor	HP	cluster	180	Itanium2	648 Gfbps/s	540 GB
Monolith		cluster	400	Xeon	740 Gfbps/s	400 GB
Mozart	SGI Altix 3700 Bx2	SMP	64	Itanium2	410 Gfbps/s	512 GB
Ra		cluster	280	AMD Opteron	1.34 Tfbps/s	688 GB
SweGrid	GRID	cluster	600	Pentium 4	-	12 GB

Storage. The current SNIC storage solutions range from local disks to separate disk systems and to tape. SNIC computing resources have today access to around 100 Tbytes disk and 400 Tbytes tape (scratch disk not included).

In the context of SweGrid, significant storage capacity has been built up. Of the total amount of storage, SweGrid delivers 80 Tbytes of disk storage, and 270 Tbytes of tape storage. Most of the storage is available through grid-enabled interfaces.

The need for large amount of storage is concentrated to a few large projects with international partners. Other SNIC-projects have very modest storage requirements.

Staffing and support. A national helpdesk has been set up with a common request tracking system that registers email correspondence. Many queries refer to the functionality of specific resources. The purpose of the national helpdesk is to unify support so that users and support personnel will only have to interact with one type of support system.

In-depth support includes activities to support specific users and projects that are not handled by helpdesk support and is performed at the user's local HPC centre. The type of support given varies and reflects the needs of the user: examples include assistance to new users to get started, porting of codes, grid-enabling of codes, parallelization and optimization of codes. The SNIC and university-funded application experts form an integral part of the in-depth support structure. The application experts also work in collaboration with projects to extend and improve existing codes.

National academic network. The Swedish University Computer Network (SUNET) interconnects all universities in Sweden in a network called GigaSunet. The HPC resources of SNIC are all connected to GigaSunet, directly or indirectly, via the local campus network of the hosting university. All HPC users at Swedish universities are also connected to GigaSunet via their campus network.

GigaSunet is a wide area network where the backbone is based on routers in a ring topology interconnected with 10 Gbit/s links. GigaSunet is connected to NORDUnet and GÉANT. The capacity of GigaSunet with its international connections is not a limiting factor for the current HPC usage. However, the campus networks and the connection to the local computer systems are often bottlenecks in data communication for the research.

New HPC resources. During 2007 and 2008, a significant increase of the SNIC resources will take place. This process has been started by the three new national Linux clusters that became operational in spring 2007. Two local HPC centers (C3SE and LUNARC) have joint their funding obtained from SNIC to purchase two systems containing 552 dual-core nodes. Every node has four Xeon 3 GHz cores and 100 GB local storage. The majority of the nodes have 4 GB memory and a Gigabit Ethernet, 24 nodes have 8 GB memory and Infinipath

(InfiniBand) interconnect.

Another HPC center that has obtained funding from SNIC for a new Linux cluster is UPPMAX. The resource that was acquired from IBM is of capacity type consisting of 200 x3455 nodes, each having 2 dual-core, 2.8 GHz, AMD2220 CPUs and 8 GB memory. The interconnect will be Gbit Ethernet. There will be a total of 800 cores and 1600 GB memory, and a theoretical peak performance of 4.5 Tflops/s.

SNIC has also granted NSC approximately 30 MSEK for a new national capacity system. This system will probably be operational at the end of 2007 and will be most powerful Swedish capability system and definitely on the list of most powerful computers worldwide.

Besides the Swedish Research Council, the Knut and Alice Wallenberg foundation has also contributed funding for the Swedish HPC infrastructure. At the end of 2006, this foundation granted 20 MSEK for the investment in new HPC equipment. SNIC will be responsible for funding the maintenance and support for this equipment.

2.2.4 Denmark

The Technical University of Denmark (DTU) hosts a computer (Niflheim) with 1082 processors, more than 2.1 TB of memory and 36 TB local disk. The total peak performance of the resource exceeds 5.9 Tflops/s. The main compute resource consists of 816 AMD Opteron processors in 204 IBM servers. Each node is equipped with 8 GB of memory and local disk. The second resource consists of 266 IBM ThinkCentre S50 professional desktop PCs with Intel Pentium-4 (3.2 GHz). Each PC has 2 GB of memory and 80 GB disk. A bioinformatics installation at DTU includes four SGI Altix 3000 shared-memory servers with 32, 64, 64, and 128 processors (and 123, 182, 184, and 734 GB memory, respectively) and a storage unit of ca. 20 TB disk. Another HPC installation at DTU consists of several SunFire servers with UltraSPARC III and UltraSPARC IV dual-core CPUs. Finally, DTU hosts a 210-node PC-cluster consisting of 210 Dell PE650 rack-mounted servers, each with one 2.4 GHz Pentium-4 CPU and 1 GB memory. The theoretical peak performance of the cluster is about 1 Tflops/s.

The University of Copenhagen hosts an SGI Altix system (Brahe) that consists of 64 1300Mhz Itanium2 processors, a total of 128 GB memory and 5.2 TB disk. The cluster Steno contains ca. 646 AMD Opteron processors (more than 760 cores) and a 30 TB SAN disk array. Another 100 TB disk storage system is also connected to Steno. In total, 118 nodes in the cluster are connected with InfiniBand.

The University of Southern Denmark hosts a compute cluster (Horseshoe) that contains 200 Dell PowerEdge 1950 servers, each with two 2.66 GHz Intel Woodcrest dual-core processors.

The University of Aarhus hosts an SGI Altix 3700 shared-memory system with 64 Itanium2 1.5 GHz processors and 256 GB memory. Peak performance at 384 Gflops/s. The university also hosts IBM pSeries systems (Fenris, Sleipner, Hugin and Munin) with a total of 80 processors. Additionally, five Sun X4100 AMD Opteron-based servers are available with a total peak performance of 1.5 Tflops/s.

National academic network. The Danish Research Network (Forskningsnettet) is a high-speed internet supplier connecting Danish universities and research institutions. Forskningsnettet is connected to the international research and education networks through NORDUnet and GÉANT.

2.3 Resource allocation

There is considerable similarity between the Nordic countries in the way researchers obtain access to the national supercomputer facilities.

2.3.1 Finland

Researchers acquire access to the supercomputer facilities by application. The basic principles and priorities of CSC's resource allocations are defined by CSC's Board of directors.

The HPC resources are allocated by the CSC Resource Allocation Group that is nominated by CSC's managing director. The group consists of a chairman, specialists in various disciplines, and persons responsible for the preparation of resource allocations. The specialists are nominated based on their scientific and technological expertise and experience.

The CSC Resource Allocation Group accepts applications from researchers continuously, without annual deadlines, and holds meetings every three weeks so that the applications can be handled in due time. The applications are evaluated by scientific merit and quality (e.g., scientific publications). The group follows closely the national science policies and priorities defined by the Finnish government. The group reports once a year to CSC's Board of Directors and to the Ministry of Education.

2.3.2 Norway

Researchers acquire access to the supercomputer facilities by application. A Resource Allocation Committee (RFK) defines the procedures for application, evaluates the applications, and decides on the allocation of resources. The RFK is given its mandate by the Board for the Division of Science of the Research Council of Norway. The RFK reports to the eVITA Programme Board. The RFK committee is composed of researchers from different scientific disciplines.

The assessment of the scientific quality of the applications is carried out by the members of the RFK. Applications for large allocations go through an external review.

The current RFK guidelines and procedures for resource allocation are described on the Notur home page (www.notur.no). Allocations are in terms of CPU hours (per facility) per 6-month period. Projects that require large allocations can apply for a 12-month period. When a project has consumed its allocation, the project must apply for new allocations.

The university partners that operate supercomputer facilities also have local routines for allocations on the local resources (which normally also includes a part of the national facilities).

2.3.3 Sweden

Researchers acquire access to the supercomputer facilities by application. The Swedish National Allocations Committee (SNAC) is appointed by SNIC and issues calls for proposals twice per year. After a reviewing process SNAC allocates available resources. The proposals are reviewed by the SNAC committee members,

each with a responsibility for a specific research area. The applications also used as a database to assess the needs of the users and identify trends.

The SNAC assessment of scientific quality is carried out through reviews of the proposals by the committee members. Most of the projects receiving time from SNAC have been reviewed when funding for the project was obtained. The scientific quality of the proposals are therefore generally high. The SNAC emphasis is on ensuring that adequate resources are allocated to each project.

The current SNAC policy and procedures are described on the SNAC home page. The allocations are currently in terms of CPU hours/month. When a user has consumed the monthly allocations, the user can still submit jobs but they may run with lower priority.

The SNIC centres also have local routines for allocating resources. At LUNARC and UNICC all computer time is handled this way, while PDC, NSC and HPC2N have both local and SNAC allocations. UPPMAX has recently moved from the former to the latter group. For smaller allocation requests on SNAC resources, up to 1000 hours per month, researchers may apply directly to a SNIC centre to become a local user.

2.3.4 Denmark

Researchers acquire access to the Danish supercomputer facilities by application. It is also possible for researchers to apply for a grant to DCSC to buy and operate a compute resource. A substantial fraction of the applications is sent out for international review.

2.4 Large computational projects

2.4.1 Finland

The discipline consuming the largest amount of CPU time on CSC's computer systems is nanoscience. It represents nearly half of all the consumed CPU time. The second largest user communities are physics and chemistry. Overall usage by discipline in 2006 is:

discipline	CPU usage
nanoscience	30%
physics	27%
chemistry	24%
biosciences	12%
fluid dynamics	3%

There are currently no active projects at CSC with data-intensive applications, but at least two projects with intense data mining requirements will become active in the near future: LHC data from CERN and Planc satellite. Possible other data intensive projects may arise in bio-informatics and medical sciences. Below is the list of the ten largest projects in terms of CPU time:

1. Numerical modelling of plasma and fusion physics; Physics
2. Ion irradiation physics in materials; Nanoscience
3. Electronic, magnetic, optical and chemical properties for nanoparticles; Nanoscience

4. Lattice simulations of relativistic theories; Physics
5. Multiscale modelling of surfaces and surface-reactions; Nanoscience
6. Ab initio molecular dynamics methods; Chemistry
7. Computational study of NMR and EPR parameters; Chemistry
8. Nanocatalysis on metal surfaces; Nanoscience
9. Interactions of molecules “in silico”; Chemistry
10. Electronic structure of condensed matter and phenomena related to it; Nanoscience

2.4.2 Norway

The usage by discipline of the Notur resources in the period from October 2005 to September 2006 can be summarized as follows:

discipline	CPU usage
geosciences	41%
chemistry	34%
physics	13%
forecast	6%
other	6%

Here, ’forecast’ stands for weatherforecast simulations (production runs) that are executed a number of times per day by the Norwegian Meteorological Institute.

The projects that were among the largest consumers of the resources in 2006 are:

- Quantumchemical calculations of molecular properties; Chemistry
- Climate modelling; Geosciences
- Quantumchemical modeling of catalysis and spectroscopy; Chemistry
- Electronic structure studies on perovskite oxides and metal hydrides; Chemistry
- Global chemistry and climate interactions; Geosciences
- Modelling of astrophysical plasmas; Astrophysics

More details of the scientific objectives and recent achievements of these projects can be found on the webpage www.notur.no/projects.

The largest single application on the Notur resources is the operational weatherforecast of the Meteorological Institute (using HIRLAM but changing to the Unified Model in 2007).

There exist research projects that have access to local resources (outside Notur II). Bioinformatics is a discipline that makes significant use of local resources.

2.4.3 Sweden

The usage by discipline of the SNIC resources in the period from October 2005 to September 2006 can be summarized as follows:

discipline	CPU usage
physics	30-60%
chemistry	20-50%
applied mechanics	5-10%

Other areas having smaller contribution are biology, computer science, electronics and geosciences.

There is a small group of users who clearly have needs that go beyond what SNAC allocation typically provides. In many cases, these users have access to other systems, such as SNIC local resources, or run their own systems. These users run scientific projects that require most HPC resources in Sweden. Some of these are

- Physical Chemistry
- Computational Fluid Dynamics
- Meteorology
- IceCube project (http://icecube.wisc.edu/)
- Bioinformatics
- Elementary Particle Physics

2.4.4 Denmark

The general computer usage in Denmark is mainly covered by researchers within the following areas:

discipline	CPU usage
physics	35%
chemistry	35%
bioinformatics	20%
applied mechanics	10%

2.4.5 WLCG collaboration

The four Nordic countries have obligations with respect to the experiments that will start at CERN's Large Hadron Collider later in 2007. Data from the LHC particle accelerator shall be distributed to the research community through the world-wide LHC Computing Grid (WLCG). The WLCG model consists of a number of levels (Tier's). At the first level (CERN), data is being gathered from the experiments, stored and processed and distributed to a number of Tier-1 centres. A Tier-1 center must have sufficient capacity to store and further process the data that comes from CERN and implement the necessary operational and support tasks. The Tier-1 centres make the data available to the Tier-2 centres that typically consist of collaborating centers and research groups that carry out specific simulations.

There are currently eleven Tier-1 centres defined world-wide. The Nordic countries together form a (distributed) Tier-1 center. The Nordic Tier-1 is coordinated by the Nordic Data Grid Facility (NDGF) that is hosted by NORDUnet. Each of the countries contributes hardware resources to the Tier-1, as well as personnel for operations and support. Hardware consists of compute facilities (x86 hardware), disk and tape storage, and adequate network capacity.

A number of Grand Challenge submissions that are described later in this report are related to the LHC data that is stored at the Tier-1 centres.

2.5 Contacts and addresses

Finland: Dr. Kimmo Koski, Managing Director Finnish IT center for science (CSC).
Email: kimmo.koski@csc.fi. URL: www.csc.fi

Norway: Dr. Jacko Koster, Managing Director UNINETT Sigma.
Email: jacko.koster@uninett.no. URL: sigma.uninett.no

Sweden: Assoc. Prof. Sverker Holmgren, Director Swedish National Infrastructure for Computing (SNIC).
Email: sverker@it.uu.se. URL: snic.vr.se

Denmark: Rene Belsø, Director Danish Center for Scientific Computing (DCSC).
Email: belso@dcsc.dk. URL: www.dcsc.dk

NDGF: Rene Buch, CEO Nordic Data Grid Facility, NORDUnet AS, Kastrup, Denmark.
Email: rene.buch@ndgf.org. URL: www.ndgf.org

3 International trends in technology

There exist a number of international trends in hardware and software technology that may influence the establishment of e-Infrastructure to support Grand Challenge projects in the Nordic countries. This section includes a non-exhaustive lists of these trends.

The scalability of several widely-used scientific applications is limited. Certain execution scenarios in popular chemistry applications such as Gaussian and ADF do not scale beyond a moderate number of processors (ca. 64) even if these are connected with a fast interconnect. Other applications do scale well up to a large number of processors (e.g., the ROMS ocean model). In general, many applications in practice are executed on 1-32 processors. Such applications may have long run-times or many instances of the same application are executed simultaneously (each instance using different data sets or parameter settings).

Technology developments are such that processor clock frequencies will no longer increase at the same rate as they have done in recent years. However, aggregate processor performance continues to increase (at fairly constant cost) due to new architectural techniques (e.g., multi-core). In 2007, cluster compute nodes with two quad-core processors will be common. All big investments in new compute resources in the Nordic countries in the coming years will have thousands of processing elements (processors/cores). Obviously, the research community must be able to make efficient use of such systems. Scientific applications that consist of multiple instances of the same binary (each of which is executed on a different data set) can most likely be migrated to such new architectures without many modifications. For parallel applications that are sensitive to inter-processor bandwidths and latencies, this may not be straightforward. Gaps between the growing numbers of processing units in modern compute resources and the limited scalability of the software running on it, must be bridged in the coming years. This will require significant effort in investigating and improving the scalability of the relevant scientific applications.

The above realization is also important for the Grand Challenges that are listed in this survey. Getting access to larger resources does not mean automatically that the research activity can be scaled up accordingly in a sensible way. Scaling up computer simulations often also requires more detailed input data (with finer grid resolutions, better accuracy, etc.) that may not be readily available.

Combining distributed resources by grid middleware technologies is also becoming increasingly common. Grid technologies can be used to provide larger and better resources to the research community, by aggregating the capacities of distributed (or loosely-coupled) resources and by optimizing the usage of existing resources. Grid technologies also provide uniform interfaces to heterogeneous resources.

Internationally, there is increasing emphasis on establishing infrastructures and standards that facilitate the sharing of data between distributed research groups and across scientific disciplines. The trend is that the research community is moving towards establishing digital repositories for scientific data that can be accessed (and are maintained) by international communities.

The national research networks are currently being upgraded with the next generation high-speed network technology. The new networks can carry multiple wavelengths (lambdas) simultaneously, each with 10 Gbit/s bandwidth. Lambdas can be established between specific end-to-end point and for specific projects. Multiple lambdas can in principle be combined to establish Tbit/s end-to-end connectivity. The large increase in network performance facilitates the aggregation of distributed compute and storage resources for single scientific applications and enables the use of new classes of distributed applications in e-Science.

4 Computational Grand Challenges

In the autumn of 2006, the Nordic countries invited the research community in the four countries to describe Grand Challenges that are of high interest to the research community and whose solution requires the availability of large-scale infrastructure resources. The survey was commissioned by the Danish Agency for Science, Technology and Innovation, the Finnish IT center for Science, the Research Council of Norway, and the Swedish Research Council.

The aim of the survey was to get an overview of

- important scientific questions that are being studied today with the use of large computational resources but that can be addressed significantly better by having access to much larger computational resources
- important scientific questions that will have an emerging need for large computational resources (in the coming ten years)

With large resources, we mean resources that are several factors larger than the largest national facilites available and which is likely to be more than the aggregate capacity of all available facilities in a single country.

A Grand Challenge is a large-scale complex problem in science and engineering with broad scientific, environmental an/or economic impact and whose solution can be advanced by applying high-performance computing techniques and resources. Grand Challenges are expected to comprise multiple investigators/groups working collaboratively as a team and a multi-year collaborative effort. In particular, a Grand Challenge should have many or all of the following characteristics:

- address a fundamental scientific question that is widely recognized as important and challenging to solve

- require large scale infrastructure resources (HPC, storage, networks)
- have international interest and relevance
- be a long-term (multi-year) sustained effort with intermediate goals
- require collaboration among research teams and communities

Compute intensive tasks that do not address a fundamental scientific problem, such as operational weather forecasting or other applications that are used for production purposes, are not considered Grand Challenges.

The problems addressed by the (existing) projects that consume large amounts of compute resources on the national facilities are not by default candidates for Grand Challenges. Such projects often consist of a number of related smaller and short-term projects that do have similar objectives and use similar computational methods and data sets, but address different scientific problems.

5 Observations and Recommendations

The invitation was sent out in August 2006 with deadline November 15. A total of 34 descriptions for Grand Challenges were received. Section 7 summarizes the Grand Challenges. In Section 7.1, we have aimed to categorize the submissions according to a number of pre-defined criteria. Section 8 contains more extensive descriptions of the submissions in terms of scientific objectives, methods, challenges, and intermediate goals.

Grand Challenges were received from various disciplines, in particular biosciences, chemistry, earth sciences, and physics. A number of Grand Challenges were submitted by individual researchers, while other Grand Challenges include collaborations from research groups from all Nordic countries.

Besides the motivations for the survey mentioned in Section 1 (Introduction), the survey also highlighted the following.

- The diversity in needs for large-scale infrastructure resources that exist within the different research communities. The requirements range from low-cost capacity resources to high-end capability resources and from low to huge demands for storage and corresponding support for (distributed) data management.
- The willingness of research groups in the Nordic countries to pursue collaborative research that requires infrastructure resources that are far beyond what is currently available. The need for large resources expressed in many of the submitted Grand Challenges, together with the collaborations that span groups of several countries, cannot be satisfied by single facilities or by single countries.
- The differences between the Nordic countries in the computationally-intensive sciences. Differences in the types of available resources between the national infrastructures can also be observed. The national resources (should) naturally reflect the requirements of the most demanding computational sciences in the country.

The Grand Challenge descriptions differ considerably in terms of the requested size (amount) of infrastructure resources. Also the lengths of the periods that are needed to address the Grand Challenges vary significantly. Some Grand Challenges require a long start-up phase to get a project going, in particular if it concerns new research activity, while other Grand Challenges are modifications (up-scaling) of on-going work.

Even though we believe that we have contacted all relevant disciplines and research groups in the Nordic countries, we do not claim that the received Grand Challenges provide a complete overview of all the Grand Challenge problems that are of interest to the Nordic research community. Researchers may not have been tempted to submit a proposal because the invitation did not include concrete promises or guarantees that the researchers would obtain funding or access to larger resources.

To create a Nordic e-Infrastructure that supports research on Grand Challenge problems, we recommend that the following points are investigated.

- Sharing of national resources between the Nordic countries. One should facilitate scientists to access compute resources in several countries. This is of particular interest for resources that are not available (in type or in size) in all countries. This requires that on the national level the user groups are identified that will benefit from using resources abroad. This can be achieved by an application and peer-review process. By establishing some exchange and accounting mechanisms, one can keep track of the cross-border usage of the resources. Buying access to compute resources in other countries for large computational tasks that require large compute power during relatively short periods should also be considered.
- For large-scale scientific problems that can be addressed with loosely-coupled compute and storage resources, one should establish Nordic grid environments that couple the various resources in the countries such that the applications have access to the desired large aggregate compute (and storage) capacity. In this way, resources in different countries can be used simultaneously for demanding projects. Such mechanisms also allow 'cycle-stealing', that is, the use of (idle) resources in other countries at times when there is a peak demand for resources in one of the Nordic countries. By combining the compute resources that are present in the Nordic countries, research communities can be given access to much larger aggregate compute capacity. This way, they are able to compete internationally with researchers in countries that can afford establishing resources that are much larger than any of the resources that are available in the Nordic countries.
- The establishment of a Nordic digital repositories for scientific data to support international research collaboration. The repositories can be used by research groups that have common goals and interests. Common repositories avoid situations where similar databases are maintained within each of the Nordic countries. Such infrastructure requires skills in distributed computing, distributed data management, and networking.

To implement one or more of these recommendations, long-term collaborations between the national infrastructure projects/organizations are needed. We encourage therefore that the national infrastructures establish more intense collaboration and address the above recommendations. In addition, funding models must be considered that support such Nordic collaboration on infrastructure and facilitates Nordic Grand Challenge projects.

The allocation of infrastructure resources in various countries to support a Grand Challenge project requires coordination and collaboration (and trust) between the Nordic countries. Issues such as how Nordic e-Infrastructure is funded, who will get access to the resources (and who decides this), and where the resources will be located have to be dealt with. Within each of the Nordic countries, committees already exist that decide on resource allocations on the national facilities. Once resources are identified for Grand Challenge projects, a similar Nordic resource allocation committee could be established.

Ultimately, one could consider the creation of a Nordic supercomputing center that builds on the national infrastructures but where the geographical location (country) of compute and storage resources is fully transparent to the researcher.

It will be interesting to identify similarities and possible synergies between the submitted Grand Challenges. This could result in increased collaborations between the research groups and scientific communities.

It may be worth considering to maintain a database of Grand Challenge problems for the Nordic region. The proposals that were received for this survey provide a basis for such database. New Grand Challenges can be added to this database (e.g., by invitation or regular calls) using a number of predefined criteria. Existing entries can be updated whenever scientific progress has been made, milestones reached, or more information is known about for example the computational requirements. The database can be used to provide scientific cases for collaboration on e-Infrastructure on the Nordic and European level. The database also highlights challenging applications that exist within the Nordic countries. The aim should be that the database contains a fairly complete overview of the Grand Challenge problems.

6 Workshop

A Grand Challenge workshop was held at CSC in Espoo on March 12, 2007. The workshop was connected to the Finnish Supercomputing Days 2007. The purpose of the workshop was to gather researchers, HPC experts as well as the representatives from the Nordic countries, HPC centers and research councils. The aim was to form a meeting point where researchers from the Nordic countries could present and discuss their needs for large-scale infrastructure.

More than fifty people attended the workshop. Nine of the submitted Grand Challenges were presented to illustrate the variety in the scientific fields, the multidisciplinary nature of the proposals and the diversity in e-Infrastructure requirements. The presentations showed that the needs for the e-Infrastructure resources are far beyond what is currently available.

The programme of the workshop is given in Appendix B.

7 Summaries of the Grand Challenges

This section gives a brief summary of the submitted Grand Challenges. For each Grand Challenge, the acronym, the full title, the main author, the main scientific discipline and a short summary is given, including the need for compute and storage resources. It should be noted that many of the Grand Challenges were submitted by multiple participating researchers or groups. The discipline is given for convenience only to help the reader select the submissions that are of his/her interest (several submissions involve researchers from different disciplines).

The Grand Challenges are ordered by acronym. More extensive descriptions are given in Section 8.

ABICAT - *Ab initio* molecular dynamics and statistical studies of catalytic reactions
Professor Kim Bolton, School of Engineering, University College of Borås, Sweden

Discipline: Physics

The objective of this project is to perform molecular dynamics and Monte carlo simulations related to catalytic processes which according to the authors have in the past been prohibited due to large computational requirements. The computer resource required for this project according to the authors is capability machine such as the IBM Blue Gene. There are no large memory and storage requirements.

See Section 8.1.

ATMOTURB - From Kolmogorov theories to fluid mechanics of climate change
E. Lindborg, Linné Flow Centre, Department of Mechanics, KTH, Sweden

Discipline: Computational Fluid Dynamics

The objective is to carry out large-scale direct numerical simulations (DNS) of stratified turbulence including scalar transport and to develop on basis of these simulations sound subgrid models for numerical models of the middle atmospheric dynamics and tracer transport. Improved subgrid models are crucial for accurate climate projections and for predictions of ozone depletion. A Linux cluster with more than 2000 CPUs and high-speed interconnect is needed for the project. The memory and storage requirements are not clear from the application.

See Section 8.2.

BGC - Biomedical Grid Computation
Gunnar Norstedt, Center for Molecular Medicine, Karolinska Institute, Karolinska Hospital, Sweden

Discipline: Bioinformatics

According to the authors this project intends to facilitate biomedical research by providing access to different Grid applications, to computer expertise, programming tools and help desk functions and attempts to achieve a higher level of sharing of the bioinformatics resources already available on SweGrid. Periodic use of large computer resources is needed according to the proposal (without prescribing the character of the resource).

See Section 8.3.

BHOO - A Brief History Of the Oceans
Laurent Bertino, Mohn-Sverdrup Center, Nansen Environmental and Remote Sensing Center (NERSC), Bergen

Discipline: Geophysics (Oceanography)

The objective is to reconstruct a 100-year history of the global oceans, on the full ocean depths and including the polar sea-ice. The project aims to combine all available ocean observations and state-of-the-art physical ocean modeling by a space-time estimation procedure (data assimilation). Running such 100-year analysis would require billions of CPU hours and PetaBytes of storage. Cluster architectures with fast interconnect seem appropriate.

See Section 8.4.

BIOGRID - Bioinformatics for the grid
Bengt Persson, Linköping University, Sweden

Discipline: Bioinformatics

This project aims to establish a Nordic bioinformatic grid infrastructure with two aims: one is to gridify heavy computational tasks so that they can be performed faster and more rationally; the other is to coordinate bioinformatic infrastructure efforts in order to use the Nordic resources more efficiently. The focus is on making the grid solution easily available and coordinated with the Nordic countries rather than on the precise resource.

See Section 8.5.

BRAINSIM - Full scale, biologically detailed neocortex simulation
Anders Lansner, Computer Science and Communication, Stockholm University and KTH, Sweden

Discipline: Neuroscience

The objective is to develop the tools necessary to support brain-scale modeling. The results will increase the understanding of brain functions and help to diagnose and treat diseases. The project requires a Linux cluster with number of CPUs ”one or two order of magnitudes of what is currently available”.

See Section 8.6.

CarboClint - Carbon Cycle Climate Interactions
Christoph Heinze, Geophysical Institute and Bjerknes Centre for Climate Research, University of Bergen, Norway

Discipline: Geophysics

The objective is to quantify the main feedbacks between the large scale carbon cycle and climate. The carbon cycles of the oceans and land react to both climate change as well as rising atmospheric CO_2 concentrations and provide a feedback to the atmospheric greenhouse gas concentration. The goal is to implement a full representation of the marine and terrestrial carbon cycles into an existing climate model (Bergen Climate Model). The major challenge is to the build up of an optimal reference state or ”standard run” for perturbation runs or scenarios. This requires the integration of the full model system for at least 1500 years (the turnover time scale for the world ocean circulation). The resulting model systems are suitable for distributed memory architectures with fast processor/node interconnect. Fast and large disks are needed plus a reliable tape archiving system to store and retrieve data.

See Section 8.7.

CATOX - Nanocatalysis at the oxide-metal interface
Hannu Häkkinen, Nanoscience Center / Physics and Chemistry Departments, University of Jyväskylä, Finland

Discipline: Physics/Nanoscience

The proposed grand challenge aims at studying size-controlled, oxide-supported metal clusters and nanoparticles that have a wide unexplored range of physical and chemical properties and a high potential for chemical, photo-chemical, and catalytic applications. The can in a flexible way utilize resources up to 1000 CPUs (both capacity and capability) for this project, concentrated in a short period of time. The requirements on the memory and the storage are not specified.

See Section 8.8.

CLIQAIMD - AIMD simulations of complex liquids
Kari Laasonen, Department of Chemistry, University of Oulu, Finland

Discipline: Chemistry

The objective is to study several aspects of complex liquids and molecules in them using ab initio molecular dynamics (AIMD). They are interested of high concentration acids, reactions in solutions and ionic liquids. The codes they use run well on several platforms including PC clusters, massively parallel systems such a IBM Blue Gene, IBM Regatta and SGI/Altix. CPMD is very efficient both on vector and micro processors. An essential requirement is fast communication between the processors. At least an InfiniBand type of connection

is needed. The requirements on the storage are not specified.

See Section 8.9.

CMPQ - Computation of novel materials with predictive quality
Olle Eriksson, Department of Physics, Uppsala University, Sweden

Discipline: Physics

The scientific objectives are to perform large scale electronic structure calculations in order to compute properties of low-dimensional, nanometer-sized and strongly correlated materials with a predictive quality. This will lead to an increased understanding of unconventional superconductivity and complex magnetism. This project requires a large number of CPUs with possibility of massive parallelization (e.g., via fast interconnect) and shared memory. The optimum solution would be a distributed supercomputer of type Cray XT3 with 2-3 thousand AMD Opteron CPUs and an SGI Altix with 512 / 4 TB memory.

See Section 8.10.

COMMIS - Charge and Optical induced Manipulation of individual Molecules on Insulating Substrates
Adam S. Foster, Laboratory of Physics, Helsinki University of Technology, Finland

Discipline: Physics/Nanoscience

The proposed grand challenge aims to controllably manipulate the electronic structure of individual molecules on insulating substrates via charging and optical excitation. It is part of a larger project combining simulations and experiments. Fundamental understanding and advanced analysis will be provided by state-of-the-art simulations. Parallel computing resources, with fast communication over a large number of nodes and large shared memory capacities are needed. The requirements for memory and storage are not specified.

See Section 8.11.

FINALICE - Exploring the nuclear matter with the ALICE experiment
Markku Oinonen, Helsinki Institute of Physics, University of Helsinki, Finland

Discipline: (High Energy) Physics

The proposal is part of the ALICE experiment at CERN. ALICE data analysis - as for all the LHC experiments - will be shared among the participating countries and exploits GRID infrastructure. Particularly, in the Nordic countries distributed Tier-1 level computing resources are anticipated together with related end-user analysis resources (Tier-2). The project needs grid type resources and large storage resources.

See Section 8.12.

FINNCMS - Finnish CMS Project
Veikko Karimäki, Helsinki Institute of Physics, University of Helsinki, Finland

Discipline: (High Energy) Physics

The proposed project is part of the CMS project at CERN, and aims at analysis on data coming from LHC (Large Hadron Collider). Computer clusters are best suited to solve I/O bound data-analysis jobs and CPU-bound Monte Carlo simulations in Experimental High Energy Physics. Because all events are independent, there is

little need for inter job communication. Each LHC experiment including CMS, needs large storage resources to be able to process PetaBytes of data / year. Since no single site can provide all the needed storage alone, a distributed hierarchical computing model based on grid middleware has been adopted. A grid environment with large storage and computer clusters provides the e-Infrastructure resources needed to solve the CMS computing needs.

See Section 8.13.

GIG - Gene Interactions in General Pedigrees
Lars Rönnegård, Linnaeus Centre for Bioinformatics, BMC, Sweden

Discipline: Bioinformatics

The objective is to test fast non-iterative statistical methods for detection of genetic interactions in general pedigrees, and compare these with currently used iterative algorithms. Their expectations are that the project will have large impact on the research community because current methods make large scale studies of genetic interaction infeasible and hinders their understanding of how multi-factorial traits are controlled. Although the authors find a capacity resource in a form of grid sufficient for their project the number of CPUs required is very large (5000-10000 CPUs). The memory and storage requirements are moderate.

See Section 8.14.

GRAMS - GRAnd and not so grand challenges in computational Material Science
Anatoly B. Belonoshko, Applied Materials Physics, Institute of Materials Science, The Royal Institute of Technology, Sweden

Discipline: Material Science / Physics

The aim is to solve several problems that will help their understanding of how material function. To achieve this they are planning to use simulation techniques different and computationally more demanding from those that are normally used today. A Linux cluster with large number of CPUs, fast interconnect and large storage is required for this challenge according to the authors. However no precise details on the size of the resource is provided.

See Section 8.15.

HetCat - Unsolved Problems in Heterogeneous Catalysis
Kersti Hermansson, Materials Chemistry, Uppsala University, Sweden

Discipline: Material Chemistry

The objective is to find the catalytic implication of defects in metal oxides at elevated temperatures. According to the authors the project will produce the knowledge required for design of new catalysts which have large technical, economical and environmental benefits. The author does not describe the resources needed for their challenge.

See Section 8.16.

HiMuMoMo - New generation of force fields for predictive hierarchical multiscale molecular modelling
Prof. Aatto Laaksonen, Div. of Physical Chemistry, Arrhenius Laboratory, Sweden

Discipline: Chemistry

The great challenge of computational molecular science is according to the authors the transition from a descriptive to predictive level of molecular modeling. In order to substantially improve reliability and predictivity of molecular simulations, a new generation of force fields seems necessary. The aim of the author's project is to use large computer resources for obtaining such a force field. The optimal architecture for this project would be either a "supercluster" of many (tens of) thousands nodes, or, which may be more efficient in terms of performance/price ratio, a large(102-103) number of such clusters organized in a GRID environment. The requirements on the memory and the storage are not specified.

See Section 8.17.

LIPOPRO - Large-scale atomistic and coarse-grained modeling of lipoproteins
Ilpo Vattulainen, Institute of Physics, Tampere University of Technology, Finland

Discipline: Biophysics

The objective is to resolve the structure of lipoproteins used as carriers of cholesterol. The description states that the proposed work is completely new science and this kind of work has not been studied prior to this work and that the project is truly multi- disciplinary and of broad interest, providing state-of-the-art results to be published in high-impact journals. For the project, they need access to a supercomputing environment characterized by fast processors and state-of-the-art network. Storage and memory requirements are moderate.

See Section 8.18.

LOWSOUND - Generation and propagation of sound in low Mach number internal flows
Dr. Gunilla Efraimsson, Linné Flow Centre, KTH Aeronautical and Vehicle Engineering, Sweden

Discipline: Computational Fluid Dynamics

The objective is to increase the knowledge in sound generation and sound propagation for internal low Mach number flows. The computational challenge consist in performing direct numerical simulations (DNS) and large eddy simulations (LES) of both incompressible and compressible flows. According to the authors, this challenge can be fulfilled using capability kind of a system with distributed memory an high-speed interconnection.

See Section 8.19.

MHDTurbu - High resolution magnetohydrodynamic turbulence simulation
Axel Brandenburg, Nordita, AlbaNova University Center, Sweden

Discipline: Computational Fluid Dynamics

The aim of this project is to improve the understanding of the asymptotic scaling of magnetohydrodynamic (MHD) turbulence. According to the authors, this can be achieved through direct numerical simulations at the largest resolution possible, combined with "calibrated" large eddy simulations. The required resource for this project is a very large Linux cluster (more than 8000 CPUs) with high-speed interconnect. Although the authors have not prescribed the memory requirements, the storage requirements are large (120-1200 TB).

See Section 8.20.

MOIS - An integrative software system for management, analysis and design in molecular-level chemical and

biological calculations
Vidar R. Jensen, Department of Chemistry, University of Bergen, Norway

Discipline: Chemistry

The objective is the development and application of a general, open source software system for the handling of execution/control, management and analysis of experiments in molecular modeling (MOIS). According to the authors, the system will enable automated ”in silico” optimization of functional molecular structures and increase the potential and applicability of molecular-level computations. Automated searches for new molecular structures bring molecular-level computations into a new era with respect to dependence on high-end computational infrastructure and consumption of CPU resources. The proposal as such does not request access to large scale compute or storage resources, but a successful system can have considerable impact on the demand for such resources. Distributed memory architectures are suitable for the MOIS system. Components needed for realizing global access to MOIS system can be realised by GRID technology.

See Section 8.21.

MomentA - Nuclear Electromagnetic Moments via Atomic Theory
Pekka Pyykkö, Department of Chemistry, University of Helsinki, Finland

Discipline: Chemistry

The objective is to calculate high precision nuclear quadrupole moments. The proposed method consists of ab initio calculations of electric field gradients generated by the electronic cloud at the site of the nucleus done by a multiconfigurational Dirac-Hartree-Fock method. The CPU demands for the project are low (32-128 nodes). On the other hand, large memory (4GB per node or more) and large distributed temporary storage (100GB per node or more) are needed. Furthermore, a need for very long computing times (weeks per run) is also indicated.

See Section 8.22.

MURPF - Microphysics in Ultra Relativistic Plasma Flows
Åke Nordlund, Niels Bohr Institute, Copenhagen, Denmark

Discipline: Astrophysics

The objective is to obtain a realistic modeling of highly relativistic collisionless (plasma) shocks, as they are believed to exist for example in Gamma-Ray Bursts. The authors propose to carry out two types of computer simulations, to understand the formation, propagation, and radiative signatures of collisionless shocks: (i) 3D relativistic streaming experiments with a novel PIC code that incorporates photon-kinetic modeling, describing the full formation of a collisionless shock, containing both ions and electrons with a large ion-to-electron mass ratio and (ii) a fully 3D photon-particle interaction experiment, to test a proposed wakefield mechanism for particle acceleration and ultra high energy photon production in shocks.

See Section 8.23.

NANOGOLD - First principle design of ligand-protected gold nanoparticles
Hannu Häkkinen, Nanoscience Center / Physics and Chemistry Departments, University of Jyväskylä, Finland

Discipline: Physics/Nanoscience

The objective is to study the electronic structure and atomic geometry of nanoclusters by using well-defined

density functional theory (DFT)-based methods. Most crucial for the project is to have access to scalable parallel machine with fast interconnect network (faster than InfiniBand). Several hundred CPUs should be available for a single run. The requirements on the memory and the storage are not specified.

See Section 8.24.

NANOQUANT - Understanding Nanomaterials from the Quantum
Hans Ågren, Dept. of Theoretical Chemistry, Royal Institute of Technology, Sweden

Discipline: Chemistry

The objective is to demonstrate the potential impact of a large scale computer facility on new applications in pure quantum mechanics in the life and materials sciences. The authors propose to connect the inherent theoretical potential of first principles quantum mechanics to applications on nano-scale systems. For this, they plan to use a toolbox that is equipped with linear scaling technology with the potential to address various nano-scale systems, like catalytic microsurfaces, nano-particles, polymers, surface adsorbates, solutions, protein segments, functional centers in enzymes etc. With this technology it will be possible not only to follow the energetics and conformations but also the general electromagnetic properties and responses of these systems as they evolve and so to learn about their functionality. Large-scale computer hardware is needed in order to move the simulations from systems with (currently) 10.000 atoms to 100.000 atoms, thereby reaching a quantum description of the fundamental properties of a full protein. The Grand Challenge is divided in Computational Proteomics and Enzymology, Computational Atmospheric Chemistry and Computational Nanoparticle Technology.

See Section 8.25.

N-CC-GC - Nordic Climate Computing Grand Challenge
Markku Rummukainen, Rossby Centre, SMHI, Sweden

Discipline: Climate modeling

The objective is to gain a better understanding of climate change with a particular interest for the Nordic region. The aim is to use increased model resolution as this will help resolving effects due to narrow ocean passages in the North Atlantic floor. A capability resource (probably with several thousands of CPUs) with fast interconnect and low latency is required for this project. National resources are sufficient to carry out some of the scientific challenges but larger resources are needed to compete internationally. The storage requirements are of the size 100 TB. No details for the memory requirements are provided.

See Section 8.26.

NPQCD - Non-Perturbative Quantum Chromodynamics
Kari Rummukainen, Dept. of physical sciences, University of Oulu, Finland

Discipline: (High Energy) Physics

The objective is to study several lattice QCD related topics. The proposal includes most of the physicists in the Nordic region working in lattice QCD. The main challenge is to achieve maximum floating-point performance; the requirements for memory and I/O (speed and volume) are typically much less demanding. The software to be used is optimised using large clusters with very fast low-latency network connections or by special parallel computers.

See Section 8.27.

Pss1Å - Protein structure from sequence to resolution 1Å
Liisa Holm and Alex Bunker, Bioinformatics group of Biotechnology Center and Computational Pharmaceutical science group of the DDTC, University of Helsinki, Finland

Discipline: Bioinformatics

The goal of this project is to develop an algorithm whereby the structure of any given protein can be determined from sequence to an accuracy of 1. The authors state that dramatically increasing the accuracy of protein structure calculations can only be achieved through hybridizing two approaches, homology based approach, and ab initio type simulations. For the first capacity type resources (or even grid type) are needed (memory less than 1GB/core). For the latter part capability resources are needed. The requirements on the storage are not specified.

See Section 8.28.

QUADCASC - Quantum dynamics of collision cascades
Kai Nordlund, Department of Physical Sciences, University of Helsinki, Finland

Discipline: (Solid State) Physics

The proposed project aims to initiate the first density functional theory quantum mechanical simulations of the development of collision cascades in important materials like Si, Fe and carbon nanotubes. The resource needed is a capability architecture with efficient communication. The requirements on the storage are not specified.

See Section 8.29.

SAM - Stellar atmosphere modeling
Mats Carlsson, Institute of Theoretical Astrophysics, University of Oslo, Norway

Discipline: Astrophysics

The objective is to test and improve the physical description of the mechanisms occurring in stars in general, the methods and approximations involved in the construction of numerical stellar models as well as their computer implementation. In particular, the proposal includes the numerical modeling of the coupled convection zone-chromosphere-corona system and the high resolution simulation of a sun spot. A 3D radiation magnetohydrodynamic code is used for this. The basic challenge is the complex interplay of several processes acting on a wide range of spatial and time scales. The need to cover spatial scales from a few tens to thousands of kilometers and temporal scales from a few milliseconds to hours makes the problems compute intensive. Computational needs are estimated to be 80-140 days on 1000 CPUs. Compute clusters with a large number of CPUs (preferably 1000 or more) will be needed, but no fast processor/node interconnect.

See Section 8.30.

StaFraLa - Statistical Fracture Laws
Mikko Alava, Laboratory of Physics, Helsinki University of Technology, Finland

Discipline: Physics

The authors propose a grand challenge in the field of applying statistical mechanics to study the fundamental laws of fracture. The aim of the project is to have established, what kind of scaling of strength one finds in the currently used statistical fracture models, what are the theoretical reasons thereof, what kind of other signatures, experimentally, they would produce, and what the implications are for engineering and materials science. The

need of resources range from grid simulations to large MPP simulations (previous usage up to 1024 processors). The requirements on the memory and the storage are not specified.

See Section 8.31.

STP - Simulation of Tsunami Propagation
Xing Cai, Simula Research Laboratory, Lysaker, Norway

Discipline: Geophysics (Oceanography)

The objective is high-resolution simulation of tsunami propagation over large distances (oceans) and in a timely fashion. Challenges are in combining large-scale parallel computing with a hybrid mathematical and numerical strategies to take into account both certain small regions and shorelines, but also deal efficiently with large ocean areas. Careful consideration is needed with respect to the decomposition of the domain, load balancing, and the number of degrees of freedom (in the order of hundreds of millions). The project needs access to multi-thousand processor distributed-memory computers with TB-scale memory and preferably storage of the order of 100 TB.

See Section 8.32.

TURBSIM - Direct numerical simulation of turbulence, transition and separation for flows around bodies
Prof. Dan Henningson, Linné Flow Centre, KTH Mechanics, Sweden

Discipline: Computational Fluid Dynamics

The objective is the fully-resolved DNS of the turbulent flow around a realistic body, such as an airfoil. The challenge in the project is the high Reynolds number that characterises the flow and increases the requirements for the resolution. The resource required for the project is a Linux cluster with high-speed interconnect. The number of CPUs required for the project is not clear from the application but from the information about the software and the current resources it can be concluded that the authors aim to use more than 2000 CPUs. The memory requirements are of the size of 500 GB and storage of 50 GB is requested.

See Section 8.33.

Vioprocg - Violent processes in geophysics
Galen Gisler, Physics of Geological Processes, University of Oslo, Norway

Discipline: Geophysics

The objective is to study violent processes in geophysics, and in particular explosive volcanic eruptions. These problems are inherently multi-scale and multi-physics. The greatest challenge is in the sheer size and complexity of such problems . Similar three-dimensional calculations carried out at Los Alamos required 1.5 million CPU hours. Capacity architectures with fast interconnect seem appropriate. Large storage for output data and visualization is important.

See Section 8.34.

7.1 Categorization

In this section, we have aimed to categorize the Grand Challenge submissions according to a small number of pre-defined criteria. These are:

Discipline:
B = Biosciences
C = Chemistry
E = Earth sciences (incl. geosciences and climate modeling)
P = Physics (incl. nanoscience, material science, computational fluid dynamics and astrophysics)

Science: theoretical, applied, socio-economical impact

Collaboration: national, nordic, other countries. The numbers in these columns represent the number of participating persons/groups named in the Grand Challenge from the same country as the author, in other Nordic countries and outside the Nordic countries.

Type of resources: capacity, capability, SMP (shared-memory), grids

Quantity of resources: CPU, storage, memory

We use some guidelines to have a consistent classification of the submission. For quantity of resources, we used the following:

Nr. of CPUs	class
0-200	0
201-500	1
501-1000	2
>1000	3

Here, a '3' can be interpreted as a large-scale application.

Similarly, for storage needs we have used:

Amount of storage (TB)	class
0 - 0.5	0
0.5 - 1	1
1 - 5	2
>5	3

and for memory usage (per core):

Amount of memory (GB)	class
0-1	0
1-2	1
2-4	2
>4	3

In the matrix below, we have tried to categorize the Grand Challenge descriptions according to these criteria. The Grand Challenge descriptions do not always include detailed information about the expected needs for resources. In many cases, the researchers probably do not know the quantity of resources. This is probably

inherent to a Grand Challenge problem. In cases where the needed resources are not indicated, we have tried to estimate them on basis of the previous usage and knowledge on software.

		Science			Collaboration			Type of resources				Quantity of res.		
Acronym	discipline	theoretical	applied	socioec. impact	national	nordic	other countries	capacity	capability	grids	SMP	CPUs	storage	memory
ABICAT	P	X	-	-	1	-	-	-	X	-	X	2	1	1
ATMOTURB	P	X	-	-	-	-	-	-	X	-	-	3	2	2
BGC	B	-	X	-	1	-	-	-	-	X	-	1	1	1
BHOO	E	-	X	X	-	-	-	X	-	-	-	3	3	-
BIOGRID	B	-	X	-	3	3	-	-	-	X	-	1	1	1
BRAINSIM	B	X	-	-	1	4	-	-	X	-	-	3	3	3
CarboClint	E	X	X	X	2	-	-	X	-	-	-	1	3	2
CATOX	P	X	-	-	-	-	-	X	-	-	-	2	-	1
CLIQAIMD	C	X	-	-	-	-	1	-	X	-	-	2	-	-
CMPQ	P	X	-	-	2	-	-	-	X	-	X	3	2	3
COMMIS	P	X	-	-	-	-	-	-	X	-	X	3	-	-
FINALICE	P	X	-	-	1	-	-	X	-	X	-	1	3	-
FINNCMS	P	X	-	-	-	-	-	X	-	X	-	1	3	-
GIG	B	X	-	-	1	-	-	-	-	X	-	3	2	2
GRAMS	P	X	-	-	-	-	-	-	-	-	X	3	3	2
HetCat	C	X	-	-	-	1	-	X	-	-	-	1	1	1
HiMuMoMo	C	X	-	-	-	-	-	-	X	X	-	3	2	2
LIPOPRO	B	X	-	-	1	1	1	-	X	-	-	2	-	-
LOWSOUND	P	X	-	-	-	-	-	-	X	-	-	3	2	3
MHDTurbu	P	X	-	-	-	4	3	-	X	-	-	3	3	2
MOIS	C	-	X	-	1	-	3	X	-	X	-	1	2	1
MomentA	C	X	-	-	-	1	-	X	-	-	-	1	2	3
MURPF	P	X	-	-	2	3	-	X	X	-	-	3	-	2
NANOGOLD	P	X	-	-	-	1	-	-	X	-	-	3	-	-
NANOQUANT	C	X	-	-	-	3	-	-	X	-	-	3	-	3
N-CC-GC	E	-	X	X	2	5	-	-	X	-	-	3	3	3
NPQCD	P	X	-	-	2	1	-	-	X	-	-	3	-	-
Pss1A	B	X	-	-	-	-	-	X	X	X	-	3	-	-
QUADCASC	P	X	-	-	-	-	-	X	X	-	-	2	-	-
SAM	P	-	X	-	-	3	-	X	-	-	-	3	3	-
StaFraLa	P	X	-	-	-	-	1	X	X	X	-	3	-	-
STP	E	-	X	X	1	-	-	X	X	-	-	3	3	-
TURBSIM	P	X	-	-	-	-	-	-	X	-	-	3	2	3
Vioprocg	E	X	-	-	-	-	-	X	X	-	-	3	3	2

8 Grand Challenge descriptions

8.1 ABICAT - *Ab initio* molecular dynamics and statistical studies of catalytic reactions

Author. *Professor Kim Bolton, School of Engineering, University College of Borås, Borås, Sweden. Email: kim.bolton@hb.se*

1. Participating groups.
Docent Henrik Grönbeck, Competence Centre for Catalysis, Chalmers University of Technology, Goteborg, Sweden. Email: ghj@chalmers.se

The text for this Grand Challenge has been removed from this version of the report.

8.2 ATMOTURB - From Kolmogorov theories to fluid mechanics of climate change

Author. *E. Lindborg, Linné Flow Centre, Department of Mechanics, KTH, Stockholm, Sweden. Email: erikl@mech.kth.se*

1. Participating groups.
The project is proposed by the research groups forming the Linné Flow Centre, Department of Mechanics, KTH, Stockholm, Sweden. Director of the centre is Prof. D. Henningson (henning@mech.kth.se)

Additional proposals from the Linné Flow Centre: Turbulent airfoils (D. Henningson) and Aeroacoustics (G. Efraimsson)

The text for this Grand Challenge has been removed from this version of the report.

8.3 BGC - Biomedical Grid Computation

Author. *Gunnar Norstedt, Center for Molecular Medicine, Karolinska Institute, CMM L8-01, Karolinska Hospital, Stockholm. Email: gunnar.norstedt@cmm.ki.se*

1. Participating groups.
Co-applicants: Amilcar Flores Morales and Jacob Odeberg. Applied Bioinformatics, CMM, Karolinska Institute. Supporting applicants: Several scientists that in common use clinical material for research

The text for this Grand Challenge has been removed from this version of the report.

8.4 BHOO - A Brief History Of the Oceans

Author. *Laurent Bertino, Mohn-Sverdrup Center, Nansen Environmental and Remote Sensing Center (NERSC), Bergen, Norway. Email: Laurent.bertino@nersc.no*

1. Participating groups.
Mohn-Sverdrup Center, NERSC

The text for this Grand Challenge has been removed from this version of the report.

8.5 BIOGRID - Bioinformatics for the grid

Author. *Bengt Persson, Linköping University, Sweden*

1. Participating groups.
Bengt Persson, Linköping University, Sweden
Erik Sonnhammer, SBC, Stockholm University, Sweden
Ann-Charlotte Berglund Sonnhammer, SNIC, Sweden
Sven Stafström, NSC, Sweden
Juha Haataja, CSC, Finland
Inge Jonassen, CBU, Bergen Centre for Computational Science, Norway

The text for this Grand Challenge has been removed from this version of the report.

8.6 BRAINSIM - Full scale, biologically detailed neocortex simulation

Author. *Anders Lansner, Computer Science and Communication, Stockholm University and KTH. Email: ala@csc.kth.se*

1. Participating groups.
Jan Bjaalie, CMBN, UiO, Norway, j.g.bjaalie@medisin.uio.no
Gaute Einevoll, UMB, Norway, Gaute.Einevoll@umb.no
Sten Grillner, Neuroscience, KI, Sten.Grillner@neuro.ki.se
John Hertz, Nordita, Sweden and Denmark, hertz@nordita.dk
Kimmo Kaski, Finland, kaski@lce.hut.fi

The text for this Grand Challenge has been removed from this version of the report.

8.7 CarboClint - Carbon Cycle Climate Interactions

Author. *Christoph Heinze, Geophysical Institute and Bjerknes Centre for Climate Research, University of Bergen. Email: christoph.heinze@gfi.uib.no. Other key people: Karen Assmann, Kristof Sturm, Christophe Bernard*

1. Participating groups.
Nansen Environmental and Remote Sensing Center (NERSC), Bioforsk - Norwegian Institute for Agricultural and Environmental Research

The text for this Grand Challenge has been removed from this version of the report.

8.8 CATOX - Nanocatalysis at the oxide-metal interface

Author. *Dr. Hannu Häkkinen, Nanoscience Center / Physics and Chemistry Departments, University of Jyväskylä*

1. Participating groups.
Computational nanoscience group, University of Jyväskylä (Docent Hannu Häkkinen, key researchers Dr. Michael Walter and Docent Karoliina Honkala/Chemistry)

The text for this Grand Challenge has been removed from this version of the report.

8.9 CLIQAIMD - AIMD simulations of complex liquids

Author. *Prof. Kari Laasonen, Department of Chemistry, University of Oulu, Finland. Kari.Laasonen@oulu.fi*

1. Participating groups.
Dr. Ari Seitsonen
Priv.-Doz. Dr. Barbara Kirchener

The text for this Grand Challenge has been removed from this version of the report.

8.10 CMPQ - Computation of novel materials with predictive quality

Author. *Olle Eriksson, Department of Physics, Uppsala University, Sweden*

1. Participating groups.
Theoretical Magnetism, Uppsala University
Condensed Matter Theory, Uppsala University

The text for this Grand Challenge has been removed from this version of the report.

8.11 COMMIS - Charge and Optical induced Manipulation of individual Molecules on Insulating Substrates

Author. *Adam S. Foster, Laboratory of Physics, Helsinki University of Technology. Email: asf@fyslab.hut.fi*

1. Participating groups.
Surfaces and Interfaces at the Nanoscale (SIN) group, Laboratory of Physics, Helsinki University of Technology (the other groups involved are experimental or would not use Nordic resources)

The text for this Grand Challenge has been removed from this version of the report.

8.12 FINALICE - Exploring the nuclear matter with the ALICE experiment

Author. *Dr. Markku Oinonen, Helsinki Institute of Physics. Email: markku.j.oinonen@helsinki.fi*

1. Participating groups.
HIP/ALICE group, Helsinki Institute of Physics, University of Helsinki
JYFL/ALICE group, Department of Physics, University of Jyväskylä

The text for this Grand Challenge has been removed from this version of the report.

8.13 FINNCMS - Finnish CMS Project

Author. *Veikko Karimäki, Helsinki Institute of Physics. Email: Veikko.Karimaki@helsinki.fi*

1. Participating groups.
Helsinki Institute of Physics

The text for this Grand Challenge has been removed from this version of the report.

8.14 GIG - Gene Interactions in General Pedigrees

Author. *Lars Rönnegård, Linnaeus Centre for Bioinformatics, BMC, Uppsala. Email: lars.ronnegard@lcb.uu.se*

1. Participating groups.
Örjan Carlborg, Uppsala University, Sweden
Sverker Holmgren, Uppsala University, Sweden
Dietrich von Rosen, SLU, Sweden

The text for this Grand Challenge has been removed from this version of the report.

8.15 GRAMS - GRAnd and not so grand challenges in computational Material Science

Author. *Anatoly B. Belonoshko, Applied Materials Physics, Institute of Materials Science, The Royal Institute of Technology, Stockholm. Email: anatoly@fysik.uu.se*

1. Participating groups.
Applied Materials Physics, Institute of Materials Science, The Royal Institute of Technology, Stockholm

The text for this Grand Challenge has been removed from this version of the report.

8.16 HetCat - Unsolved Problems in Heterogeneous Catalysis

Author. *Kersti Hermansson, Materials Chemistry, Uppsala University. Email: kersti@mkem.uu.se*

1. Participating groups.
Knut Børve, Kjemisk Institutt, University of Bergen, phone: +47 555 8336, e-mail: knut.borve@kj.uib.no

The text for this Grand Challenge has been removed from this version of the report.

8.17 HiMuMoMo - New generation of force fields for predictive hierarchical multiscale molecular modeling

Author. *Prof. Aatto Laaksonen, Div. of Physical Chemistry, Arrhenius Laboratory, Stockholm University. Email: aatto@physc.su.se*

1. Participating groups.
Dr. Alexander Lyubartsev (Email: sasha@physc.su.se) Div. of Physical Chemistry, Arrhenius Laboratory, Stockholm University

The text for this Grand Challenge has been removed from this version of the report.

8.18 LIPOPRO - Large-scale atomistic and coarse-grained modeling of lipoproteins

Author. *Prof. Ilpo Vattulainen, Institute of Physics, Tampere University of Technology, Finland. Web: www.tut.fi/biophys and www.fyslab.hut.fi/bio/*

1. Participating groups.
The main collaborators of this research proposal are the groups of:

Prof. Mikko Karttunen, Department of Applied Mathematics, The University of Western Ontario, Canada. Web: www.softsimu.org

Ph.D. Marja Hyvönen, Wihuri Research Institute, Helsinki; together with the Institute. Web: www.wri.fi

Prof. Siewert-Jan Marrink, University of Groningen, The Netherlands.
Web: www.rug.nl/gbb/research/researchgroups/moleculardynamics/

Also, the proposal is part of the activities of MEMPHYSCenter for Biomembrane Physics, in the University of Southern Denmark, Odense, led by Prof. Ole G. Mouritsen. Web: www.memphys.sdu.dk

The text for this Grand Challenge has been removed from this version of the report.

8.19 LOWSOUND - Generation and propagation of sound in low Mach number internal flows

Author. *Dr Gunilla Efraimsson, Linné Flow Centre, KTH Aeronautical and Vehicle Engineering, Stockholm, Sweden. Email: gef@kth.se*

1. Participating groups.
The project is proposed by the research groups forming the Linné Flow Centre at KTH Stockholm, Sweden. The Director of the centre is Prof. Dan Henningson, henning@mech.kth.se.

Additional proposals from Linné Flow Centre: Turbulence, transition and separation (D. Henningson) and Atmospheric turbulence (E. Lindborg)

The text for this Grand Challenge has been removed from this version of the report.

8.20 MHDTurbu - High resolution magnetohydrodynamic turbulence simulation

Author. *Axel Brandenburg, Nordita, AlbaNova University Center. Email: brandenb@nordita.dk*

1. Participating groups.
Dr Maarit Korpi, University of Helsinki
Professor Åke Nordlund, NBI, Copenhagen
Dr Tarek Yousef, Cambridge and Trondheim
Dr Alexander Schekochihin, University of Cambridge
Professor Kandaswamy Subramanian, Pune
Dr Nils Erland Haugen, Trondheim
Dr John Brooke, University of Manchester

The text for this Grand Challenge has been removed from this version of the report.

8.21 MOIS - An integrative software system for management, analysis and design in molecular-level chemical and biological calculations

Author. *Vidar R. Jensen, Department of Chemistry, University of Bergen*

1. Participating groups.
Bjørn K. Alsberg (NTNU), Vidar R. Jensen (UiB), Nathalie Reuter (UiB/Unifob), Agnar Aamodt (NTNU), Helge Langseth (NTNU), Geerd H. F. Diercksen (Max-Planck-Institut für Astrophysik), Ross King (University of Wales), Hai Lin (University of Colorado at Denver).

The text for this Grand Challenge has been removed from this version of the report.

8.22 MomentA - Nuclear Electromagnetic Moments via Atomic Theory

Author. *Prof. Pekka Pyykkö, Department of Chemistry, University of Helsinki, Helsinki.*

1. Participating groups.
Jacek Bieron, University of Cracow

The text for this Grand Challenge has been removed from this version of the report.

8.23 MURPF - Microphysics in Ultra Relativistic Plasma Flows

Author. *Åke Nordlund, Niels Bohr Institute, Copenhagen University. Email: aake@astro.ku.dk*

1. Participating groups.
Niels Bohr Institute, University of Copenhagen, Denmark
Dep of Physics & Astronomy, University of Aarhus, Denmark
Stockholm Observatory, Stockholm, Sweden
Institute of Theor. Astrophysics, University of Oslo, Norway
Department of Physical Sciences, University of Oulu, Finland

The text for this Grand Challenge has been removed from this version of the report.

8.24 NANOGOLD - First principle design of ligand-protected gold nanoparticles

Author. *Dr. Hannu Häkkinen, Nanoscience Center / Physics and Chemistry Departments, University of Jyväskylä, Jyväskylä*

1. Participating groups.
Computational nanoscience group, University of Jyväskylä (Docent Hannu Häkkinen, key researchers Dr. Michael Walter and Docent Jaakko Akola)

Theory group at Competence Center for Catalysis (KCK), Chalmers, Sweden (Docent Henrik Grønbeck)

The text for this Grand Challenge has been removed from this version of the report.

8.25 NANOQUANT - Understanding Nanomaterials from the Quantum Perspective

Author. *Hans Ågren, Dept. of Theoretical Chemistry, Royal Institute of Technology, Stockholm. Email: agren@theochem.kth.se*

1. Participating groups.
Lektor Hans Jørgen Aa. Jensen, Dept of Chemistry, University of Southern Denmark Odense, Denmark.
Prof. Kenneth Ruud, Dept of Chemistry, University of Tromsø, Norway.
Prof. Dage Sundholm, Dept of Chemistry, University of Helsinki, Finland.

The text for this Grand Challenge has been removed from this version of the report.

8.26 N-CC-GC - Nordic Climate Computing Grand Challenge

Author. *Markku Rummukainen, Rossby Centre, SMHI, Norrköping, Sweden. Email: Markku.Rummukainen@smhi.se*

1. Participating groups.
Rossby Centre, SMHI, Sweden
University of Copenhagen, Denmark
Norwegian Meteorological Institute, Norway
University of Oslo, Norway
Stockholm University, Sweden
Lund University, Sweden
Finnish Meteorological Institute, Finland
Danish Meteorological Institute, Denmark

The text for this Grand Challenge has been removed from this version of the report.

8.27 NPQCD - Non-Perturbative Quantum Chromodynamics

Author. *Prof. Kari Rummukainen, Dept. of physical sciences, University of Oulu, Finland*

1. Participating groups.
University of Oulu (Prof. K. Rummukainen)
University of Helsinki (Prof. K. Kajantie)
University of Copenhagen (Dr. P.H. Damgaard)
University of Jyväskylä (Dr. K.J. Eskola, Dr. K. Tuominen)

The text for this Grand Challenge has been removed from this version of the report.

8.28 Pss1Å - Protein structure from sequence to resolution 1Å

Author. *Liisa Holm and Alex Bunker*

1. Participating groups.
Bioinformatics group of Biotechnology Center and Computational Pharmaceutical science group of the DDTC (Faculty of Pharmacy)

The text for this Grand Challenge has been removed from this version of the report.

8.29 QUADCASC - Quantum dynamics of collision cascades

Author. *Kai Nordlund, University of Helsinki, Finland*

1. Participating groups.
Materials science simulation team, Accelerator Laboratory, University of Helsinki

The text for this Grand Challenge has been removed from this version of the report.

8.30 SAM - Stellar atmosphere modeling

Author. *Mats Carlsson, Institute of Theoretical Astrophysics, University of Oslo, mats.carlsson@astro.uio.no*

1. Participating groups.
Institute of Theoretical Astrophysics, University of Oslo (ITA) Niels Bohr Institute, Copenhagen University, Denmark (NBI) Institute of Solar Physics, Stockholm, Sweden (ISP) Astronomical Observatory, Uppsala University, Sweden (AO)

The text for this Grand Challenge has been removed from this version of the report.

8.31 StaFraLa - Statistical Fracture Laws

Author. *Mikko Alava, Laboratory of Physics and COMP, Helsinki University of Technology, Finland. Email: mja@fyslab.hut.fi*

1. Participating groups.
Mikko Alava, HUT
Phani Nukala, Oak Ridge National Laboratory, USA
Stefano Zapperi, Univ. Roma I, La Sapienza & CNR-ISC, Italy

The text for this Grand Challenge has been removed from this version of the report.

8.32 STP - Simulation of Tsunami Propagation

Author. *Xing Cai, Simula Research Laboratory, Lysaker, Norway. Email: xingca@simula.no*

1. Participating groups.
Department of Scientific Computing, Simula Research Lab; Department of Mathematics, University of Oslo; Norwegian Geotechnical Institute

The text for this Grand Challenge has been removed from this version of the report.

8.33 TURBSIM - Direct numerical simulation of turbulence, transition and separation for flows around bodies

Author. *Prof. Dan Henningson, Linné Flow Centre, KTH Mechanics, Stockholm. Email: henning@mech.kth.se*

1. Participating groups.
The project is proposed by the research groups forming the Linné Flow Centre at KTH Stockholm, Sweden. Director of the centre is Prof. Dan Henningson.

Additional proposals from Linné Flow Centre: Atmospheric turbulence (E. Lindborg) and Aeroacoustics (G. Efraimsson)

The text for this Grand Challenge has been removed from this version of the report.

8.34 Vioprocg - Violent processes in geophysics

Author. *Galen Gisler, Physics of Geological Processes, University of Oslo. Email: galen.gisler@fys.uio.no*

1. Participating groups.
Physics of Geological Processes, University of Oslo

The text for this Grand Challenge has been removed from this version of the report.

A Proposal form

A preformatted form had to be used to submit a Grand Challenge description. The form asked for the information listed below. The items 1-7 have been included in this report.

1. **Title.**
2. **Acronym.** Up to 8 characters.
3. **Author.** Name and contact information.
4. **Participating groups.** List of groups that would participate in the Grand Challenge project.
5. **Scientific objectives.** A description of the Grand Challenge and its relevance. The description must be concrete and well-defined and address a specific unresolved scientific problem. (max 500 words)
6. **Method.** The proposed method to address the Grand Challenge problem. Include a brief (but correct) overview of the current state-of-the-art in the Nordic countries and internationally for tackling the problem and how the proposed method compares to this. The description of the proposed method must be concrete. (max 500 words)
7. **Challenges.** A description of the challenges in terms of computation and/or data to address the Grand Challenge. (max 250 words)
8. **Current activities.** A description of how the current research activity is related and suitable to address the Grand Challenge problem. What will be the benefits to the research activity in case you would get access to much larger (compute and data) resources. Include a justification that the results are very likely to appear in the most prestigious scientific journals in the field and will bring the research and group(s) to the forefront of international research. (max 300 words)
9. **Intermediate goals.** Intermediate goals (and a tentative time schedule) that are required to address the Grand Challenge. (max 200 words)
10. **Current resources.** A short account of the computational resources used today for addressing the Grand Challenge. This includes CPU-hours per year, size of the (parallel) applications, memory and storage demands, and the names of the facilities that are currently used. (max 100 words)
11. **Software.** A justification that the application software that is available today is suitable to address the Grand Challenge and will be able to make efficient use of large-scale e-Infrastructure. In particular, the scalability of the application(s) in terms of number of processors should be detailed. (max 200 words)
12. **Type of resources.** What are the type of e-Infrastructure resources that will be most beneficial for addressing the problem, e.g., capacity architectures (e.g., compute clusters), capability architectures (e.g., shared-memory or vector-processor architectures), large storage, or a grid environment. (max 100 words)
13. **Support.** A list of key people that support the proposal (provide name, affiliation, and email address).
14. **Other groups.** A list of research groups in the Nordic countries (and optionally other countries in Europe) with related research activity that may have an interest in the proposal.

B Workshop

The programme of the workshop that was held at CSC in Espoo on March 12, 2007.

09:30-10:00	Registration and coffee
10:00-10:10	Leif Laaksonen, CSC, Opening and welcome
10:10-10:30	Jura Tarus, Siniša Krajnović, Summary on the survey
10:30-11:00	Heikki Järvinen, Finnish Meteorological Institute, Nordic Climate Computing Grand Challenge
11:00-11:30	Aatto Laaksonen, Stockholm University, New generation of force fields for predictive hierarchical multiscale molecular modeling
11:30-12:00	Boris Gudiksen, University of Oslo, Stellar Atmosphere Modeling
12:00-13:00	Lunch
13:00-13:30	Geir Kleivstul Pedersen, University of Oslo, Simulation of tsunami propagation
13:30-14:00	Alex Bunker, University of Helsinki, A new advanced hybrid Monte Carlo molecular dynamics algorithm for protein structure refinement and its application to the ABC transporter super-family of membrane proteins
14:00-14:30	Bengt Persson, Linköping University, Bioinformatics from the grid
14:30-15:00	Coffee
15:00-15:30	Anders Lansner, Stockholm University and KTH, Full scale, biological detailed neocortex simulation
15:30-16:00	Laurent Bertino, Mohn-Sverdrup Center / Nansen Environmental and Remote Sensing Center, Bergen, A brief history of the oceans
16:00-16:30	Dage Sundholm, University of Helsinki, Understanding nanomaterials from the quantum perspective
16:30	Informal gathering