CCS Report: PART III

Vision and Strategies of Center for Computational Sciences

October 2007

Center for Computational Sciences

University of Tsukuba

Table of Contents

1. Introduction

This document, CCS report: PART III, presents the vision and strategies of the Center for Computational Sciences, University of Tsukuba. The vision focuses on the time frame of 2007-2011.

The report has been prepared for review of the Center scheduled from 30 October to 1 November 2007.

The purpose of the review is to receive an outside examination on the research activities and their outcomes of the Center in view of the founding objectives, and to incorporate the recommendations from the review for future developments of the Center.

Since details of the research results are prepared as a separate report, this report puts more focuses on the vision, strategies and plans of the Center. An overview, chronology and the organizational aspects of CCS are described in CCS Report: PART I, and details of the research plans of each division and group are described in CCS Report: PART II.

2. Mission and Role of CCS

The mission of the Center of Computational Sciences is to promote scientific discovery by computational science through the application of advanced computing technologies, and to support research of computational science in Japanese universities and institutes by operating leading-edge computing systems.

The center carries out research in material, life and environmental sciences using large-scale numerical simulations and data analyses. To advance these studies, the center carries out research and development in high-performance computing and networking systems, and applications of novel information technologies. The center is an inter-university facility open to researchers throughout Japan.

From the nationwide viewpoint of the development of scientific research, the CCS is a unique research “inter-university” organization which jointly carries out “interdisciplinary” research in the computational sciences, including development of cutting-edge high-performance computer systems. While the computational science activities in each of these disciplines is conducted by university labs and institutes specific to that discipline, we conduct computational science in each discipline, as well as “interdisciplinary” in an attempt to integrate many research fields. In contrast to institutes that are organized with a focus on computer sciences, our center is concerned with both computer sciences and computational sciences. For example, the National Institute of Informatics (NII) focuses on information science, but not on computational science or the development of high-performance computer systems. In Japan, there are supercomputer centers in major universities that only focus on the operation and service of their computing facilities, and not on computational science research.

From the viewpoint of the development of scientific research within our university, our center is the topmost research center, and it actively supports budgets, faculty positions, and facilities. The CCS faculty is a graduate school and contributes to the education of computational science.

3. “Interdisciplinary Computational Science” and Vision of the Center for Computational Sciences

Computational science, which enables us to explore uncharted fields of science through applications of high-performance computing, is a third paradigm of scientific research that has become indispensable for the development of science and technology in the 21st century. The methods of large-scale numerical analyses that form the pillar of computational science are common over a wide range of scientific disciplines, from fundamental science to material, life and geo-environmental sciences. This common foundation promotes the development of mutual and intimate connections among the various scientific disciplines and allows us to approach science holistically from a global perspective. This in turn leads to a global approach to nature, life and the environment, which is indispensable for the harmonious advance of science and technology, and consequently, of mankind. Furthermore, progress in computational science will help produce the next generation of scientists that will conduct research from a global perspective.

Development of computational science cannot be satisfactorily achieved solely by the scientists who use computers, and close collaboration and synergy is required with computer scientists with expertise in computer hardware, software, algorithms and programming, and information scientists with expertise in data and media engineering and technologies. Therefore, in order to establish a solid basis for the development of computational science, it is necessary to build a system that integrates the activities at the forefront of scientific research with those of computer science and information science.

Computational science is thus a multi-disciplinary area: each discipline X may have a “computational X” where X is an established discipline such as mathematics, biology, finance, physics, chemistry, or social science. For example, physics has “computational physics”. To date, most researchers of “computational X” make use of an existing supercomputing center to execute their tasks. Often these activities are short-term projects that have nothing to do with other computational sciences, such as “computational Y.” However, some knowledge, models and algorithms, can sometimes be shared among the different computational sciences. In view of solutions to problems, extensive knowledge and integration of computational sciences from different disciplines are necessary.

Computational science is a cutting-edge, indispensable multidisciplinary and interdisciplinary area for the development of science and technology in the 21st century. We define “Interdisciplinary Computational Science” using the following two aspects:

- Integration and collaboration of computational science and computer science:
  - High-performance computing systems (HPC) will be increasing in size and become more complex in configuration. HPC technologies are indispensable for optimizing the possible benefits associated with these advances.
  - Continuous development and investment are required to facilitate the next generation of computational science.

- The common base of computational science brings mutual intimate connections among various science disciplines and allows us to approach science as a whole from a global perspective.

University of Tsukuba and the Center for Computational Sciences have undertaken a long-standing effort to establish an international and national research center for the advancement of computational science by taking advantage of the top-level research achievements produced and accumulated at the Center.

The CCS’s vision is as follows:

The CCS aims to be a worldwide COE that develops “Interdisciplinary Computational Science” as a new area essential to 21st century advances in science and technology, and support it continuously from global and long-term perspectives.

The vision of CCS is to further enhance and develop the close collaboration among various science disciplines by applying the activities undertaken in the computer and information sciences pursued at the Center, which is quite unique in the world. In doing so the Center hopes to establish a system that advances computational science from a global perspective and simultaneously produces the next generation of scientists who will push science forward within a global context.

We aim to establish an organization for computational science research at the world’s highest level through a close integration of science and computer and information sciences. On this basis, we aim to meet the grand challenges associated with various scientific disciplines such as fundamental, material, life and geo-environmental sciences. Simultaneously, the organization should function as an advanced educational system for educating the next generation of computational scientists.

4. Strategies and Plan

To realize our vision and accomplish our mission, we will implement the following strategies and plans. The vision and this plan focus on the time frame of 2007-2011.

[Research & Development] Integrate advances of scientific research with those of computer science and information science, and thereby establish and promote new interdisciplinary computational science.

- Execute “Interdisciplinary Computational Science Promotion Program” or the operation of inter-university facilities such as PACS-CS for promotion and acceleration of interdisciplinary computational science.
- Integrate computational science with grid/network computing, sensing technologies, data engineering and informatics to explore “new” concepts of computational science.

[Education & Public Relations] Educate and produce the next generation of scientists who push forward new interdisciplinary computational science from global viewpoints.

- Dual-degree (double major) program
- Campus-wide courses on “computational science” for graduate students
- HPC seminar series
- Workshops and symposiums

[Infrastructure & Service] Support research by operating a leading-edge, advanced, large-scale computing facility for inter-university computing resources, and by development of an advanced computing infrastructure.

- Acquire & operate the next supercomputer as a part of the inter-university computing resources, making an “open supercomputer” alliance with the University of Tokyo and Kyoto University (T2K).
- Contribute to the “national” next-generation supercomputer project, and take leadership of the nationwide computing infrastructure, and explore the next-generation high-performance platform.

[Collaboration & Alliance] Promote alliances and collaboration, and establish international/national networks of researches on computational science.

- Ally the Center with “supercomputer” centers at other universities, the next-generation supercomputer center of RIKEN, and research laboratories in Tsukuba
- Promote international collaborations by “Computational Science Researchers Invitation Program,” and support virtual organization using Grid technologies.

4.1 Research and Development

The Interdisciplinary Computational Science Promotion Program will be one of the main activities for promotion and acceleration of interdisciplinary computational science. The purpose of this program is to encourage research activities of computational sciences. The computational power of the PACS-CS is provided to researchers throughout Japan through this program.

The Interdisciplinary Computational Science Promotion Program has two categories:

- The **"**Interdisciplinary Collaboration Program" will promote the interdisciplinary research activities of different disciplines. We will support the establishment of a network of interdisciplinary research projects and encourage all challenges of interdisciplinary computational science. For this program, we will organize an "Interdisciplinary Collaboration Program Support Committee" to help match research groups from different fields. We will follow up the research activity produced by the interdisciplinary collaboration.
- A large-scale scientific simulation program will push forward the challenges of several fields in computational sciences by providing the computational power of the PACS-CS. We will review proposals and concentrate our computational power to make new scientific discoveries and to follow up the scientific results.

We already started the program in October 2007 for the PACS-CS. The detail is described in CCS Report: PART I. Also, opening the FIRST simulator that was previously dedicated to computational astrophysics at University of Tsukuba is under consideration for the next academic year.

Computational science has made rapid and considerable progress. The computational capability of supercomputers is now on the verge of surpassing a petaflops (10^{15} floating operations per second), which is roughly a millions times faster than today's PCs. This advance allows us to explore a wide range of phenomena through computer simulations, which was impossible in the past, e.g., the creation and evolution of the universe, the quantum properties of nano- and biomaterials and their contribution to life, and global climate changes.

At the same time, the development of intelligent information processing technologies is beginning to enable the management and analysis of enormous amounts of data in the petabyte ranges, and this development is accelerating discoveries in many science disciplines, such as genome analyses, high-energy accelerator experiments, astronomical

observations and satellite observation of geo-environments. Furthermore, the recent growth of high-speed Internet is creating a communication environment that has enabled scientific research and education to be carried out on a global scale.

Data, data analysis, and visualization are vital for progress in the increasingly data-intensive realm of science and engineering research and education. Any cogent plan addressing cyberinfrastructure must address the phenomenal growth of data in all its forms. Scientists and engineers are producing, accessing, analyzing, integrating, storing and retrieving massive amounts of data daily. Furthermore, this is a trend that is expected to see significant growth in the very near future as advances in sensors and sensor networks, high-throughput technologies and instrumentation, automated data acquisition, computational modeling and simulation, and other methods and technologies materialize.

In short, computational science is rapidly developing into a unified framework, or a "new" concept of computational science, in which large-scale modeling and simulation, large-scale data analysis, and experiments and observations are integrated to solve challenging issues in various branches of science. As such, computational science has now become an indispensable tool to better understand 21st century sciences.

Several challenges have already begun. The Computational Media Group and the Grid Computing Group have started the project, "Sensing Web," in collaboration with Kyoto University and other universities for new integration of sensing technology, the Internet and grids. Also, the Computational Informatics Division and High-performance Computing System Division, Global Environment and Biological Sciences Division have proposed an idea to integrate sensing, databases and computation to capture and observe phenomena of humans and nature in the real world; this would be called "Real-world Computational Informatics."

4.2 Educational Activities

The Computational Science Dual-Degree (double major) Program will enable a graduate student in a doctoral program to simultaneously belong to a masters program of a different graduate school, and receive both a doctoral degree in science and a master's degree in computer science, or vice versa, upon graduation. The program is intended to educate researchers who can advance new interdisciplinary computational science from global viewpoints. We will design curriculum and courses for advanced computational science and begin this program in 2008. The first expected combination of majors is a Ph. D in Physics and a master's degree in Computer Sciences.

Campus-wide courses on computational sciences for graduate students are other planned

educational activities. Currently, we propose two courses:

- Computational Science Literacy
- High-performance Parallel Computing Technology for Computational Sciences (overlapped with the HPC Seminar)

The lectures will be given by faculty members of CCS and will be worth one unit of graduate credit. These courses will start in the summer of 2008.

The HPC Seminar presents knowledge, methods and techniques for programming modern high-performance computer systems, including recent microprocessors and their performance, and parallel programming. This is an inter-university activity open for researchers outside the Center, even researchers in companies not related to the university. The seminar is held during 2 or 3 days in the summer. The number of participants in the last seminar in 2007 was more than 80. It is also was broadcasted via the Internet.

We will organize workshops and symposiums. We have a plan for a “Long-term Computational Science Workshop,” in which an advanced subject encompassing multiple disciplines of science and computer science will be chosen, and participants who can stay from a few weeks to a few months are encouraged to carry out interdisciplinary research collaborations. In addition to this, we plan to have an annual international symposium, which will be part of an existing symposia series covering various areas of computational science that is held every year in Tsukuba (there have been 10 symposia over the past five years).

4.3 Acquisition and Operation of the Next Supercomputer and T2K Project

In University of Tsukuba, we had a supercomputer system (VPP5500) which was operated by the Academic Computing and Communications Center until 2006. The university decided that the Center of Computational Sciences should acquire and operate the next supercomputer. This next supercomputer may be used as part of the inter-university computing resources, like PACS-CS.

For acquisition of the system, we, University of Tsukuba, have made an alliance with University of Tokyo and Kyoto University, since the three universities were planning to procure their next supercomputers in the same time frame to replace their current ones. We agree to cooperate and make an alliance to share the main system architecture on three machines. The alliance should take the following topics into account:

- Common Architecture - Nodes and interconnections

- Use of Open System/Software – based on a Linux cluster
- Source code level compatibility – easy code porting & grids
- Grid computing – resource sharing both CPU and data

We call this conceptual design "Open Supercomputer," and our alliance "T2K Open Supercomputer Alliance." We define "Open Supercomputer" as follows:

- Openness of basic architecture
  - Use of a high-performance commodity processor
  - Use of the leading processor in the market facilitates the high-performance, high-density and low-power characteristics computing facility in the supercomputer center

- Openness of system software
  - Use of open source software based operating system and libraries
  - Provide a seamless environment for both multi-scale and multi-sites to enable users to develop various fields and sizes of applications

- Openness of users' needs
  - Support much wider area of users than traditional supercomputer centers
  - Explore new application fields

We expect that this architecture will be useful for bridging the gap between/connecting small laboratories and supercomputing centers.

We expect a performance of more than 80 TFLOPS performance for our site. According to the schedule, the bid will be done at the end of December 2007, and installation will be completed by the end of May. The operation will be started in the beginning of June 2008.

We will continue the alliance with University of Tokyo and Kyoto University also for operation and research on computational science.

4.4 Contribution to the Next-generation Supercomputer Project

In 2006, the Japanese government launched the National Project (The Next-Generation Supercomputer Project) to develop a peta-scale computer system (called the "Keisoku" computer), which has the performance of more than 10 PFLOPS. The Project spans seven years, from 2006 to 2012, and RIKEN (Research Institute for Science and Chemistry) has been selected to conduct the project.

In September 2006, RIKEN and University of Tsukuba signed a collaboration agreement on the Next-Generation Supercomputer Project. Based on this agreement, several members of the Center for Computational Sciences have been appointed as visiting fellows of RIKEN, and have been involved in the project. More information about related projects are described in CCS Report: PART I.

CCS will contribute not only to the design of the system but also to the evaluation and performance improvement of real target application codes. The Center and RIKEN are setting up a research collaboration program under the General Agreement for tuning performance of QCD, RS-DFT and FFT codes for the Next-Generation Supercomputer as peta-scale applications. Such a contribution is made possible by the unique environment of CCS, where researchers of computer science and computational science are working together.

In 2012, the peta-scale system is scheduled to become operational. The system will have a great impact on the communities of computational science, because the performance of this system will be huge in the nationwide computing infrastructure. As a leading research organization, it is a very important issue for us to take leadership regarding the operation of the peta-scale system in the nationwide computing infrastructure.

Besides the contribution to the next-generation supercomputer project, we have to explore the next-generation high-performance computing platforms, including arithmetic accelerators such as GRAPE-DR, Clear-speed, GPU and many-core architectures, which integrate many functions on one chip. These architectures are promising for some special-purpose applications. Since our center has a long history of collaboration with computational science and computer science, this approach is possible just in our center. The FIRST project is a very good example of what we have achieved so far in this approach.

4.5 Collaboration and Alliance

Finally, collaborations and alliance on the research of computational science are important factors for promoting and accelerating interdisciplinary computational science.

For nationwide alliance, we are going to strengthen the T2K alliance with Tokyo and Kyoto, as well as the connection to the coming peta-scale system, not just for computing facilities, but also for research on computational science.

Our university is located in Tsukuba Science City, which has many government research organizations. We already have connections to the major research institutions in Tsukuba

(e.g., High Energy Accelerator Research Organization (KEK), Advanced Institute for Science and Technology (AIST), National Institute for Materials Science (NIMS), Meteorological Research Institute (MRI), and National Institute for Environmental Studies (NIES)). The strategy is to strengthen research connections among the institutions in Tsukuba for the research of computational science.

National Institute of Informatics has been carrying out the "Cyber Science Infrastructure Project," which is building a grid infrastructure for academic research between the universities and research institutions. Because of its well-known research achievement record, the Center for Computational Sciences, which forms one of the centers, has been requested to participate in the project, and have been playing an active role. We will host the virtual organization (VO) of several disciplines by grid technology. We already host the lattice QCD communities' VO through the JLDG activities.

We will strengthen and expand the wide spectrum of international and domestic collaborations already underway, and use these connections as bridges to send and invite researchers and students for research collaborations. We have a plan to run the "Computational Science Researchers Invitation Program," which will invite researchers who are at the forefront of research in science and computational science.

The list of international collaborations are as follows:

i) **International Lattice Data Grid** in particle physics (Edinburgh University (UK), DESY (Germany), Univ. di Roma (Italy), FNAL/BNL/JLAB (USA))
ii) **Astronomical Radiation Hydrodynamics Research** (National Astronomical Observatory (Japan), SISSA (Italy))
iii) **Nano-Bio Science Collaboration** (ETH (Switzerland), Princeton University and Vanderbilt University (USA), Advanced Institute for Science and Technology (AIST), National Institute for Materials Science (NIMS))
iv) **Quantum Structural Biological Science Collaboration** (AIST, Hyogo Prefectural University (Japan), ETH and IBM Zurich (Switzerland))
v) **Quantum Dynamics Collaboration** (RIKEN (Japan), Washington University (USA))
vi) **Warming of Arctic Research** (IARC of Alaska University/Hawaii University (USA), Pusan University (Korea), Academia Sinica (China))
vii) **France-Japan Grid Computing** (National Institute of Informatics (Japan), INRIA (France), Pennsylvania State University (USA))
viii) **Information Integration/Knowledge Discovery Research** (Carnegie-Mellon University and Georgia Institute of Technology (USA), Chubun University (Hong-Kong)).

Within the Center, we will continue collaborations between different divisions and groups to bring synergy. The following collaborations are possible and some of them have already begun:

i) **Particle Physics and Astrophysics division** collaborates with **High-performance Computing Systems and Computational Informatics divisions** on the development of peta-scale computing systems, algorithms, and a global-scale data grid formation

ii) **Astrophysics group** and **Geo-environment group** collaborate on the development of physical modeling which relies principally on radiation transport

iii) **Materials and Life Science division** and **High-performance Computing Systems division** collaborate to develop, implement and optimize quantum mechanical first-principle methods for multi-atom systems such as the real-space and real-time density functional methods which are scalable to petaflops range

iv) **Life Science group** and **Biological Science group** collaborate on the inter-relation at the molecular level between biological functions and the evolution tree

v) **Computational Informatics division** and **Global environment group** collaborate on the development and operation of the climate/weather database and use it as a testbed for innovations in knowledge discovery from large-scale databases.